# Identification of Differentially Expressed Genes in HPV-Associated Cancers Using Gene Expression, Tissue, and MicroRNA Microarrays

by

**Ivan Martinez**

M.Sc. Cell Biology, National Autonomous University of Mexico, 2001

Submitted to the Graduate Faculty of

the School of Medicine in partial fulfillment

of the requirements for the degree of

Doctor of Philosophy

University of Pittsburgh

2007

**UNIVERSITY OF PITTSBURGH**

**SCHOOL OF MEDICINE**

This dissertation was presented

by

**Ivan Martinez**

It was defended on

March 30, 2007

and approved by

**Dr. Panagiotis Benos**, Assistant Professor, Human Genetics

**Dr. Neal DeLuca**, Professor, Molecular Genetics and Biochemistry

**Dr. Stefan M. Duensing**, Assistant Professor, Molecular Genetics and
Biochemistry

**Dr. Robert L. Ferris**, Associate Professor, Otolaryngology and of Immunology

**Dr. Saleem A. Khan**, Professor, Molecular Genetics and Biochemistry

# IDENTIFICATION OF DIFFERENTIALLY EXPRESSED GENES IN HPV-ASSOCIATED CANCERS USING GENE EXPRESSION, TISSUE, AND MICRORNA MICROARRAYS

**Ivan Martinez, PhD**

**University of Pittsburgh, 2007**

## ABSTRACT

Infections with high-risk human papillomaviruses (HPVs) have been implicated in the pathogenesis of cervical carcinoma and a subset of squamous cell carcinoma of the head and neck (SCCHN). In this study, we compared the cellular gene expression profiles of HPV16-positive and HPV-negative oropharyngeal carcinomas with those of the normal oral epithelium. Using a high-density oligonucleotide microarray containing 22,215 human transcripts, we showed that 397 and 162 genes were differentially expressed in HPV16-positive and HPV-negative SCCHN, respectively, compared to the normal oral epithelium. Our studies also identified 59 differentially expressed genes in HPV16-positive SCCHN as compared to both HPV-negative SCCHN and normal oropharyngeal tissues. Such up-regulated genes included those involved in nuclear structure and meiosis (*SYCP2*), DNA repair (*RFC5*), and transcription regulation

(*ZNF238*). Genes involved in proteolysis (*KLK8*) and signal transduction (*CRABP2*) were found to be down-regulated in HPV-positive SCCHN. Our results reveal specific gene expression patterns in HPV16-positive oropharyngeal squamous carcinomas and suggest that HPV infection could play an important etiologic role in these tumors. In another study using the same high-density microarray platform, we have analyzed the cellular gene expression profiles of five HPV-16 and two HPV-18 positive cervical cell lines, one HPV-negative cervical carcinoma cell line, and normal cervical tissue. Our results showed that 877 and 536 genes were differentially expressed in the HPV-positive cell lines compared to the normal cervix tissue and the HPV-negative cervical carcinoma cell line C-33A, respectively. We also found that a total of 57 genes were differentially expressed in the HPV-positive cell lines as compared to both the normal cervix and C-33A. Differentially expressed genes including those involved in cell proliferation such as the L-type amino acid transporter 1 (LAT1, also known as SLC7A5) and gene expression regulation like nucleosome assembly protein 1-like 3 (NAP1L3) were found to be affected for the first time in cervical cell lines. *In situ* hybridization of LAT1 and NAP1L3 mRNA performed using tissue-arrays (containing ~50 different cervical tumor samples per slide) showed that these genes are also affected in their expression in tumor tissues. These results could lead to the identification of new cellular pathways affected by the presence of HPV in cervical cells. We have also carried out studies to determine whether the expression of human microRNAs (miRNAs; small non-coding RNAs that have the ability to regulate gene expression) are affected by the presence of

HPV DNA. For this purpose, we analyzed the expression of miRNAs in HPV-16 positive cervical cell lines and tissues. Twenty-seven miRNAs were differentially expressed in cervical cell lines containing integrated HPV-16 DNA compared to the normal cervix, while, only 6 miRNAs were differentially expressed in a cell line containing episomal HPV-16 DNA compared to the normal cervix. Furthermore, 10 miRNAs were affected in their expression in cell lines containing integrated HPV-16 DNA compared to C-33A. Interestingly, microRNA-218 (miR-218) was specifically underexpressed in cell lines, cervical lesions and cancer tissues containing integrated HPV-16 DNA as compared to both the HPV-negative cell line C-33A and the normal cervix. Expression of the HPV-16 E6 oncogene in transfected cells reduced miR-218 expression, and conversely, RNA interference of E6/E7 oncogenes in an HPV-16 positive cell line increased miR-218 expression. We also showed that miR-218 expression parallels that of the tumor suppressor gene *SLIT2* whose intron encodes miR-218. Furthermore, exogenous expression of miR-218 in HPV-16 positive cell lines decreased expression of the epithelial-specific gene *LAMB3* which is involved in cell migration and tumorigenicity. These findings demonstrate specific regulation of cellular miRNAs in the presence of an HPV oncogene and may contribute to a better understanding of molecular mechanisms involved in cervical carcinogenesis.

# TABLE OF CONTENTS

# LIST OF TABLES

# LIST OF FIGURES

**Chapter 1**


**GENERAL INTRODUCTION**

## 1.1    THE BIOLOGY OF HUMAN PAPILLOMAVIRUSES

Papillomaviruses (from the Latin *papilla*: 'pustule' and the Greek suffix –*oma*: 'tumor') are members of the *papovaviridae* family which also include simian vacuolating virus and polyomavirus (Fields*, et al.*, 1990). Papillomaviruses have been found in a wide variety of animals including humans where they can infect basal epithelial cells of mucosal or cutaneous tissues (Burd, 2003). Human Papillomavirus (HPV) infection is one of the most common sexually transmitted diseases not only in the United Sates but worldwide. In the United States, the incidence of new HPV infections is probably 1 to 5.5 million people per year and the prevalence could be as high as 20 million people (Cates, 1999). There are more than 200 different types of HPVs where only 85 have been well characterized at the genomic level (Burd, 2003).

HPVs are small double-stranded circular DNA viruses of approximately 7,900 base pairs (bp).  HPV DNA interacts with cellular histone proteins to form a chromatin-like complex (Favre, 1975). HPV particles are non-enveloped with a diameter between 55-60 nm and are composed of 72 capsomers (Baker*, et al.*, 1991). Only one of the DNA strand of HPVs encodes proteins. The coding strand contains early and late genes depending upon their time of expression during the viral life cycle (Sanclemente*, et al.*, 2002). The HPV-16 genome can be divided into three different regions: First, the ~1,000 bp non-coding Long Control Region (LCR) that contains the core promoter (p97) of early genes, enhancer and

2

silencer sequences, origin of replication and DNA replication regulatory sequences; second, an ~4,000 bp early region that contains the open reading frames [ORFs] of E1, E2, E4, E5, E6, and E7 genes that are involved in viral replication, immortalization and proliferation of the host cell; and third, a ~3,000 bp region that contains the L1 and L2 ORFs that express the viral capsid proteins (Fig. 1) (Burd, 2003). The HPV early region includes the E6 and E7 genes that are involved in cellular transformation. The E6 and E7 genes have overlapping sequences that are differentially spliced (sharing the same 3' end) and use the same polyadenylation signal; (Seedorf, *et al.*, 1985).

The HPV life cycle is very unique in comparison to other virus families. HPV replication process requires epidermal or mucosal epithelial cells that still retain the ability to proliferate. Since epithelial cells undergo proliferation only near the basal membrane and then differentiate during migration to the suprabasal layers, infection of basal epithelium cells with HPV occurs only upon micro-abrasion of the epithelium (for example during sexual intercourse). The virus is known to bind to cell surface ligands (Sanclemente, *et al.*, 2002; zur Hausen, 2002). There is no evidence of specific HPV receptors, but some of the $\alpha_6$-integrin and heparan sulphate family molecules may be involved in this process (Evander, *et al.*, 1997;Giroglou, *et al.*, 2001). After viral-receptor recognition, HPV is decapsidated and the genome is transported inside the

# Fig. 1.    THE HUMAN PAPILLOMAVIRUS GENOME

## A



## B
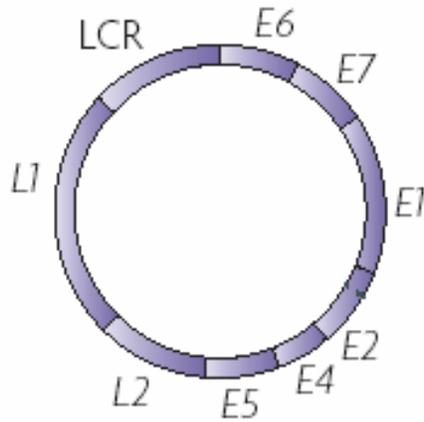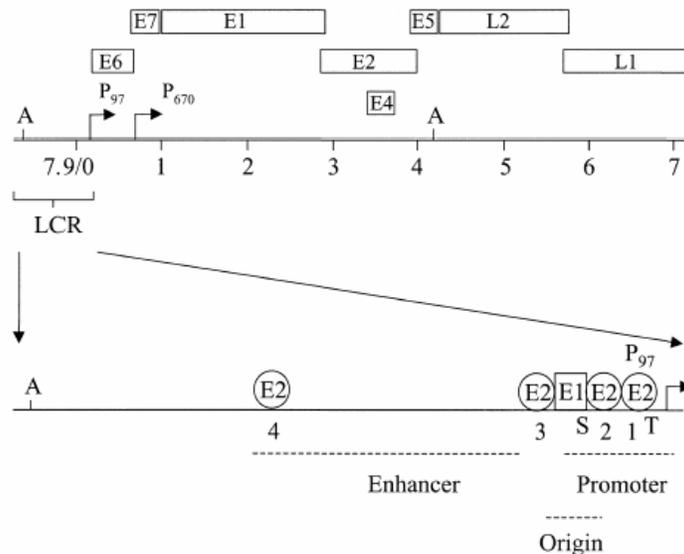
**(A) Circular HPV DNA showing the positions of the early expressed genes (E6, E7, E1, E2, E4 and E5), the late expressed gene (L1 and L2) and the long control region (LCR). (B) Top line shows the arrangement of the ORFs within the HPV genome. The bent arrows represent transcription start sites. Polyadenylation signals are indicated by an A. The bottom line shows an expanded view of the long control region (LCR) with the TATA box and Sp1-binding site indicated by T and S, respectively.**

nucleus (Duensing*, et al.*, 2002). Several studies have shown that the HPV life cycle is divided into non-productive and productive stages. The non-productive stage occurs in the basal layer of the epithelium. HPVs express early genes such as E1 and E2 in a regulated manner and maintain a low genomic copy number between 50 to 100 extrachromosomal molecules per cell (McMurray*, et al.*, 2001). When infected cells start their migration to the suprabasal layer during the process of differentiation, HPV initiates a productive stage of infection which manifests as an enhancement of genomic replication and the expression of late genes to produce viral particles (Chow*, et al.*, 1994). Some of the cellular transcription factors that regulate HPV gene expression in this process are the Activator Protein-1 (AP-1), Yin-Yang-1 (YY1), octamer binding factor family members, nuclear factor-1 (NF-1) and stimulatory protein-1 (Sp1) (Andersen*, et al.*, 1997;Saunders*, et al.*, 1998). The ability of HPV to produce virions without cell lysis allows it to produce more than one cycle of infection in the same epithelium (Fig. 2) (Sanclemente*, et al.*, 2002).

## 1.2 HUMAN PAPILLOMAVIRUSES AND CERVICAL CANCER

Cervical cancer is the second most common cause of death in women after breast cancer world-wide (Jin*, et al.*, 1999). Cervical carcinogenesis involves two main types of epithelial transformation. The transformation of basal cells in the squamous epithelium starts with a noninvasive low-grade squamous intraepithelial lesion (LG-SIL), leading to a high-grade squamous intraepithelial

# Fig. 2.     THE HUMAN PAPILLOMAVIRUS LIFE CYCLE
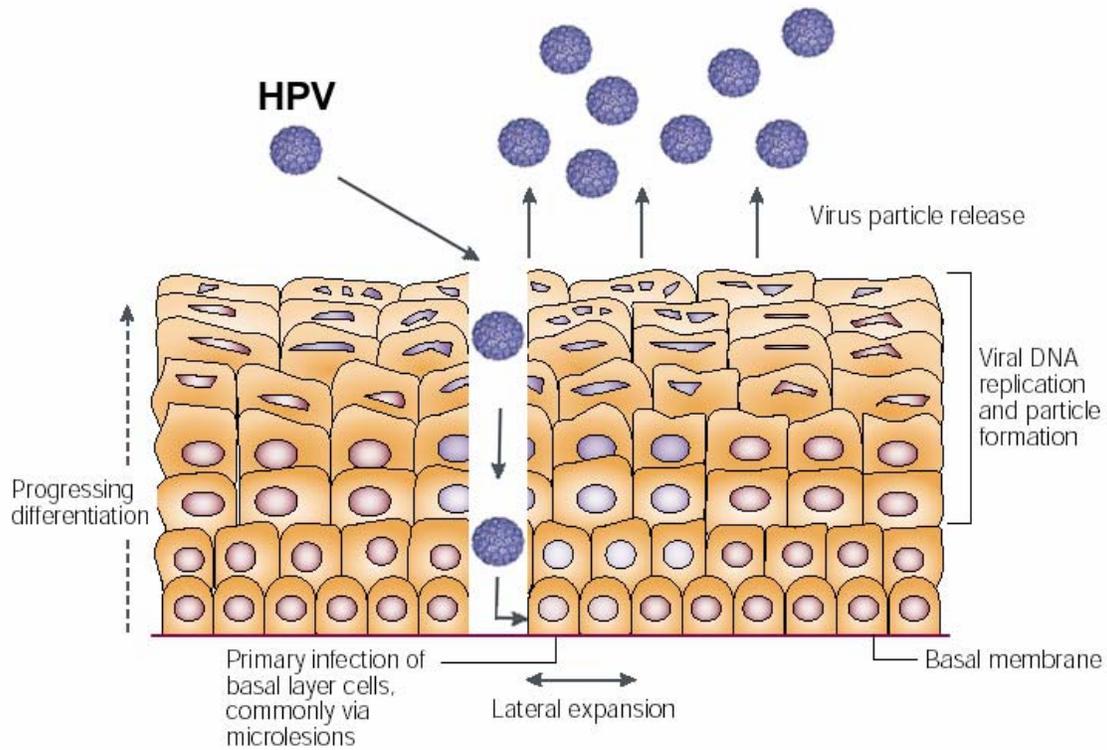


Modified from © Zur Hausen et. al. 2002 Nature Reviews Cancer

**Infection requires the availability of a basal layer cell. This usually occurs in microlesions of skin or mucosa. The infected cell divides and the population spreads laterally. Some of the progeny migrate into the suprabasal differentiating cell layers, where viral genes are activated, viral DNA is replicated and capsid proteins are formed. Viral particle formation ensues. Particles are released at the surface and might then infect additional tissues.**

lesion (HG-SIL), and ultimately evolving into invasive squamous cell carcinoma (SCC) (Alazawi, *et al.*, 2002). Adenocarcinomas and adenosquamous carcinomas (AC) originate from glandular endocervical columnar cells (Burk, *et al.*, 2003). Molecular and clinical evidence suggest a tight link between some types of HPV infection and cervical neoplasias. HPVs have been classified into two groups depending on their association with cervical cancer or precursor lesions. Low-risk HPVs (6, 11, 42, 43, and 44) are often found in cervical intraepithelial neoplasias grade 1 (CIN1), or CIN2 (also known as LG-SIL). High-risk HPVs (16, 18, 31, 33, 34, 35, 39, 45, 51, 52, 56, 58, 59, 66, 68, and 70) are often found in CIN3 (HG-SIL) or cervical cancer (Burd, 2003). The most common type of HPV found in cervical cancer in the United States and Europe is type 16 (50%), followed by 18, 31 and 45 (30%) (Walboomers, *et al.*, 1999). High-risk HPV DNA has been found in 99.7% of cervical squamous cell cancers world-wide (Walboomers, *et al.*, 1999). Therefore, the association between HPV infection and human cervical cancer is provides an important model to understand the interaction between human carcinogenesis and etiological agents.

Three viral genes of HPVs, E5, E6 and E7 are involved in cellular transformation. The HPV-16 E5 gene (located downstream of the E2 ORF)

encodes a 84 amino acid hydrophobic membrane protein that can dimerize (Kell, *et al.*, 1994). This protein is localized in the endoplasmic reticulum and golgi, as well as in the cytoplasmic membrane (Burkhardt, *et al.*, 1989;Conrad, *et al.*, 1993;Sparkowski, *et al.*, 1995). The main function of E5 is the stimulation of cell growth via the up-regulation in the activity of growth factor receptors (McMurray, *et al.*, 2001). Recent studies have shown that E5 inhibits the induction of apoptosis after DNA damage (Zhang, *et al.*, 2002). E5 expression may be important only in the early steps of transformation since this gene is normally disrupted during HPV integration (zur Hausen, 2002). The HPV-16 E6 gene encodes a 151 amino acid protein with two zinc finger domains that is expressed at the earliest stages of HPV infection (McMurray, *et al.*, 2001). The main function of E6 protein is to modify the cell growth-regulatory pathways creating a cellular environment that facilitates viral replication in cells that are already differentiated (Syrjanen, *et al.*, 1999). Some of the cellular pathways affected by E6 involve its binding and degradation of the tumor suppressor gene p53 via ubiquitin-protein ligase (Thomas, *et al.*, 1999), the up-regulation of telomerase activity (Klingelhutz, *et al.*, 1996), and the alteration of transcription regulation by binding the transcriptional activator p300/CBP (Patel, *et al.*, 1999). There is also evidence that E6 protein can regulate the transcription of minimal promoters that contain a TATA box (Sedman, *et al.*, 1991). The HPV-16 E7 gene encodes a 98 amino acid protein that contains casein kinase II phosphorylation sites at serine residues 31 and 32 (McMurray, *et al.*, 2001). Similar to E6, E7 also alters cellular pathways by several means: binding to the retinoblastoma tumor suppressor

protein, pRB, and destabilizing and degrading this cell cycle regulatory protein (Dyson*, et al.*, 1989), binding to histone deacetylase 1 (HDAC-1) to repress transcription (Brehm*, et al.*, 1999), binding to AP-1 transcription factors such as c-Jun (Antinore*, et al.*, 1996), direct interaction with the TATA-binding protein (TBP) and the TBP-associated factor-110 to modify transcriptional regulation (TAF-110) (Mazzarelli*, et al.*, 1995), and reactivation of DNA synthesis by creating a physical complex with cyclin E-cdk2 (McIntyre*, et al.*, 1996).

The precise molecular mechanisms involved in HPV-dependent cellular transformation are not fully undestood. An important molecular event involves the relatively random integration of the HPV DNA into the host genome in cervical tumors (Durst*, et al.*, 1985). This integration event is not a normal part of the viral life cycle but is an apparently accidental event that promotes cell transformation (Munger*, et al.*, 2002). Upon integration of the HPV DNA into the genome, the E1 and E2 genes are usually disrupted, leading to the loss of expression of the E1 and E2 proteins (Fields*, et al.*, 1990). The expression of E2 is required not only for HPV DNA replication, but also to repress transcription of the E6 and E7 oncogenes intiating from the p97 promoter (zur Hausen, 1991). Moreover, recent studies have shown that even the presence of HPV-16 in an episomal state can create chromosomal instability in an organotypic raft culture model (Duensing*, et al.*, 2001). This early event during cell proliferation by extrachromosomal HPV-16 probably occurs when the basal expression of E6 and E7 is sufficient to deregulate the mitotic machinery and induce the production of abnormal

9

centrosome numbers (Duensing, *et al.*, 2002). Chromosomal abnormalities such as aneuploidy or tetraploidy resulting from a reduction in mitotic fidelity are present in cervical cell lines and cervical cancers (Southern, *et al.*, 2001). It has been proposed that this type of genomic instability can promote integration of HPV genome (Kessis, *et al.*, 1996). HPV-induced genomic instability and viral integration may mimic DNA damage and may lead to the activation of various excision-repair and homologous recombination processes.

## 1.3    HUMAN PAPILLOMAVIRUSES AND ORAL CANCER

Every year ~40,000 new cases of Squamous Cell Carcinoma of the Head and Neck (SCCHN) are diagnosed in the United States (Furniss, *et al.*, 2007). Estimates for 2006 showed 40,500 new cases with 11,170 deaths (27.6%) (American Cancer Society, Cancer Facts and Figures 2006, Atlanta; American Cancer Society; 2006). The two main risk factors for head and neck cancer development are tobacco and alcohol, although some other factors such as radiation exposure, poor oral hygiene and unbalanced nutrient diets have also been associated with such cancers (Dobrossy, 2005;Portugal, *et al.*, 1997). Although the first suggestion that HPV may play a role in the development of oral cancer was proposed by Lüning and collaborators in 1985 (Loning, *et al.*, 1985), only recently a number of molecular and epidemiological studies have provided a strong correlation between high-risk HPV infection of the oral mucosa (specifically the oropharyngeal region) and the development of SCCHN (Gillison,

*et al.*, 2000;Schwartz*, et al.*, 1998;Steinberg*, et al.*, 1996). The International Agency for Research into cancer (IARC) in 2003 determined that the infection of HPV is highly likely to play an etiologic role in SCCHN (Herrero*, et al.*, 2003). Interestingly, several studies showed that there are clear differences in HPV positivity between regions of the oral cavity. The association of HPV infection is notably strong in the tonsil mucosa (Dahlstrand*, et al.*, 2004; Klussmann*, et al.*, 2003;Venuti*, et al.*, 2004). One cause for this predisposition may be a favorable environment for HPV replication in the invaginated folds within the tonsillar crypts (Tran*, et al.*, 2007). Another difference between HPV-associated SCCHN and other oral cancers is a relatively better prognosis and risk reduction from death of approximately 50% in HPV-positive tumors (Ritchie*, et al.*, 2003). One possible explanation for this is the fact that the tumor suppresor gene p53 (frequently mutated in HPV-negative SCCHN) is generally not mutated in HPV-associated SCCHN.  Since the function of p53 is not totally inactivated by the HPV E6 oncogene,the response of such tumors to radiotherapy is better since p53 may help in cellular apoptosis (Kumar*, et al.*, 2003). Also, there is a background prevalence of HPV DNA in the oral mucosa of otherwise healthy individuals without any evidence of cancer (Gillison*, et al.*, 2001).

The notion that HPV-associated SCCHN is completely analogous to cervical carcinoma may not be very accurate, and more information on the molecular carcinogenesis and gene expression profiles may yield clues into the similarities and differences between these two diseases. For example unlike

cervical cancer, SCCHN is primarily associated with a single oncogenic subtype HPV-16 (Gillison*, et al.*, 2000;Schwartz*, et al.*, 1998). Also, the apparent time interval between infection and development of carcinoma appears to differ between SCCHN and cervical carcinoma, and epidemiologic analyses have indicated an older age of patients comprising the SCCHN subgroup versus cervical carcinoma patients (Gillison*, et al.*, 2000;Schwartz*, et al.*, 1998). Finally, the prognosis for head and neck cancer has not changed greatly in the last two decades, so a better understanding of the molecular differences between HPV-positive and HPV-negative SCCHN will be useful for future treatment procedures such as HPV vaccination or closer follow-up of specific therapies.

## 1.4    MICROARRAY TECHNOLOGY

One of the most recent technologies used to understand complex global mRNA expression profiles is the microarray or "gene chips" (Schena*, et al.*, 1996; Zhu*, et al.*, 1998). Despite being a very recent technology, it has quickly become a powerful tool for research, including disease characterization, biological classification and pharmacogenomics. A microarray is defined as a collection of microscopic spots of specific nucleic acid sequences (typically referred to as probes) distributed in a grid-like format that is attached to a solid surface (Butte, 2002;Conway*, et al.*, 2003). The sample or samples that are hybridized to these probes (typically referred to as targets) are normally labeled with some types of fluorescent dyes, which emit signals that can be quantified at different wave

lengths using scanners (Li, *et al.*, 2002;Nakanishi, *et al.*, 2001). There are two main types of microarray platforms: spotted microarrays (low density) and oligonucleotide microarrays (high density) (Brown, *et al.*, 1999). In the spotted microarrays, the probes that are spotted on the surface of the slide or membrane can be single- or double-stranded oligonucleotides, complementary DNA (cDNA) or PCR fragments. In contrast, the high density oligonucleotide microarrays normally use probes that are synthezised *in situ* by different chemical reaction like photolithography (Butte, 2002;Li, *et al.*, 2002). The differences in spot density between these two types of microarrays are shown in Fig. 3. At the same time, there are several types of labeled targets that can be used for the hybridization of the probes, including RNA, genomic DNA, cDNA, complementary RNA (cRNA) or PCR products (Slonim, 2002;Xiang, *et al.*, 2003). In all the cases, the goal is to hybridize the target sample(s) specifically to the complementary sequence of different probes present in the microarray. After comparing the differences in fluorescent intensities between the samples and the respective controls, the microarray experiments reveal differential expression of genes between the samples (overexpression or underexpression) (Li, *et al.*, 2002;Nakanishi, *et al.*, 2001).

Before identifying differential expression between several samples, normalization procedures are utilized in order to normalize the data from different samples and reduce diffrerences in hybridization efficency, dye incorporation, degradation of the sample, etc. There are different types of

**Fig. 3.    DIFFERENCES BETWEEN LOW AND HIGH DENSITY MICROARRAYS**
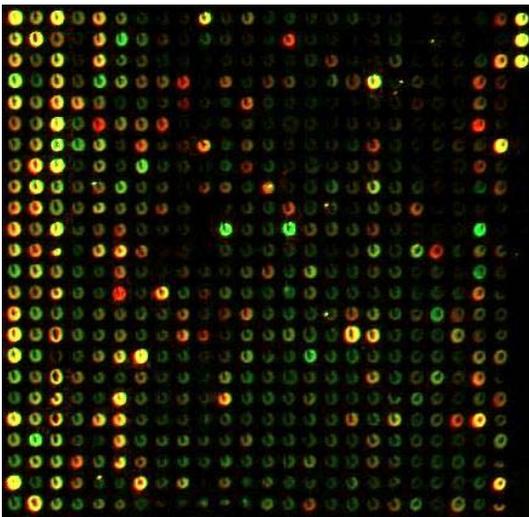


**Low density microarray**
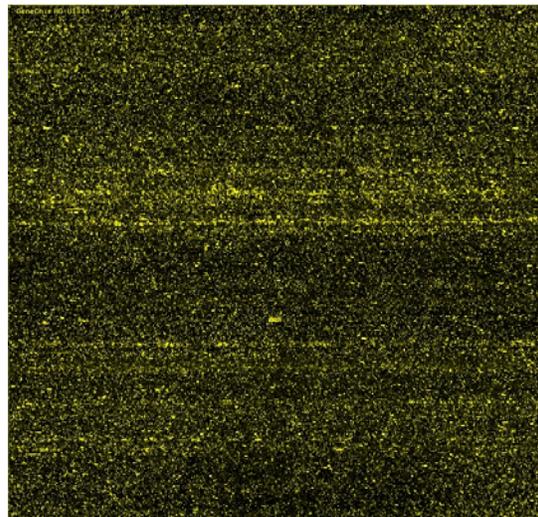


**High density microarray**

normalization strategies. The most commonly used are: total intensity normalization (assumes that all the samples have the same amount of RNA, therefore, the total intensity calculated from all the spots on an array should be the same), regression normalization (assumes that a significant number of genes are expressed similarly between the samples, so the data is modified to be equal to one) and ratio statistics (assumes that house-keeping genes have the same expression level in different samples, so the data is adjusted to such genes after obtaining the mean expression ratios) (Datta*, et al.*, 2003;Dopazo*, et al.*, 2001;Eisen*, et al.*, 1998).

After normalization, the data can be manipulated using statistical analyses such as clustering algorithms (to obtain cluster of genes that are expressed in a similar fashion) or using differential expression algorithms (to obtain genes that are expressed at different levels with some degree of statistical significance) (Saeed*, et al.*, 2003). Finally, after statistical analyses of the microarray data, the most common techniques to validate this information (to avoid false positive results) are Northern blotting and semi-quantitative or quantitative real-time RT-PCR. The genes selected for such verification may include those that are highly overexpressed or underexpressed in the microarray analysis, genes with a level of expression that is known from previous studies, or genes selected for a specific purpose of the study.

## 1.5    MicroRNAs AND CANCER

The recent discovery of small non-coding RNAs named microRNAs (miRNAs) that have the ability to regulate different cellular processes in several organisms including humans, is probably one of the most important discoveries in biology. This "second RNA revolution" (after the discovery of enzymatic activity of some RNAs in the 1980s) (Zaug, *et al.*, 1986) could potentially open a new layer of complexity in the regulation of gene expression in eukaryotic cells. The first miRNA discovered was *lin-4* in *C. elegans* in the year 1993 by Lee, Feinbaum and Ambros (Lee, *et al.*, 1993). Most of the scientific community ignored the relevance of this discovery for several years (predicting that this type of genetic regulation was exclusive to *C. elegans*) until the second miRNA named *let-7* was discovered in 2000, and soon afterwards it was found that this miRNA was conserved across species from *D. melanogaster* to humans (Pasquinelli, *et al.*, 2000). Recently, there has been an explosion in miRNA research. The number of human miRNAs reported in the miRbase database ([http://microrna.sanger.ac.uk/sequences/](http://microrna.sanger.ac.uk/sequences/)) up to March 2007 has reached 482, and an additional 1,000 miRNAs have been predicted based on computational analysis.

The processing and maturation of miRNAs occurs in different cellular compartments. Most miRNAs are transcribed by RNA polymerase II, have a cap, and are polyadenylated (Kim, 2005). Nevertheless, there is still the possibility

16

that some miRNAs may be transcribed by RNA polymerase III. Following transcription, the large primary (pri)-miRNA transcripts (several kilobases) are processed into precursor (pre)-miRNAs by the protein Drosha in the nucleus (Lewis, *et al.*, 2005). Pre-miRNAs are hairpin-like structures of ~80 nt with characteristic 2 nt 3' overhangs. In animals, pre-miRNAs are exported to the cytoplasm by exportin 5 (Khvorova, *et al.*, 2003;Kim, 2005), with GTP-binding co-factor Ran, where they are further processed to generate mature miRNA duplexes by the protein Dicer. MiRNA duplexes (~ 22 nt) associate with the RNA-induced silencing complex  (RISC) but only one strand, the mature miRNA (normally the strand with less thermodynamic stability in the 5' end), remains bound to RISC and is delivered to its target (Khvorova, *et al.*, 2003;Kim, 2005;Schwarz, *et al.*, 2003). The fate of target mRNAs depends on the degree of complementarity with the miRNA. Commonly, the 5' 2-8 nt of the miRNA (called the seed sequence) is complementary to the target, and the remaining miRNA contains many mismatches. A low degree of complementarity results in translational repression, while a high degree of complementarity results in cleavage of the mRNA followed by its eventual destruction (Fig. 4) (Kim, 2005).


The location of miRNAs in the genome is variable but around 70% of known mammalian miRNAs are located in three different types of transcription units (Kim, 2005;Rodriguez, *et al.*, 2004): in the exonic regions of a non-coding transcript, in the intronic region of a non-coding transcript, and in the intronic

# Fig. 4.    THE BIOGENESIS OF MicroRNAs
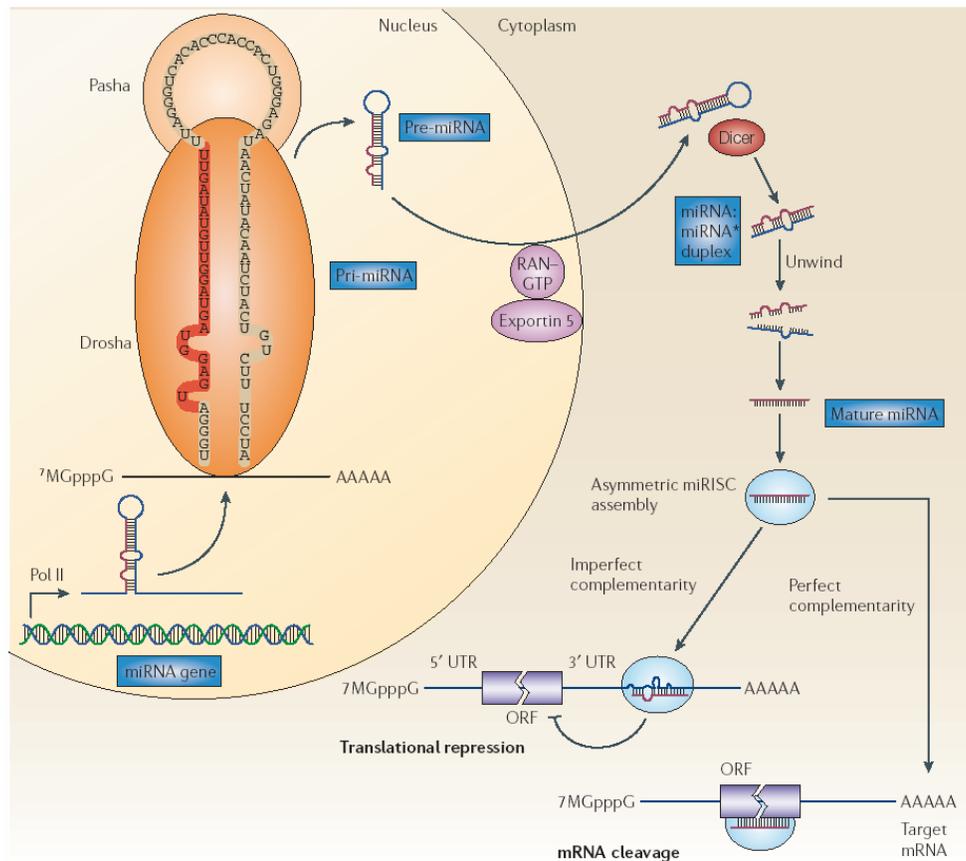


© Esquela-Kercher A. and Slack FJ. 2006. Nature Reviews Cancer

**MicroRNA (miRNA) genes are generally transcribed by RNA Polymerase II (Pol II) in the nucleus to form large pri-miRNA transcripts. These pri-miRNA transcripts are processed by the RNase III enzyme Drosha to release the ~70-nucleotide pre-miRNA precursor product. RAN–GTP and exportin 5 transport the pre-miRNA into the cytoplasm. Subsequently, another RNase III enzyme, Dicer, processes the pre-miRNA to generate a transient ~22- nucleotide miRNA:miRNA\* duplex. This duplex is then loaded into the miRNA-associated multiprotein RNA-induced silencing complex (miRISC) (light blue), and the mature single-stranded miRNA (red) is preferentially retained in this complex. The mature miRNA then binds to complementary sites in the mRNA target to negatively regulate gene expression.**

region of a protein coding transcript. Additionally, there are a small number of miRNAs that overlap with the exons or introns of alterantively spliced genes. Also, other miRNAs are occasionally found in the 3' UTR of coding mRNAs. One interesting characteristic of mammalian miRNAs is that approximately 50% of them are clustered with other miRNAs, suggesting that these groups of miRNAs are transcribed as a single polycistronic transcription unit.

The first demonstration of a possible link between miRNAs and human cancer was provided by Croche and collaborators in 2002 when they discovered a frequent down-regulation of miR-15 and miR-16 in chronic lymphocytic leukemias (Calin*, et al.*, 2002). MiRNAs have also been proposed to function as oncogenes or tumor suppressor genes. For example, reduced expression of miRNAs that target oncogenes will result in tumor suppression. Conversely, overexpression of miRNAs that target tumor suppressors will result in oncogenic activity and tumor formation. Recently, some miRNAs such as miR-15a and miR-16-1 have been shown to negatively regulate the *Bcl-2* oncogene that is commonly overexpressed in many cancers (Cimmino*, et al.*, 2005). The *let-7* family of miRNAs regulates the *ras* oncogenes (Johnson*, et al.*, 2005). *Ras* genes contain activating mutations in about 15-30% of cancers. The miRNAs controlling the expression of this oncogene therefore serve as important regulators of cellular proliferation. In some cancers, downregulation of the *let-7* family miRNAs results in reciprocal upregulation of *Ras*. This downregulation is

most pronounced in lung cancers (Johnson, *et al.*, 2005), but is also sporadically observed in other tissues (Johnson, *et al.*, 2005). Oncogenes can also regulate the expression of miRNAs. C-myc regulates the miRNA cluster miR-17-92 that in turn targets E2F1 (O'Donnell, *et al.*, 2005). Upregulation of this cluster was demonstrated for B cell lymphomas. C-myc is often upregulated in cancers, and it serves as a transcription factor for the miR-17-92 cluster (O'Donnell, *et al.*, 2005). Because C-myc can also directly regulate E2F1, C-myc-mediated repression of E2F1 via the miR-17-92 cluster serves to fine-tune E2F1 regulation (O'Donnell, *et al.*, 2005).

Another model to study the influence of miRNAs in cancer has been provided by transgenic and knock-out mouse models. Recently, a transgenic mouse that specifically overexpresses miR-155 in B cells showed the development of polyclonal pre-leukaemic pre-B-cell proliferation followed by B-cell malignancy (Costinean, *et al.*, 2006). This model suggests that the deregulation of some miRNAs in cancer could be one of the first steps in the process of carcinogenesis. Finally, miRNA microarray profiling has become an important tool in an effort to understand the importance of miRNA expression patterns and their implication in human cancers. The first study of global miRNA expression in human and mouse tissue using miRNA microarrays was published by Liu and collaborators in 2004 (Liu, *et al.*, 2004).

## 1.6    THESIS OUTLINE

**Chapter 1** consists of a general introduction to the biology and viral life cycle of HPVs followed by the correlation between infection by these viruses and the pathology of cervical and head and neck carcinomas. This chapter also describes the evolution of the microarray technology specifically used for global gene expression analysis. Finally, we describe the biology of microRNAs and its implication in human cancer.

**Chapter 2** describes the differential gene expression profiles of squamous cell carcinoma of the head and neck (SCCHN) tumors containing HPV DNA compared to HPV-negative SCCHN tumors or normal head and neck tissues by using high-density microarrays. After the verification of the microarray data by other types of analyses, we found specific gene expression changes and gene clustering in HPV-positive SCCHN supporting the idea that this type of head and neck malignancy may contain alterations in different pathways compared to HPV-negative SCCHN.

**Chapter 3** describes the differential gene expression profiles of HPV-positive cervical cell lines in comparison to an HPV-negative cell line (C-33A) and normal cervical tissue. After the detection of viral oncogene expression and physical state of the HPV DNA (episomal or integrated) in the HPV-positive cell lines, we

analyzed the global expression of these cell lines by using high-density microarrays. After the verification of the microarray data with other molecular tools, we found specific gene expression changes in HPV-positive cervical cell lines supporting the idea that the presence of HPV DNA is enough to change the expression landscape of the host cells.

**Chapter 4** describes the differential expression of microRNAs (miRNAs) in cervical carcinoma cell lines and cervical carcinomas containing HPV-16 DNA in comparison to a HPV-negative cervical carcinoma cell line or normal cervical tissue. After the verification of the microarray data with other molecular tools, we found that one miRNA named miR-218 was specifically downregulated by the HPV-16 E6 oncogene. Finally, we found that miR-218 possibly targets the mRNA of the Laminin 5 β3 gene (LAMB3) which as been implicated in cervical cancer and HPV infection.

**Chapter 5** discusses the results of the research described in previous chapters and describes the future goals of these different projects.

**Chapter 2**


**IDENTIFICATION OF DIFFERENTIALLY EXPRESSED GENES IN HPV-POSITIVE AND HPV-NEGATIVE OROPHARYNGEAL SQUAMOUS CELL CARCINOMAS**


Work described in this section was published in the European Journal of Cancer (Eur J Cancer. 2007, 2:415-432) with authors Ivan Martinez, Jun Wang, Kenosha F. Hobson, Robert L. Ferris and Saleem A. Khan

R.L. Ferris and J. Wang provided tissue samples and performed RNA extraction. K.F. Hobson performed part of data analysis. I. Martinez performed all the experiments described in this section. I. Martinez, R.L. Ferris and S.A. Khan wrote the manuscript.

**2.1 INTRODUCTION**

There are approximately 500,000 new cases of squamous cell carcinoma of the head and neck (SCCHN) worldwide and 45,000 cases in the United States per year (Greenlee*, et al.*, 2001;Vikram*, et al.*, 1984). SCCHN is usually associated with such risk factors as heavy consumption of alcohol and/or tobacco. The survival in SCCHN patients is still poor and has not improved recently despite the advances in detection and therapies. It is well-established that human papillomaviruses (HPVs) are involved in the pathogenesis of cervical cancer (Dell*, et al.*, 2001;Walboomers*, et al.*, 1999;zur Hausen, 2000;zur Hausen, 2002). Recently, molecular epidemiologic studies have shown a strong correlation between oncogenic HPV infections and a subset of oropharyngeal cancers (Gillison*, et al.*, 2000;Ritchie*, et al.*, 2003;Schwartz*, et al.*, 1998;Steinberg*, et al.*, 1996;Wong*, et al.*, 2000). It is not currently understood whether these HPV infections are an independent etiologic factor or a co-factor in the development of such tumours. Interestingly, a relative risk reduction from death of approximately 50% has been observed in HPV-associated tumours compared to those without detectable HPV DNA (Ritchie*, et al.*, 2003).

HPVs are small double-stranded DNA viruses of approximately 7900 base pairs (bp). At present, more than 150 different types of HPV have been identified, and the high risk HPV types 16 and 18 are associated with a majority of cases of

cervical cancer (zur Hausen, 2000;zur Hausen, 2002). Replication of HPVs requires the viral E1 and E2 proteins as well as host replication factors (Chow, *et al.*, 1994;Fuchs, *et al.*, 1994). The E2 protein also downregulates the transcription of the viral E6 and E7 oncogenes, which are transcribed from the p97 promoter (Mansur, *et al.*, 1993;McBride, *et al.*, 1991;McCance, 2005). The E6 and E7 oncoproteins of high-risk HPVs are involved in cellular transformation (Mansur, *et al.*, 1993;McBride, *et al.*, 1991;McCance, 2005;Munger, *et al.*, 1993;Stoppler, *et al.*, 1994). Most benign and low-grade cervical lesions contain HPV DNA in an extrachromosomal state (Durst, *et al.*, 1985). However, in most cases of cervical carcinomas the HPV DNA is usually found integrated into the host chromosomes, frequently disrupting the E1 and E2 genes (Durst, *et al.*, 1985;Meissner, 1999;zur Hausen, 2000;zur Hausen, 2002). This results in increased expression of the viral E6 and E7 oncogenes (Yee, *et al.*, 1985). The E6 protein promotes ubiquitination and consequently proteasomal degradation of the cellular tumour suppressor proteins p53 and PDZ domain-containing disc large protein (DLG) (Dell, *et al.*, 2001;Lechner, *et al.*, 1992;Mantovani, *et al.*, 1999;Scheffner, *et al.*, 1990;zur Hausen, 2002). E6 is also known to interact with a number of other cellular proteins and activates telomerase (zur Hausen, 2000;zur Hausen, 2002). The E7 protein binds to and inactivates the function of pRB and related p107 and p130 proteins (Dell, *et al.*, 2001;zur Hausen, 2002). E7 also interacts with additional cellular proteins such as TBP, histone H1 kinase, cyclin E, etc (Fuchs, *et al.*, 1994;Mansur, *et al.*, 1993;McCance, 2005;Munger, *et al.*, 1993;Stoppler, *et al.*, 1994). The E6 and E7 proteins are

also known to alter cellular gene expression but the precise molecular mechanisms involved are not well-understood (Mansur*, et al.*, 1993;Munger*, et al.*, 1993;Stoppler*, et al.*, 1994). In addition, E6/E7 expression promotes chromosomal instability, foreign DNA integration and other mutagenic events in the cell (Dell*, et al.*, 2001;Jeon*, et al.*, 1995;McCance, 2005;zur Hausen, 2000). Although several studies have suggested a possible role for HPV infection in a subset of SCCHN, very little is known about the molecular events involved in carcinogenesis. Although the distribution of episomal and integrated HPV forms in precancerous and cancerous lesions of the head and neck has not been determined, limited evidence suggests a similar physical state as observed in cervical carcinoma. Molecular studies of HPV-associated SCCHN are necessary for a better understanding of the physical state and potential role of this virus in carcinogenesis, and for the development of new, more targeted therapeutic strategies.

Recently, DNA microarrays have been successfully used to identify global patterns of gene expression in different human neoplasias, including head and neck cancers (Golub*, et al.*, 1999;Jeon*, et al.*, 2004;van 't Veer*, et al.*, 2002;van de Vijver*, et al.*, 2002). In the case of breast cancer studies, a set of 70 genes correctly predicted the nature and progression of the disease as well as the outcome (van 't Veer*, et al.*, 2002;van de Vijver*, et al.*, 2002). These studies support the idea that changes in the molecular profiles of gene expression are early events during carcinogenesis and such global expression profiles can be

used effectively to predict the course of the disease. Recently, investigators have used microarrays to analyze gene expression changes in SCCHN tissues and cell lines, but little is known about the gene expression changes in HPV-associated SCCHN (Gillison*, et al.*, 2001;Guo*, et al.*, 2002;Ibrahim*, et al.*, 2003;Kuo*, et al.*, 2002;Leethanakul*, et al.*, 2003). The identification of molecular portrait of gene expression profiles in HPV-positive and HPV-negative SCCHN, including their differences, could result in a better understanding of critical events during carcinogenesis.

This study was undertaken to identify changes in cellular gene expression profiles in HPV-positive and HPV-negative SCCHN as compared to normal tissues as well as to each other. Our microarray analysis showed considerable differences in the gene expression profiles that were specifically associated with these two types of cancers. Several of the differentially expressed genes were found to be involved in cell cycle regulation (*CDKN2A*), nuclear structure and meiosis (*SYCP2*), DNA replication and repair (*RFC5*), transcription regulation (*ZNF238*), cell differentiation (*KLK8*) and epidermis development (*CRABP2*). Unsupervised clustering analysis also showed that genes located in specific chromosomal regions such as 1p31-p36 and 12q21-24 were specifically overexpressed in HPV-positive oropharyngeal squamous carcinomas.

## 2.2    MATERIAL AND METHODS

*Tissue samples and cell lines*

A total of 11 samples were analysed in this study: Three HPV-positive (SK20, SK30, SK31), four HPV-negative (SK32, SK33, SK34, SK35), and four normal oral mucosa (SK16, SK17, SK36, SK37).  Tumours and normal mucosal specimens were snap-frozen and stored at $-80^0$C until further use. Collection of tissues was performed under an IRB-approved Tissue Banking protocol, and written informed consent was obtained from each patient prior to sample collection.  The SCCHN samples were from the oropharynx (tonsil and base of tongue), while the normal oral mucosa specimens were obtained from patients undergoing removal of the oropharyngeal tissues: tonsils, soft palate and uvula for sleep apnea. None of the patients received prior chemotherapy or radiotherapy. Cervical carcinoma cell lines were obtained from the American Type Culture Collection (ATCC) and maintained in Dulbecco's modified Eagle medium (C-33A) or RPMI 1640 (CaSki) supplemented with 10% fetal bovine serum (FBS) at 37°C in the presence of 5% $CO_2$.

*Isolation of RNA and RT-PCR analysis*

Total RNA from all frozen tumour samples and oropharyngeal mucosal tissues was isolated using the RNeasy Mini Kit (Qiagen Inc.) according to the

manufacturer's protocol. The mucosa was carefully dissected from any subjacent muscle or lymphoid tissue before RNA extraction, so that only normal squamous epithelium was studied. The RNA pellet was dried under a vacuum and resuspended in 30-50 μl of DEPC treated water. The integrity of RNA samples was confirmed by analysing 1 μg of total RNA on 1.2% (w/v) denaturing formaldehyde-agarose gels. Before DNA synthesis, RNA was treated with deoxyribonuclease I, Amplification Grade (Invitrogen) for 15 min at room temperature to avoid DNA contamination. DNase I was inactivated by incubation with 25 mM EDTA at 65ºC for 10 min. We examined the expression of the HPV-16 E6 and E7 genes in the above samples by RT-PCR (Ferris, *et al.*, 2005) using the following primers: HPV-16 E6 (forward) 5'- ATGCACCAAAAGAGAACTGC -3' and (reverse) 5'- TTACAGCTGGGTTTCTCTAC -3' (477-bp PCR product); HPV-16 E7 (forward) 5'-GTAACCTTTTGTTGCAAGTGTGACT-3' and (reverse) 5'-GATTATGGTTTCTGAGAACAGATGG-3' (137-bp RT-PCR product). The expression of the cellular glyceraldehyde-3-phosphate dehydrogenase (*GAPDH*) house-keeping gene was used as a positive control using the following primers: *GAPDH* (forward) 5'- ACCACAGTCCATGCCATCAC-3' and (reverse) 5'-TCCACCACCCTGTTGCTGTA-3' (452-bp RT-PCR product). All RT-PCR reactions were done using the Thermoscript One-step System kit (Invitrogen Corp.). The cDNA synthesis was performed at 37°C for 1 hour in a final volume of 20 μl using 1 μg of total RNA template, 0.5 μg of oligo (dT)$_{15}$, 10 mM dNTPs, 30 U of RNase inhibitor and 200 units of MMLV reverse transcriptase. PCR was performed in a 50 μl volume containing 50 mM KCl, 10 mM Tris (pH 8.3), 1.5 mM

MgCl$_2$, 0.01 % gelatin (w/v), 200 $\mu$M deoxynucleoside triphosphate (dNTP) mix, 0.4 $\mu$M of each primer and 2.5 units of the Taq DNA polymerase. The DNA was denatured at 94°C for 5 min, followed by 40 PCR amplification cycles that consisted of denaturation (94°C, 1 min), annealing (58-60°C, 1 min) and extension (72°C, 2 min). An additional extension step of 72°C for 5 min was included at the end of the reaction.

*Quantitative real-time RT-PCR analysis*

We validated the microarray data by quantitative real-time RT-PCR (QRT-PCR) analysis of a few representative genes using the QuantiTect SYBR Green one-step RT-PCR kit (QIAGEN, Valencia, CA).  Approximately 400 ng of DNase I-treated total RNA from the samples was mixed with 25 $\mu$l of 2 X QuantiTect SYBR Green RT-PCR Master Mix, 0.5 $\mu$M of specific forward and reverse primers and RNase-free water in a final volume of 50 $\mu$l. The amount of fluorescence emitted by SYBR Green I was quantified by the ABI Prism 7700 system software (Applied Biosystems, Foster City, CA). The genes and primers used for this analysis are shown in Table 1. The following RT-PCR cycle parameters were used: reverse transcription at 50°C for 30 min, hot-start DNA polymerase activation 95°C for 15 min, 40 cycles of denaturation at 94°C for 15 sec each, annealing at 58°C for 30 sec and extension at 72°C for 30 sec. Each reaction was run in triplicate in a 96-well plate. Relative expression of the target

gene was calculated using the 2 delta CT method described previously (Livak, *et al.*, 2001): (Relative expression = $2^{-\Delta CT}$; where $\Delta CT = C_{T\,(Target\,gene)} - C_{T\,(endogenous\,control\,gene)}$) where GAPDH was used as the endogenous control gene.

*cRNA synthesis and microarray analysis*

Microarray analysis was carried out using the high-density oligonucleotide U133A GeneChip® expression microarray (Affymetrix, Santa Clara, CA) that contains 22,215 transcript sets representing 14,820 human genes. RNA isolated from the samples described above was further purified by using the RNeasy Total RNA Isolation Kit (QIAGEN, Valencia, CA) to remove contaminating DNA. Twenty micrograms of total RNA from each sample were used to generate double-stranded (ds) cDNA using the Superscript II Choice system (GIBCO-BRL, Rockville, MD). First-strand synthesis was carried out using a T7-(dT)$_{24}$ primer (Sigma-GenoSys, The Woodlands, TX) and Superscript II reverse transcriptase. This primer includes the promoter sequence for the T7 RNA polymerase. The second-strand synthesis was performed using RNase H and DNA polymerase I. The resulting ds cDNA was used to synthesize cRNA using biotin-labeled ribonucleotides and T7 RNA polymerase utilizing the BioArray HighYield RNA transcript labelling kit (ENZO Life Sciences, Inc. Farmingdale, NY). The cRNA was fragmented to 35-200 bases length and hybridized to the Human Genome U133A GeneChip® expression microarray in the Affymetrix Hybridization Oven (Affymetrix, Santa Clara, CA). The GeneChips were washed, and then stained

# TABLE 1.  PRIMER SEQUENCES FOR QUANTITATIVE RT-PCR

| Gene title | Forward primer 5'-3' | Reverse primer 5'-3' |
| --- | --- | --- |
| cellular retinoic acid binding protein 2 (CRABP2) | TGCAGGGTCTTGCTTTCTTT | GGGCTAGGACTGCTGACTTG |
| cyclin-dependent kinase inhibitor 2A (CDKN2A) | AGTGAGCACTCACGCCCTAA | CCCCTGAGCTTCCCTAGTTC |
| kallikrein 8 (KLK8) | ACCAGTCCCCGAGAGAATTT | ACAGACCATGCCATCTGTGA |
| replication factor C (activator 1) 5, 36.5kDa (RFC5) | GTCACCAGCAGGTTCCAAAT | TCCTGCATGAAAACAAGTGC |
| synaptonemal complex protein 2 (SYCP2) | CATGTCACCGAAGCAAGTGT | GAAGTCTTCTGGGCTTGGACT |
| zinc finger protein 238 (ZNF238) | TTTCCTTTGAGGGGATAGGG | TCAAAAGACAATGCAGTGTTGA |
| glyceraldehyde-3-phosphate dehydrogenase (G3PDH) | CAGCCTCAAGATCATCAGCA | TGTGGTCATGAGTCCTTCCA |

with streptavidin-phycoerythrin (SAPE) to generate fluorescent signals from biotin-labeled cRNA. The GeneChips were scanned using an Affymetrix GeneArray® scanner (Affymetrix, Santa Clara, CA). Each probed array was scanned twice to calculate an average of two images, define the probe cells and compute intensity for each cell. Fragmentation, hybridization, staining and scan were performed by the Pitt Array Facility, University of Pittsburgh.

*Microarray data analysis*

The .CEL image files obtained from the Affymetrix Microarray Suite 5.0 were converted to .DCP files using the program dChip version 1.2 (www.dchip.org). Two different statistical tests, Significance Analysis of Microarray (SAM) program version 1.21 (www.stat.stanford.edu/~tibs/SAM/) and Gene Expression Data Analysis (GEDA) program, were used to analyze the microarray data in order to identify differentially expressed genes that were robust enough to appear in both the analyses.  For SAM analysis, the dChip program was used to generate normalized intensity signal data. The invariant set normalization method used by dChip program is one of the most commonly used normalization method for analysing the data in the SAM program (Saviozzi*, et al.*, 2004). In SAM analysis, a false discovery rate (FDR) of less than 5 percent and a difference of at least 2-fold was chosen to identify the total number of differentially expressed genes. We also utilized the GEDA tool developed at the University of Pittsburgh (James Lyons-Weiler, University of Pittsburgh Cancer

Institute, http://bioinformatics.upmc.edu) for statistical analysis. For this, non-normalized intensity signal data from dChip (PM-only model) were normalized (log-base2 transformation, median within arrays, Global Mean Adjustment among arrays) and analysed with a t-Test at level alpha = 0.05. This normalization method is specifically used by the GEDA program. Genes with an expression ratio of at least 2-fold difference relative to the controls were considered to be differentially expressed in GEDA analysis. After analyzing the data with the two statistical tests, we overlapped the lists of differentially expressed genes in order to generate a list of differentially expressed genes using software developed by Dr. Jim Lund at the University of Kentucky (http://elegans.uky.edu/MA/progs/Compare.html). All the microarray experiments were designed in accordance with the MIAME guidelines (Brazma, *et al.*, 2001).

For unsupervised hierarchical clustering analysis, we first filtered the genes using a threshold of 0.05 in the ratio of the standard deviation and the mean of a gene's expression values across all samples (Coefficient of Variation), and then applied an algorithm for similarity measurements using Pearson's correlation coefficient (Li, *et al.*, 2001).

## 2.3 RESULTS

*Expression of the HPV-16 early genes in SCCHN samples by RT-PCR analysis*

RT-PCR analysis was carried out to test for the presence and expression of HPV early genes in all the SCCHN samples and the control normal oral tissues. RNA from CasKi and C-33A cervical carcinoma cell lines was used as positive and negative control, respectively, for HPV-16 gene expression. The E6 (with different splicing variants) and E7 viral oncogene transcripts were present in the HPV-16 positive SCCHN samples SK20, SK30 and SK31 but not in the HPV-negative SCCHN samples SK32, SK33, SK34, and SK35 or the normal oral mucosa samples SK16, SK17, SK36 and SK37 (Fig. 5). The above results confirmed the appropriate expression of the HPV-16 oncogenes in the three HPV-positive SCCHN samples.

*Microarray analysis of genes differentially expressed in the HPV-positive SCCHN samples compared to the normal oral tissue*

In order to identify differentially expressed genes in HPV-16 positive SCCHN samples (SK20, SK30 and SK31) as compared to the normal oral tissue samples (SK16, SK17, SK36 and SK37), we utilized the Affymetrix U133A GeneChip. Normalization and differential gene expression analyses were carried out as described in material and methods. A total of 1,329 and 1,431 genes showed up- and down-regulation, respectively, using a pooled variance t-Test analysis when all the HPV-16 positive SCCHN samples were compared to the normal oral tissues (supplementary data, Table A). On the other hand, 470 and

**Fig. 5.      ANALYSIS OF HPV16 E6 AND E7 GENE EXPRESSION IN OROPHARYNGEAL TISSUE SAMPLES BY RT-PCR**
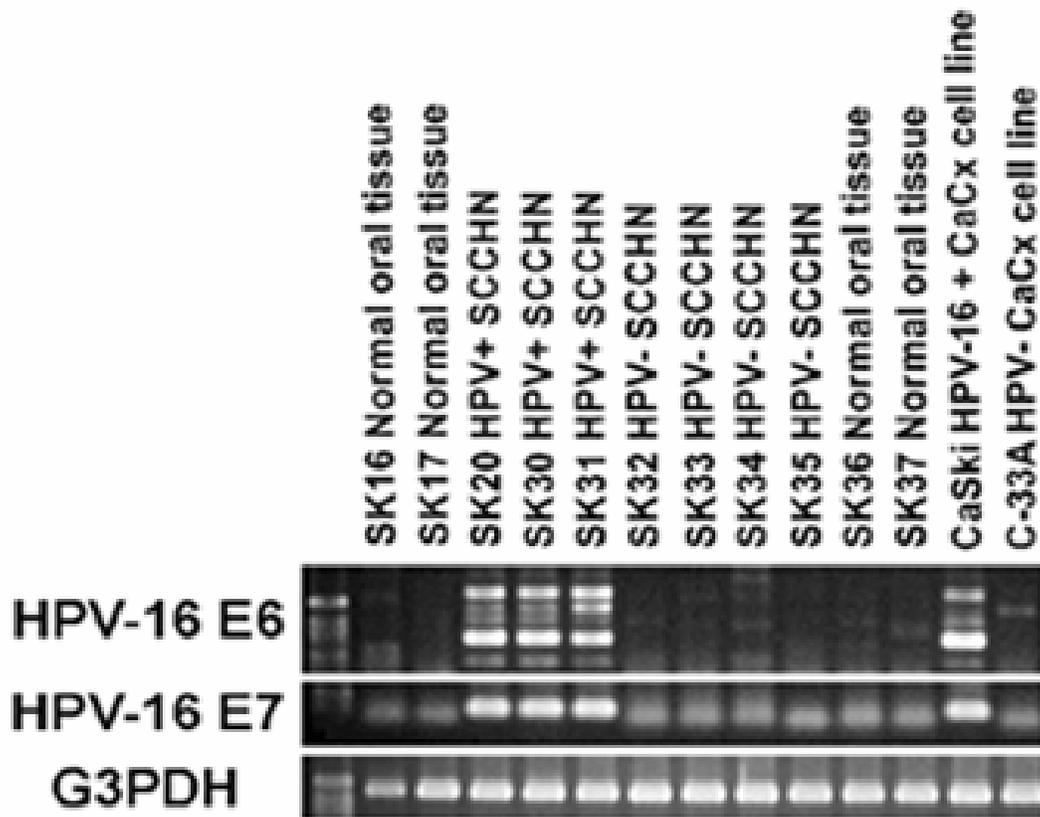


**CaSki (HPV-16 positive) and C-33A (HPV-negative) cervical cell lines were used as controls. Expression of E6 splicing variants (E6, 477 bp, E6*I, 293 bp, E6*II, 176 bp) and E7 (137 bp) genes was found only in SK20, SK30 and SK31 SCCHN samples. The integrity of all the RNA samples was validated by amplification of the house-keeping gene *GAPDH* (452 bp).**

424 genes were found to be up- and down-regulated, respectively, using the SAM-Test analysis (supplementary data, Table B). When the results of the above two statistical tests were overlapped, 228 genes were found to be upregulated and 169 downregulated in HPV-16 positive SCCHN samples as compared to the normal oral tissues (Fig. 6). A list of differentially expressed genes representing various cellular pathways is shown in Table 2.  A complete list of genes is shown in the supplementary data, Table C. Our analysis identified genes that have previously been shown to be overexpressed in HPV-infected cervical or oral epithelial cells such as those involved in DNA replication and cell cycle regulation: cyclin-dependent kinase inhibitor 2A (*CDKN2A*), minichromosome maintenance deficient genes (*MCM2, MCM3*), topoisomerase (DNA) II alpha 170-kDa (*TOP2A*) and proliferating cell nuclear antigen (*PCNA*) (Li*, et al.*, 2004;Ren*, et al.*, 2002;Santin*, et al.*, 2005). These findings suggest that some carcinogenic pathways are frequently affected in HPV-positive cervical as well as HPV-positive oral cancers. Our experiments also identified several genes that were previously implicated in tumourigenesis of different tissues but not in SCCHN. Examples of these include the upregulation of genes involved in cell differentiation or DNA repair such as secreted frizzled-related protein 4 (*SFRP4*), and RAD51 associated protein 1 (*RAD51AP1*) (Horvath*, et al.*, 2004;Iwanaga*, et al.*, 2006). Our results also showed the downregulation of genes involved in proteolysis such as protease serine 3 (*PRSS3*) and chemotaxis, such as chemokine (C-C motif) ligand 14 (*CCL14*), for the first time in HPV-positive SCCHN (Table 2).

**Fig. 6.** **Summary of differentially expressed genes identified by microarray analysis using two different statistical analyses**
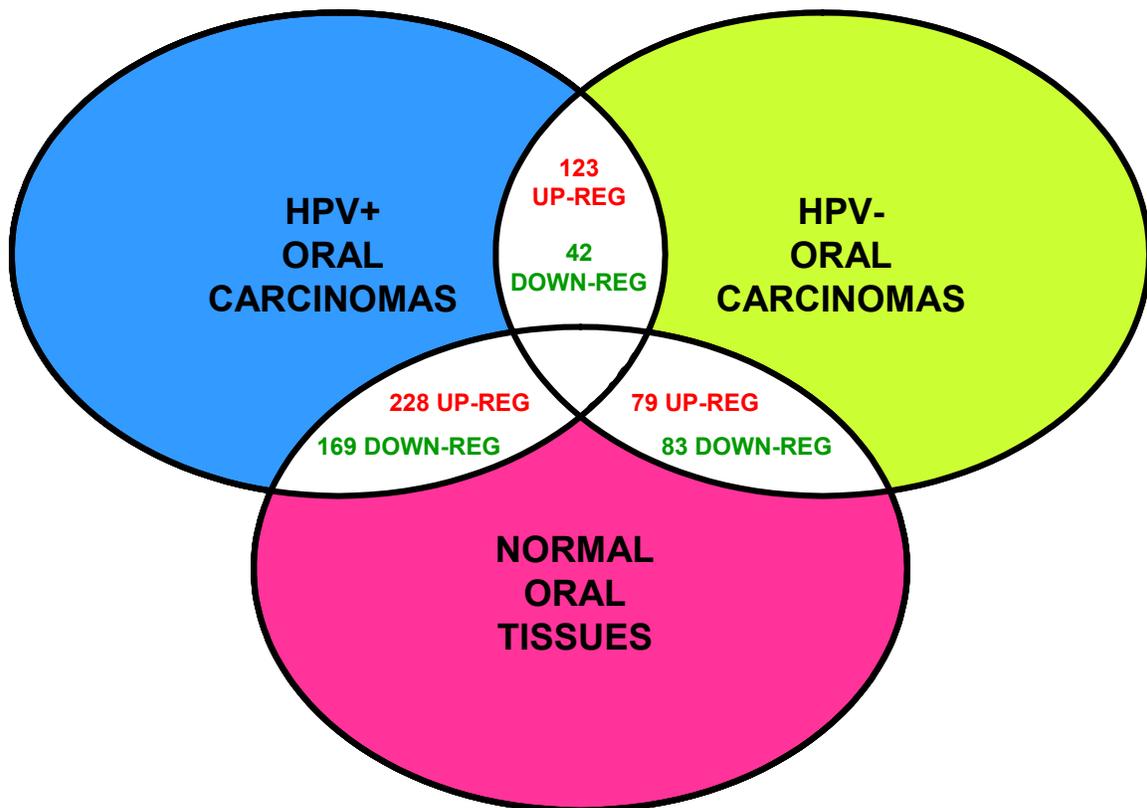


The ovals represent the 3 groups of samples used in our study. The overlapping regions indicate the number of genes found to be up- or down-regulated between two different groups of samples.

*Genes differentially expressed in HPV-negative SCCHN compared to the normal oral mucosa*

Using the pooled variance t-Test analysis, we found that 1,052 genes were upregulated and 850 downregulated in HPV-negative SCCHN samples compared to the normal oral tissues (supplementary data, Table D). The SAM test identified 244 upregulated and 310 downregulated genes in the HPV-negative SCCHN samples (supplementary data, Table E). A comparison of the differentially expressed genes revealed that 79 upregulated and 83 downregulated genes were common in both the pooled variance t-Test and SAM analyses (supplementary data, Table F) (Fig. 6). A list of differentially expressed genes representing various cellular pathways is shown in Table 3. Upregulated genes involved in cell-cell signaling such as parathyroid hormone-like hormone (*PTHLH* ) or angiogenesis such as endothelial cell growth factor 1 (*ECGF1*), as well as downregulated genes involved in signal transduction or apoptosis such as insulin-like growth factor binding protein 5 (*IGFBP5*) and programmed cell death 4 (neoplastic transformation inhibitor) (*PDCD4*) were identified in our studies similar to those found in previous SCCHN studies, providing a good correlation with our results (Kornberg*, et al.*, 2005;Lin*, et al.*, 2002). We also found differentially expressed genes that were not previously reported in SCCHN, although they were implicated in other types of tumours. Such upregulated genes included  those involved in cell proliferation and transcription regulation such as aldo-keto reductase family 1, member C3 (*AKR1C3*), and small nuclear RNA

## TABLE 2. GENES FOUND DIFFERENTIALLY EXPRESSED IN HPV-POSITIVE SCCHN VERSUS NORMAL ORAL EPITHELIUM

| Gene Title | Gene Symbol | Pooled variance t-Test fold change | SAM-Test fold change |
|---|---|---|---|
| **Up-regulated genes** | | | |
| Apoptosis | | | |
| lectin, galactoside-binding, soluble, 1 (galectin 1) | LGALS1 | 3.22 | 2.44 |
| Cell Adhesion | | | |
| collagen, type XI, alpha 1 | COL11A1 | 6.29 | 8.84 |
| Development | | | |
| IGF-II mRNA-binding protein 3 | IMP-3 | 3.28 | 5.06 |
| DNA Replication | | | |
| CDC7 cell division cycle 7 (S. cerevisiae) | CDC7 | 5.38 | 2.34 |
| MCM2 minichromosome maintenance deficient 2, mitotin | MCM2 | 4.39 | 3.36 |
| MCM3 minichromosome maintenance deficient 3 | MCM3 | 3.86 | 2.00 |
| Cell Cycle Regulation | | | |
| cyclin E2 | CCNE2 | 4.23 | 2.23 |
| cyclin-dependent kinase inhibitor 2A | CDKN2A | 16.59 | 8.04 |
| proliferating cell nuclear antigen | PCNA | 7.78 | 2.04 |
| retinoblastoma-like 1 (p107) | RBL1 | 6.09 | 2.37 |
| Transcription Regulation | | | |
| interferon regulatory factor 1 | IRF1 | 6.28 | 2.66 |
| Wnt Receptor Signaling Pathway | | | |
| secreted frizzled-related protein 4 | SFRP4 | 4.45 | 5.61 |
| DNA Repair | | | |
| RAD51 associated protein 1 | RAD51AP1 | 3.65 | 2.14 |
| topoisomerase (DNA) II alpha 170kDa | TOP2A | 4.07 | 3.60 |
| Miscellaneous | | | |
| lumican | LUM | 4.18 | 3.41 |
| mesoderm specific transcript homolog (mouse) | MEST | 8.88 | 3.01 |
| squamous cell carcinoma antigen recognised by T cells 3 | SART3 | 8.44 | 2.10 |
| Chromatin assembly and modification | | | |
| centromere protein F, 350/400ka (mitosin) | CENPF | 3.60 | 2.00 |
| SWI/SNF related, matrix associated, member 3 | SMARCA3 | 5.54 | 2.45 |
| **Down-regulated genes** | | | |
| Apoptosis | | | |
| programmed cell death 4 (neoplastic transformation inhibitor) | PDCD4 | -3.02 | -3.84 |
| Immune response | | | |
| chemokine (C-C motif) ligand 14 / chemokine ligand 15 | CCL14 / | -5.23 | -2.10 |

| | | | |
|---|---|---|---|
| | CCL15 | | |
| interleukin 18 (interferon-gamma-inducing factor) | IL18 | -4.47 | -4.18 |
| trefoil factor 3 (intestinal) | TFF3 | -2.61 | -10.42 |
| Cell Adhesion | | | |
| alpha-2-glycoprotein 1, zinc | AZGP1 | -3.07 | -13.77 |
| Cell Cycle Regulation | | | |
| Ras association (RalGDS/AF-6) domain family 1 | RASSF1 | -3.96 | -2.51 |
| Cell Differentiation | | | |
| deleted in malignant brain tumors 1 | DMBT1 | -4.04 | -9.31 |
| NDRG family member 2 | NDRG2 | -2.64 | -2.56 |
| transglutaminase 3 | TGM3 | -7.14 | -7.55 |
| Cellular Metabolism | | | |
| aldehyde dehydrogenase 1 family, member A1 | ALDH1A1 | -4.60 | -4.85 |
| sulfotransferase family, cytosolic, 2B, member 1 | SULT2B1 | -2.33 | -3.00 |
| UDP glucuronosyltransferase 1 family, A10 / A8 / A7 / A6 | UGT1A | -3.71 | -2.56 |
| DNA Repair | | | |
| mutL homolog 3 (E. coli) | MLH3 | -2.53 | -2.06 |
| Signal Transduction | | | |
| insulin-like growth factor binding protein 6 | IGFBP6 | -3.69 | -7.74 |
| Transcription Regulation | | | |
| BarH-like homeobox 2 | BARX2 | -2.96 | -3.71 |
| Miscellaneous | | | |
| envoplakin | EVPL | -3.26 | -3.55 |
| huntingtin interacting protein-1-related | HIP1R | -3.72 | -2.65 |
| protease, serine, 3 (mesotrypsin) | PRSS3 | -3.88 | -7.21 |
| kallikrein 11 | KLK11 | -3.70 | -7.28 |
| kallikrein 13 | KLK13 | -4.41 | -12.15 |

activating complex, polypeptide 1, 43-kDa (*SNAPC1*), while downregulated genes included those involved in apoptosis or RNA processing such as clusterin (*CLU*) and RNA binding motif protein 3 (*RBM3*) (Dresios*, et al.*, 2005;Penning*, et al.*, 2006;Scaltriti*, et al.*, 2004;Sutherland*, et al.*, 2005;Xie*, et al.*, 2002).

*Identification of genes differentially expressed between HPV-positive and HPV-negative SCCHN*

To identify pathways unique to the pathogenesis of HPV positive and negative SCCHN, we compared the gene expression profiles in these two types of cancers. The pooled t-Test analysis revealed that a total of 1,040 genes were upregulated and 1,363 downregulated in the HPV-positive SCCHN samples compared to the HPV-negative SCCHN samples (supplementary data, Table G). Using the SAM test, we identified 347 upregulated and 175 downregulated genes in the HPV-positive SCCHN compared to HPV-negative SCCHN (supplementary data, Table H). A comparison of the results obtained using the above two tests was done to identify genes present in both of these analyses. This showed that 124 genes were upregulated and 42 downregulated in HPV-16 positive SCCHN as compared to the HPV-negative SCCHN (Fig. 6). A list of differentially expressed genes representing various cellular pathways is shown in Table 4 (a complete list of genes is shown in the supplementary data, Table I).

# TABLE 3. GENES FOUND DIFFERENTIALLY EXPRESSED IN HPV-NEGATIVE SCCHN VERSUS NORMAL ORAL EPITHELIUM

| Gene Title | Gene Symbol | Pooled variance t-Test fold change | SAM-Test fold change |
|---|---|---|---|
| **Up-regulated genes** | | | |
| Angiogenesis | | | |
| endothelial cell growth factor 1 (platelet-derived) | ECGF1 | 2.90 | 2.23 |
| Cell Adhesion | | | |
| Cadherin, EGF LAG seven-pass G-type receptor 1 | CELSR1 | 2.67 | 2.21 |
| cadherin 3, type 1, P-cadherin (placental) | CDH3 | 2.51 | 3.72 |
| calcium/calmodulin-dependent serine protein kinase | CASK | 3.47 | 2.05 |
| laminin, beta 1 | LAMB1 | 3.36 | 2.18 |
| Cell Cycle Regulation | | | |
| cyclin A1 | CCNA1 | 2.16 | 2.03 |
| exostoses (multiple) 1 | EXT1 | 3.25 | 2.99 |
| Cell-Cell Signaling | | | |
| parathyroid hormone-like hormone | PTHLH | 2.61 | 9.26 |
| DNA Replication | | | |
| DNA replication complex GINS protein PSF1 | PSF1 | 2.03 | 2.27 |
| Transcription Regulation | | | |
| basonuclin 1 | BNC1 | 2.83 | 2.27 |
| polymerase (RNA) II (DNA directed) polypeptide H | POLR2H | 2.59 | 2.03 |
| small nuclear RNA activating complex, polypeptide 1, 43kDa | SNAPC1 | 2.16 | 2.42 |
| Immune Response | | | |
| CCAAT/enhancer binding protein (C/EBP), beta | CEBPB | 2.11 | 2.24 |
| Fc fragment of IgG, low affinity IIa, receptor (CD32) | FCGR2A | 2.89 | 2.39 |
| interferon-induced protein with tetratricopeptide repeats 1 | IFIT1 | 2.81 | 2.07 |
| Development | | | |
| S100 calcium binding protein A7 (psoriasin 1) | S100A7 | 2.58 | 4.15 |
| Miscellaneous | | | |
| aldo-keto reductase family 1, member B10 (aldose reductase) | AKR1B10 | 3.05 | 4.47 |
| aldo-keto reductase family 1, member C3 | AKR1C3 | 2.69 | 2.90 |
| chromosome 5 open reading frame 13 | C5orf13 | 2.16 | 2.09 |
| phosphoribosylglycinamide formyltransferase | GART | 2.55 | 2.13 |
| **Down-regulated genes** | | | |
| Cell Adhesion | | | |
| chemokine (C-X-C motif) ligand 12 | CXCL12 | -2.18 | -2.84 |
| claudin 10 | CLDN10 | -2.62 | -4.50 |
| Cell-Cell Signaling | | | |

| | | | |
|---|---|---|---|
| sprouty homolog 2 (Drosophila) | SPRY2 | -2.65 | -2.37 |
| **Cell Cycle Regulation** | | | |
| transforming, acidic coiled-coil containing protein 1 | TACC1 | -2.58 | -2.01 |
| **Cell Differentiation** | | | |
| four and a half LIM domains 1 | FHL1 | -2.74 | -2.03 |
| **Cellular Metabolism** | | | |
| glucosamine (UDP-N-acetyl)-2-epimerase/N-acetylmannosamine kinase | GNE | -2.46 | -2.00 |
| glycine amidinotransferase | GATM | -3.71 | -3.72 |
| **Transcription Regulation** | | | |
| single-stranded DNA binding protein 2 | SSBP2 | -2.77 | -2.02 |
| thioredoxin interacting protein | TXNIP | -2.43 | -2.03 |
| **Miscellaneous** | | | |
| cytochrome P450, family 3, subfamily A, polypeptide 5 | CYP3A5 | -2.13 | -4.45 |
| heat shock 70kDa protein 1A | HSPA1A | -4.49 | -2.04 |
| lymphoid-restricted membrane protein | LRMP | -3.06 | -4.24 |
| serine threonine kinase 39 (STE20/SPS1 homolog, yeast) | STK39 | -2.24 | -2.28 |
| **RNA Processing** | | | |
| RNA binding motif (RNP1, RRM) protein 3 | RBM3 | -2.32 | -2.56 |
| **Signal Transduction** | | | |
| insulin-like growth factor binding protein 5 | IGFBP5 | -3.04 | -2.36 |
| protein kinase, cAMP-dependent, catalytic, beta | PRKACB | -3.36 | -2.29 |
| Ras-related associated with diabetes | RRAD | -3.02 | -3.86 |
| RAP1A, member of RAS oncogene family | RAP1A | -3.77 | -2.27 |
| **Development** | | | |
| chordin-like 1 | CHRDL1 | -2.44 | -2.44 |
| dystrophin (muscular dystrophy, Duchenne and Becker types) | DMD | -2.45 | -2.14 |

Interestingly, a subgroup of 40 upregulated and 19 downregulated genes identified in HPV-positive versus HPV-negative SCCHN analysis, was also found in a comparison between HPV-positive SCCHN and normal oral tissues. Thus, it is likely that the expression of this group of genes (Table 5) is specifically affected by the presence of HPV. Such upregulated genes included those involved in nuclear structure and meiosis (synaptonemal complex protein 2 [*SYCP2*]), DNA repair (replication factor C 5 [*RFC5*]), and DNA methylation (DNA [cytosine-5-]-methyltransferase 1 [*DNMT1*]). Genes involved in proteolysis (kallikrein 7, 8, 10 [*KLK7, KLK8, KLK10*]) and signal transduction (cellular retinoic acid binding protein 2 [*CRABP2*]) were found to be downregulated in HPV-positive SCCHN compared to both HPV-negative SCCHN and normal oral tissues in these analyses.

*Identification of differentially expressed genes present in all SCCHN samples in comparison to normal oral tissues*

We also identified genes whose expression is similarly affected in both HPV- positive and HPV-negative SCCHN as compared to the normal oral mucosa. Based on the pooled t-Test and SAM test analyses, we found that 55 genes were upregulated and 42 downregulated in HPV-positive as well as HPV-negative SCCHN as compared to the normal oral mucosa. A list of differentially expressed genes representing various cellular pathways is shown in Table 6 (a complete list of genes is shown in the supplementary data, Table J). These

## TABLE 4.  GENES FOUND DIFFERENTIALLY EXPRESSED IN HPV-POSITIVE SCCHN VERSUS HPV-NEGATIVE SCCHN

| Gene Title | Gene Symbol | Pooled variance t-Test fold change | SAM-Test fold change |
|---|---|---|---|
| **Up-regulated genes** | | | |
| Apoptosis | | | |
| baculoviral IAP repeat-containing 1 | BIRC1 | 3.08 | 2.11 |
| baculoviral IAP repeat-containing 3 | BIRC3 | 4.65 | 4.56 |
| CD2 antigen (p50), sheep red blood cell receptor | CD2 | 3.56 | 2.60 |
| Cell Cycle | | | |
| B-cell CLL/lymphoma 2 | BCL2 | 2.43 | 2.53 |
| cyclin-dependent kinase inhibitor 1B (p27, Kip1) | CDKN1B | 2.87 | 2.07 |
| cyclin-dependent kinase inhibitor 2C (p18, inhibits CDK4) | CDKN2C | 2.83 | 3.04 |
| dual specificity phosphatase 4 | DUSP4 | 3.56 | 2.07 |
| Transcription Regulation | | | |
| BCL2-associated transcription factor 1 | BCLAF1 | 7.29 | 2.70 |
| Splicing factor proline/glutamine rich | SFPQ | 3.64 | 2.57 |
| thymus high mobility group box protein TOX | TOX | 3.64 | 2.16 |
| zinc finger protein 238 | ZNF238 | 6.28 | 2.04 |
| Immune response | | | |
| lymphotoxin beta (TNF superfamily, member 3) | LTB | 2.56 | 2.22 |
| major histocompatibility complex, class II, DP beta 1 | HLA-DPB1 | 2.51 | 2.40 |
| major histocompatibility complex, class II, DQ alpha 1/ 2 | HLA-DQA1/A2 | 2.47 | 3.29 |
| Cell Adhesion | | | |
| vascular cell adhesion molecule 1 | VCAM1 | 4.63 | 3.96 |
| Signal Transduction | | | |
| phosphoinositide-3-kinase, catalytic, gamma polypeptide | PIK3CG | 2.87 | 2.35 |
| RAS guanyl releasing protein 1 | RASGRP1 | 2.50 | 2.09 |
| DNA Repair | | | |
| RecQ protein-like (DNA helicase Q1-like) | RECQL | 3.90 | 2.16 |
| Miscellaneous | | | |
| ecotropic viral integration site 2A | EVI2A | 3.33 | 2.11 |
| tubulin, gamma complex associated protein 3 | TUBGCP3 | 4.92 | 2.14 |
| **Down-regulated genes** | | | |
| Cellular Metabolism | | | |
| aldo-keto reductase family 1, member B10 | AKR1B10 | -3.73 | -7.30 |
| aldo-keto reductase family 1, member C3 | AKR1C3 | -3.23 | -4.44 |
| solute carrier family 7, member 5 | SLC7A5 | -2.25 | -2.08 |
| UDP glucuronosyltransferase 2 family, polypeptide B4 | UGT2B4 | -4.97 | -2.71 |

| Cell Proliferation | | | |
|---|---|---|---|
| block of proliferation 1 | BOP1 | -2.48 | -2.19 |
| **Signal Transduction** | | | |
| inositol(myo)-1(or 4)-monophosphatase 2 | IMPA2 | -2.05 | -2.05 |
| **Cell Cycle** | | | |
| cyclin A1 | CCNA1 | -2.18 | -2.31 |
| **Chromatin assembly and modification** | | | |
| histone 1, H2bg | HIST1H2BG | -2.63 | -2.70 |
| **Protein modification** | | | |
| carboxypeptidase A4 | CPA4 | -2.58 | -2.14 |
| **Immune response** | | | |
| interleukin 1 receptor accessory protein | IL1RAP | -2.24 | -2.01 |
| killer cell lectin-like receptor subfamily G, member 1 | KLRG1 | -5.70 | -2.18 |
| serine (or cysteine) proteinase inhibitor, clade B, member 4 | SERPINB4 | -2.0 | -2.38 |
| **Cell Differentiation** | | | |
| S100 calcium binding protein A7 (psoriasin 1) | S100A7 | -2.7 | -6.69 |
| small proline-rich protein 1B (cornifin) | SPRR1B | -3.40 | -2.28 |
| **Translation Regulation** | | | |
| eukaryotic translation initiation factor 3, subunit 8, 110kDa | EIF3S8 | -2.27 | -2.14 |
| eukaryotic translation initiation factor 5A | EIF5A | -6.05 | -2.79 |
| **Transcription Regulation** | | | |
| achaete-scute complex-like 2 (Drosophila) | ASCL2 | -4.19 | -2.08 |
| bolA-like 2 (E. coli) | BOLA2 | -2.09 | -2.02 |
| **Miscellaneous** | | | |
| corneodesmosin | CDSN | -2.97 | -2.62 |
| seven transmembrane domain protein | NIFIE14 | -5.28 | -2.05 |

**TABLE 5. GENES FOUND DIFFERENTIALLY EXPRESSED EXCLUSIVELY IN HPV-POSITIVE SCCHN VERSUS HPV-NEGATIVE SCCHN OR NORMAL ORAL EPITHELIUM**

| Gene Title | Gene Symbol | Pooled variance t-Test fold change | SAM-Test fold change |
|---|---|---|---|
| **Up-regulated genes** | | | |
| Apoptosis | | | |
| tumor necrosis factor, alpha-induced protein 3 | TNFAIP3 | 2.40 | 2.58 |
| Cell Cycle | | | |
| synaptonemal complex protein 2 | SYCP2 | 13.58 | 12.54 |
| Transcription Regulation | | | |
| DNA (cytosine-5-)-methyltransferase 1 | DNMT1 | 6.22 | 2.22 |
| nuclear factor (erythroid-derived 2)-like 3 | NFE2L3 | 2.85 | 2.67 |
| sin3-associated polypeptide, 30kDa | SAP30 | 2.23 | 2.12 |
| suppressor of zeste 12 homolog (Drosophila) | SUZ12 | 2.86 | 2.17 |
| Cell Differentiation | | | |
| B cell RAG associated protein | GALNAC4S-6ST | 2.47 | 2.62 |
| Immune response | | | |
| CD200 antigen | CD200 | 3.10 | 3.00 |
| chemokine (C-X-C motif) ligand 11 | CXCL11 | 2.05 | 3.52 |
| chemokine (C-X-C motif) ligand 9 | CXCL9 | 2.87 | 3.93 |
| lymphocyte antigen 75 | LY75 | 4.19 | 4.41 |
| phospholipase A2, group IVC | PLA2G4C | 4.12 | 2.27 |
| Cell Proliferation | | | |
| chemokine (C-X-C motif) ligand 10 | CXCL10 | 2.18 | 2.55 |
| cysteine and glycine-rich protein 2 | CSRP2 | 3.36 | 2.34 |
| pim-2 oncogene | PIM2 | 2.44 | 2.07 |
| Cell Adhesion | | | |
| collagen, type VII, alpha 1 | COL7A1 | 2.61 | 3.35 |
| collagen, type XXI, alpha 1 | COL21A1 | 15.43 | 3.80 |
| thrombospondin 4 | THBS4 | 3.09 | 3.30 |
| Signal Transduction | | | |
| chemokine (C-X-C motif) receptor 4 | CXCR4 | 3.37 | 4.18 |
| chemokine orphan receptor 1 | CMKOR1 | 2.34 | 2.67 |
| inositol 1,4,5-trisphosphate 3-kinase B | ITPKB | 2.83 | 2.14 |
| lymphoid enhancer-binding factor 1 | LEF1 | 3.27 | 3.87 |
| phosphoinositide-3-kinase, regulatory subunit 3 | PIK3R3 | 3.15 | 2.37 |
| Ras association (RalGDS/AF-6) domain family 4 | RASSF4 | 3.06 | 3.45 |
| SMAD, mothers against DPP homolog 5 (Drosophila) | SMAD5 | 2.88 | 2.67 |
| DNA Repair | | | |
| replication factor C (activator 1) 5, 36.5kDa | RFC5 | 5.01 | 2.18 |
| thymidylate synthetase | TYMS | 3.01 | 2.09 |

| Chromatin assembly and modification | | | |
|---|---|---|---|
| SMC5 structural maintenance of chromosomes 5-like 1 | SMC5L1 | 2.88 | 2.14 |
| Protein modification | | | |
| CDC-like kinase 4 | CLK4 | 7.15 | 3.21 |
| glutaminyl-peptide cyclotransferase (glutaminyl cyclase) | QPCT | 2.37 | 3.43 |
| proteasome subunit, beta type, 9 | PSMB9 | 2.15 | 2.07 |
| tyrosine kinase 2 | TYK2 | 3.68 | 2.66 |
| Miscellaneous | | | |
| butyrophilin, subfamily 3, member A3 | BTN3A3 | 2.04 | 2.20 |
| chromosome 18 open reading frame 1 | C18orf1 | 4.05 | 3.89 |
| chromosome X open reading frame 45 | CXorf45 | 5.59 | 3.53 |
| Endothelin converting enzyme 1 | ECE1 | 2.68 | 2.30 |
| fatty acid desaturase 2 | FADS2 | 9.05 | 2.23 |
| inositol polyphosphate-5-phosphatase F | INPP5F | 3.67 | 2.92 |
| spastin | SPAST | 4.91 | 2.15 |
| stress 70 protein chaperone, microsome-associated, 60kDa | STCH | 2.57 | 2.33 |
| **Down-regulated genes** | | | |
| Cellular Metabolism | | | |
| cytidine deaminase | CDA | -2.07 | -5.50 |
| monoamine oxidase A | MAOA | -2.69 | -2.34 |
| UDP glucuronosyltransferase 1 family, polypeptide A6 | UGT1A6 | -2.04 | -3.85 |
| Cell Proliferation | | | |
| amphiregulin (schwannoma-derived growth factor) | AREG | -3.12 | -2.55 |
| CD5 antigen (p56-62) | CD5 | -4.31 | -2.15 |
| Signal Transduction | | | |
| cellular retinoic acid binding protein 2 | CRABP2 | -4.32 | -6.28 |
| wingless-type MMTV integration site family, member 4 | WNT4 | -3.91 | -2.13 |
| Chromatin assembly and modification | | | |
| H2B histone family, member S | H2BFS | -2.91 | -2.67 |
| histone 1, H2bd | HIST1H2BD | -2.59 | -2.47 |
| Cell Differentiation | | | |
| glucosaminyl (N-acetyl) transferase 2, I-branching enzyme | GCNT2 | -3.32 | -2.51 |
| kallikrein 7 (chymotryptic, stratum corneum) | KLK7 | -2.49 | -4.68 |
| kallikrein 8 (neuropsin/ovasin) | KLK8 | -10.36 | -4.6 |
| kallikrein 10 | KLK10 | -4.23 | -6.75 |
| keratin, hair, acidic, 1 | KRTHA1 | -2.49 | -2.29 |
| Transcription Regulation | | | |
| paired-like homeodomain transcription factor 2 | PITX2 | -3.08 | -6.26 |
| Cell Adhesion | | | |
| lymphocyte antigen 6 complex, locus D | LY6D | -2.24 | -4.29 |
| plakophilin 3 | PKP3 | -3.08 | -2.57 |
| Miscellaneous | | | |
| kinesin family member 1C | KIF1C | -2.62 | -2.04 |
| solute carrier family 24, member 3 | SLC24A3 | -3.01 | -2.04 |

## TABLE 6.   GENES FOUND DIFFERENTIALLY EXPRESSED IN ALL SCCHN SAMPLES VERSUS NORMAL ORAL EPITHELIUM

| Gene Title | Gene Symbol | Pooled variance t-Test fold change | SAM-Test fold change |
|---|---|---|---|
| **Up-regulated genes** | | | |
| DNA Replication | | | |
| CDC6 cell division cycle 6 homolog (S. cerevisiae) | CDC6 | 3.75 | 2.07 |
| Cell Proliferation | | | |
| hepatoma-derived growth factor, related protein 3 | HDGFRP3 | 6.32 | 2.87 |
| keratin 16 (focal non-epidermolytic palmoplantar keratoderma) | KRT16 | 2.78 | 3.51 |
| Transcription Regulation | | | |
| signal transducer and activator of transcription 1, 91kDa | STAT1 | 5.01 | 2.94 |
| Signal Transduction | | | |
| CD14 antigen | CD14 | 2.15 | 2.23 |
| insulin-like growth factor 2 receptor | IGF2R | 4.72 | 2.52 |
| secreted protein, acidic, cysteine-rich (osteonectin) | SPARC | 6.22 | 4.18 |
| Immune Response | | | |
| Fc fragment of IgG, low affinity IIIb, receptor (CD16b) | FCGR3B | 2.63 | 3.29 |
| interferon, alpha-inducible protein 27 | IFI27 | 2.21 | 2.68 |
| interferon-induced protein 44 | IFI44 | 2.88 | 2.51 |
| Development | | | |
| chondroitin sulfate proteoglycan 2 (versican) | CSPG2 | 2.85 | 3.59 |
| collagen, type I, alpha 1 | COL1A1 | 2.23 | 3.39 |
| keratin 14 (epidermolysis bullosa simplex) | KRT14 | 6.93 | 3.12 |
| Cell adhesion | | | |
| matrix metalloproteinase 3 (stromelysin 1, progelatinase) | MMP3 | 5.02 | 2.55 |
| periostin, osteoblast specific factor | POSTN | 5.47 | 10.09 |
| transforming growth factor, beta-induced, 68kDa | TGFBI | 2.76 | 2.05 |
| Miscellaneous | | | |
| cathepsin L2 | CTSL2 | 2.20 | 6.18 |
| plasminogen activator, urokinase | PLAU | 4.01 | 3.33 |
| secreted phosphoprotein 1 (osteopontin, bone sialoprotein I) | SPP1 | 7.00 | 18.90 |
| serine (or cysteine) proteinase inhibitor, clade H | SERPINH1 | 5.74 | 2.43 |
| **Down-regulated genes** | | | |
| Apoptosis | | | |
| clusterin | CLU | -3.20 | -4.54 |
| growth arrest and DNA-damage-inducible, beta | GADD45B | -3.10 | -2.55 |
| mal, T-cell differentiation protein | MAL | -2.14 | -2.50 |
| Cell adhesion | | | |

| | | | |
|---|---|---|---|
| annexin A9 | ANXA9 | -2.54 | -2.17 |
| multimerin 1 | MMRN1 | -3.70 | -4.51 |
| Immune response | | | |
| cysteine-rich protein 1 (intestinal) | CRIP1 | -7.21 | -5.16 |
| cysteine-rich secretory protein 3 | CRISP3 | -5.75 | -69.68 |
| interleukin 2 | IL2 | -2.97 | -3.16 |
| prostate stem cell antigen | PSCA | -4.63 | -8.65 |
| Signal Transduction | | | |
| proline-rich protein HaeIII subfamily 1-2 | PRH1-2 | -3.17 | -20.84 |
| Cellular Metabolism | | | |
| amylase, alpha 1A; salivary / 1B / 1C / 2A/ 2B | AMY1 | -5.15 | -7.67 |
| glycerol-3-phosphate dehydrogenase 1-like | GPD1L | -3.25 | -3.43 |
| Protein Modification | | | |
| ubiquitin-like 3 | UBL3 | -2.28 | -2.67 |
| Transcription Regulation | | | |
| p300/CBP-associated factor | PCAF | -2.51 | -2.43 |
| paternally expressed 3 | PEG3 | -6.73 | -6.01 |
| Development | | | |
| keratin 4 | KRT4 | -2.72 | -2.82 |
| sciellin | SCEL | -2.28 | -2.55 |
| Miscellaneous | | | |
| ATP-binding cassette, sub-family A (ABC1), member 8 | ABCA8 | -2.54 | -3.05 |
| dual specificity phosphatase 5 | DUSP5 | -2.54 | -2.97 |
| nucleobindin 2 | NUCB2 | -3.50 | -3.12 |

genes may represent common pathways altered during SCCHN irrespective of the presence or absence of HPVs. The upregulated genes included those involved in cell cycle regulation, epidermis development and cell adhesion such as signal transducer and activator of transcription 1 (*STAT1*), keratin 14, 16 (*KRT14, KRT16*) and transforming growth factor, beta-induced, 68-kDa (*TGFBI*). Also, downregulated genes found in both types of SCCHN samples included those involved in chromatin remodeling, cell-cell adhesion and apoptosis such as p300/CBP-associated factor (*PCAF*), annexin A9 (*ANXA9*) and mal, T-cell differentiation protein (*MAL*).

*Identification of several gene expression clusters in HPV-positive, HPV-negative SCCHN and normal oral tissues*

Unsupervised hierarchical cluster analysis was developed using all SCCHN specimens as well as normal oral tissues. After normalization and filtering with different criteria (described in materials and methods), we obtained a set of 8,286 genes with significant variation in their expression in one or more of the total of three types of samples (HPV-positive SCCHN, HPV-negative SCCHN, and control oral mucosa). From these filtered genes, we identified five clusters that showed differences in gene expression not only between all the SCCHN samples and the normal oral epithelium but also between HPV-positive SCCHN and HPV-negative SCCHN (Fig. 7). Interestingly, some of the genes found in this analysis were also found in the supervised statistical analysis such

52

as *ZNF238, DNMT1, MCM3*, aldo-keto reductase family 1, member C3 (*AKR1C3*), growth factor receptor-bound protein 10 (*GRB10*), tyrosine kinase 2 (*TYK2*) and squamous cell carcinoma antigen recognized by T cells 3 (*SART3*) (underlined in Fig. 7). After this analysis, we mapped the chromosomal location of the genes found in these clusters in order to visualize genomic regions that could contain a significant number of genes altered in their expression in HPV-positive SCCHN as compared to HPV-negative SCCHN as well as the normal oral epithelium. Fig. 8 shows that several such genes are present in chromosomal regions 1p34-p36 and 12q21-q24 (a complete list of the clustered genes is shown in the supplementary data, Table K). The genes found in the 1p34-p36 region are involved in repression of splicing (FUS interacting protein 1 [*FUSIP1*]), histone and nucleosomal binding (nuclear autoantigenic sperm protein [*NASP*], high-mobility group nucleosomal binding domain 2 [*HMGN2*]) and intracellular vesicle traffic (taxilin alpha [*TXLNA*]). On the other hand, the genes found in the chromosome region 12q21-q24 are involved in *JAK/STAT*-signaling pathway (cysteine and glycine-rich protein 2 [*CSRP2*]), regulation of nuclear architecture (thymopoietin [*TMPO/LAP2*]), myotonic dystrophy kinase cascade (citron [*CIT*]) and regulation of mRNA splicing and HIV gene expression and replication (squamous cell carcinoma antigen recognized by T cells 3 [*SART3*]).

# Fig. 7.     UNSUPERVISED HIERARCHICAL CLUSTER ANALYSIS IN OROPHARYNGEAL TISSUE SAMPLES
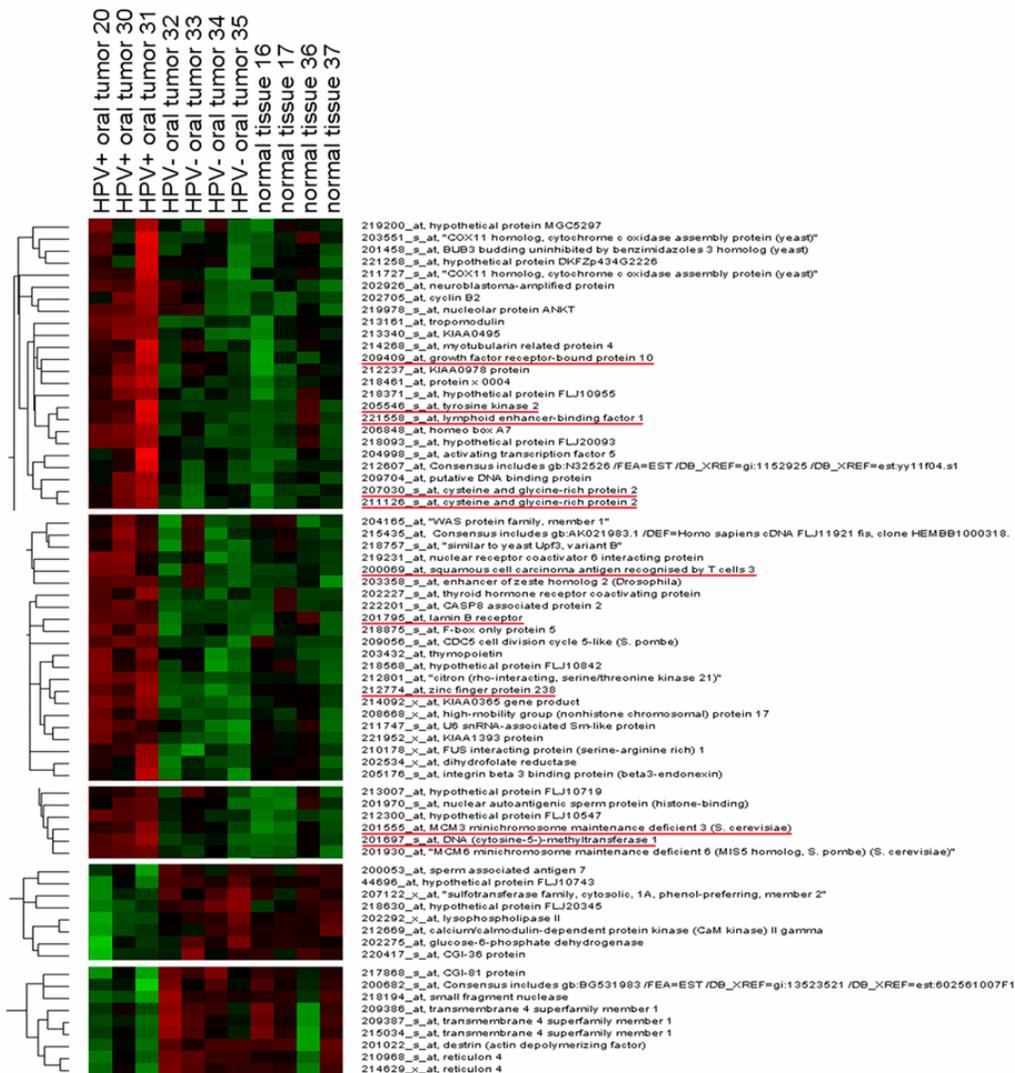


From a total of 22,215 transcripts on the microarray, 8,286 genes showed variations in expression across all the samples. From this group of filter genes, we identified 7 clusters of genes that show differential expression between the 3 groups of samples (HPV-positive SCCHN, HPV-negative SCCHN and normal oral mucosa). Genes also found in the supervised statistical analysis are underlined.

**Fig. 8.    CHROMOSOMAL LOCATION OF GENES FOUND TO BE DIFFERENTIALLY EXPRESSED IN THE HIERARCHICAL CLUSTER ANALYSIS**
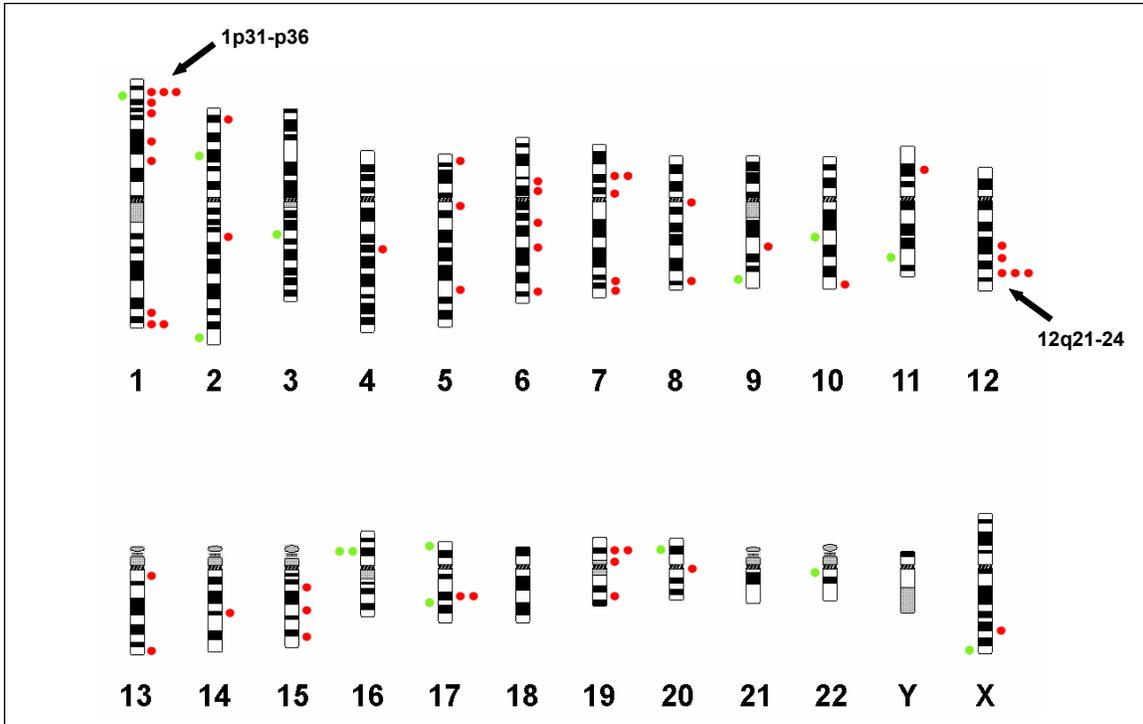


Squares to the right of the chromosomes indicate genes with increased expression in HPV-positive oropharyngeal squamous carcinomas as compared to HPV-negative carcinomas or the normal oral epithelium (clusters 1, 2 and 3). Circles to the left of the chromosomes indicate genes with reduced expression in HPV-positive oropharyngeal squamous carcinomas as compared to the HPV-negative and normal samples (clusters 4 and 5). The chromosome map was obtained and modified from the Department of Pathology, University of Washington web page http://www.pathology.washington.edu/research/cytopages/idiograms/human/ (Idiogram Album: Human copyright © 1994 David Adler).

*Validation of microarray expression data by real-time quantitative RT-PCR analysis*

We carried out quantitative real-time RT-PCR (QRT-PCR) analysis of a few representative genes in order to validate the differential gene expression profiles obtained by microarray analysis. Four genes found to be upregulated (*CDKN2A, SYCP2, RFC5, ZNF238*) and two genes found to be downregulated (KLK8, CRABP2) in HPV-positive SCCHN samples as compared to both the HPV-negative SCCHN as well as normal oral tissues were selected for validation (Fig. 9). The QRT-PCR analysis confirmed that *CDKN2A* and *SYCP2* were up-regulated in all the HPV-positive SCCHN samples as compared to the HPV-negative SCCHN and normal epithelium, whereas *RFC5* and *ZNF238* were found to be up-regulated in two out of three HPV-positive SCCHN (Fig. 9). Similarly, the *KLK8* and *CRABP2* genes were downregulated in HPV-positive SCCHN. However, there was variation in the extent of differential expression of the above genes in individual samples (Fig. 9). The above data showed that the results of QRT-PCR were consistent with those of the microarray analysis, although variations were observed in the fold-difference between these two types of analyses.

**2.4 DISCUSSION**

Recently, an improved understanding of SCCHN has resulted from studies on carcinogenic (tobacco, alcohol, etc.) versus viral (HPV, EBV, etc.) exposures. Several investigators have compared these subgroups in order to identify the pathway (s) and potential therapeutic targets which are unique to these tumours, which are often localized to distinct head and neck sub sites. Indeed, our study is unique since we focused specifically on tumours from the oropharyngeal site, in order to compare gene expression profiles between HPV-16 positive SCCHN tumours and normal squamous oropharyngeal mucosa.

To date, no study has described such a tri-partite analysis, including the epithelial squamous cells from which both tumours were derived. This significantly strengthens our ability to conclude that oncogenic signaling identified in HPV-positive and HPV-negative SCCHN is appropriately attributed to the transformation process and not to the genes expressed in oropharyngeal squamous cells. Although the number of specimens used in our study is small, a unique feature of this study is that we have exclusively utilized tissues derived from the oropharyngeal site. While HPV infection can be identified throughout tumours of the head and neck, a disproportionate frequency is found in the oropharynx (approximately 50%). Thus, identifying tumours in this region eliminates other sub site causes of tumour heterogeneity which have been observed by others. Finally, detailed statistical analysis using multiple tests of

# Fig. 9. VALIDATION OF MICROARRAY DATA IN OROPHARYNGEAL TISSUE SAMPLES BY QUANTITATIVE RT-PCR
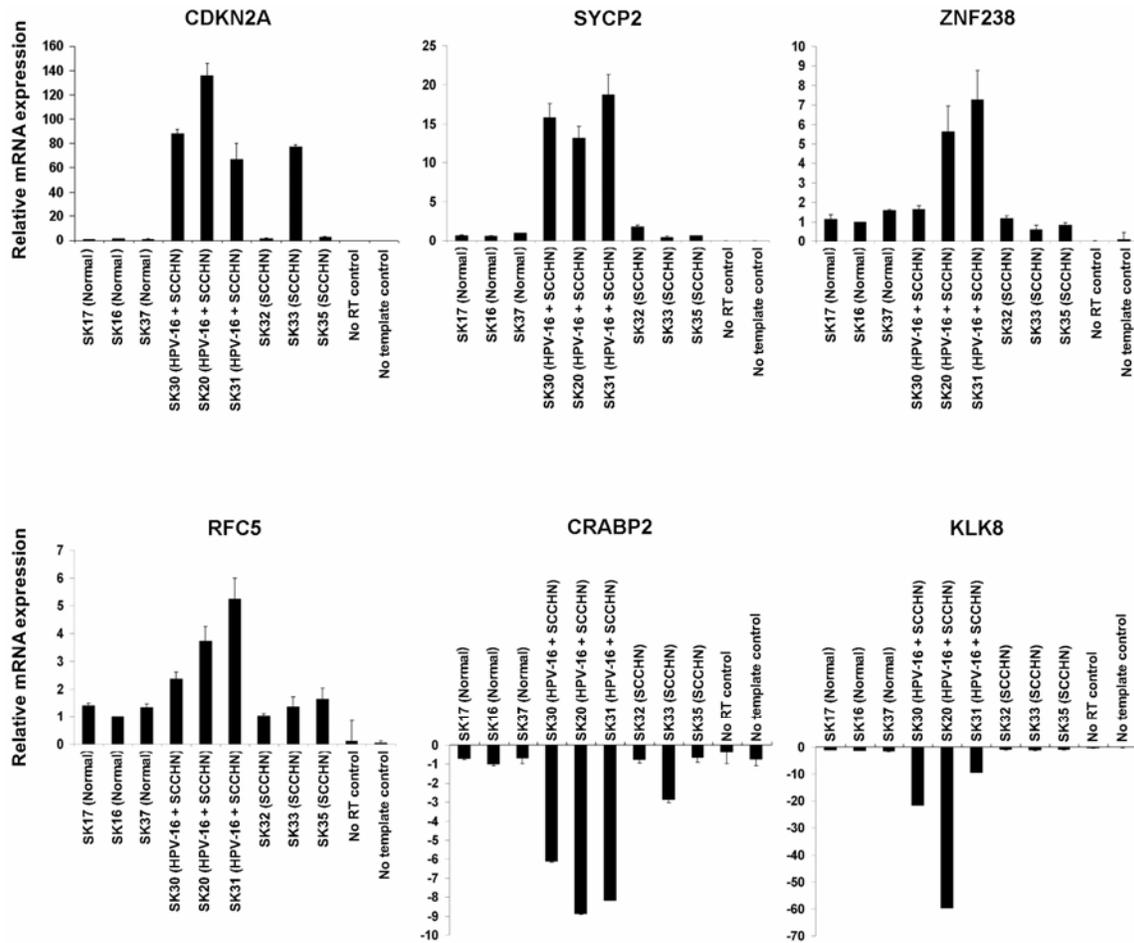


Total RNA from tissue samples was used to verify four upregulated genes (cyclin-dependent kinase inhibitor 2A [*CDKN2A*], synaptonemal complex protein 2 [*SYCP2*], zinc finger protein 238 [*ZNF238*] and replication factor C 5 [*RFC5*]), and two downregulated genes (cellular retinoic acid binding protein 2 [*CRABP2*], kallikrein 8 [*KLK8*]) that were differentially expressed based on the microarray data. All reactions were performed in triplicates and the error bar represents the standard deviation. Relative expression of the target gene was calculated using the 2 delta CT method, where *GAPDH* was used as the endogenous control gene.

significance enhances our ability to identify genes that are expressed uniquely in HPV-16 positive tumours. We have also carried out quantitative RT-PCR analysis to validate the expression of some of the genes found to be differentially expressed in the microarray analysis (Fig. 9). This provides additional confidence in the genes identified by the microarray analysis.

After confirming the expression of the HPV-16 E6 and E7 genes in HPV-positive, HPV-negative SCCHN and normal squamous mucosal samples, we proceeded with the global gene expression analysis. After analyzing the microarray data with two different normalizations and statistical tests (pooled variance t-Test and SAM-Test) in order to identify differentially expressed genes that were robust enough to appear in both analyses, we found that 228 genes were upregulated and 169 downregulated in HPV-positive SCCHN samples as compared to the normal oral tissues. The upregulated genes included those involved in DNA replication and cell cycle regulation such as *CDKN2A* (*p16INK4a*), *MCM2, PCNA, TOP2A* and *RBL1* that have been previously identified to be affected in HPV-infected cervical and oral carcinomas (Li*, et al.*, 2004;Ren*, et al.*, 2002;Santin*, et al.*, 2005). Several studies have shown either direct or indirect regulation of these genes by the tumour suppressor protein p53 (*RBL1, PCNA, TOP2A*) or by E2F transcription factors (*CDKN2A* and *MCM2*) (Ren*, et al.*, 2002;Wei*, et al.*, 2006). One possible explanation for the transcriptional deregulation of this group of genes may be the degradation of p53 and pRB (that bind and repress the activity of the family of E2F transcription

factors) by the viral oncogenes E6 and E7, respectively. Interestingly, we also found several genes differentially expressed in our microarray analysis in HPV-positive SCCHN compared to the normal tissue that have been shown to be regulated by p53 or E2F such as the tumour necrosis factor receptor superfamily, member 10b (*TNFRSF10B*), mutS homolog 6 (*MSH6*), caldesmon 1 (*CALD1*), p300/CBP-associated factor (*PCAF*), collagen, type IV, alpha 1 (*COL4A1*), nidogen 2 (*NID2*), astrotactin 2 (*ASTN2*), cytochrome P450, family 4, subfamily F, polypeptide 3 (*CYP4F3*), DNA-damage-inducible transcript 4 (*DDIT4*), chondroitin sulfate proteoglycan 2 (*CSPG2*), insulin-like growth factor binding protein 6 (*IGFBP6*), keratin 17 (*KRT17*), CDC6 cell division cycle 6 homolog (*CDC6*), replication factor C4, 37kDa (*RFC4*), flap structure-specific endonuclease 1 (*FEN1*) and SWI/SNF related, matrix associated, actin dependent regulator of chromatin, subfamily a, member 3 (*SMARCA3*) (see supplementary data, table C) (Kannan*, et al.*, 2001;Ren*, et al.*, 2002;Wei*, et al.*, 2006;Zhao*, et al.*, 2000). These data are consistent with the results of previous studies and support the notion that HPV-positive SCCHN has a specific transcription profile depends partly on E6 and E7 expression. Specific E6 or E7 knockdown studies will clarify the validity of this conclusion. After comparing the HPV-positive SCCHN samples against the normal oral tissues, we also identified a group of differentially expressed genes previously implicated in tumourigenesis of different tissues but not in SCCHN. These included the upregulated genes *SFRP4* and *RAD51AP1* and the downregulated gene *PRSS3* (Table 2). *SFRP4* has been found to be upregulated in prostate cancers and in microsatellite stable

endometrial cancers (Horvath, *et al.*, 2004;Risinger, *et al.*, 2005). This protein has also been implicated in the inhibition of the Wnt-signaling cascade that is involved in the regulation of cellular proliferation (Horvath, *et al.*, 2004). *RAD51AP1* is an S phase cell cycle checkpoint gene that has recently been found to be directly induced by E2F1 (Iwanaga, *et al.*, 2006). The overexpression of *RAD51AP1* further supports the idea that dysregulation of E2F inducible transcripts occurs in HPV-positive SCCHN due to the inactivation of pRB by E7 expression. *PRSS3* is a member of the serine protease family that shares significant homology with trypsinogen 1 and 2 and has been classified as a tumour suppressor gene in several types of cancers such as esophageal squamous cell carcinomas, gastric adenocarcinomas and non-small cell lung cancers (Marsit, *et al.*, 2005). There is also some evidence that *PRSS3* is downregulated by promoter hypermethylation in lung cancers (Marsit, *et al.*, 2005). Interestingly our results show that *DNMT1*, which is responsible for transcriptional silencing through DNA methylation, is upregulated in HPV-positive SCCHN suggesting possible hypermethylation of genes including *PRSS3* by this enzyme.

After statistical analysis of the data obtained with HPV-negative SCCHN, we found 79 upregulated and 83 downregulated genes as compared to the normal oropharyngeal squamous mucosa (supplementary data, Table F). The upregulated genes such as *PTHLH* and *ECGF1*, and downregulated genes such as *IGFBP5* and *PDCD4* were also reported in previous SCCHN studies providing

a validation of our results (Kornberg, *et al.*, 2005;Lin, *et al.*, 2002). At the same time, we found novel upregulated genes such as *AKR1C3* and *SNAPC1* and downregulated genes such as *CLU* and *RBM3* that were previously not known to be differentially expressed in SCCHN. *AKR1C3* is a hydroxysteroid dehydrogenases involved in the regulation of local concentration of androgens and estrogens in hormone dependent tissues like prostate, breast and endometrium (Penning, *et al.*, 2006). *AKR1C3* has been found overexpressed at the mRNA and protein level in prostate cancer specifically in the stromal cells (Penning, *et al.*, 2006). *SNAPC1* is a subunit of the TBP-TAF complex involved in the transcription of both RNA polymerase II and III dependent small nuclear RNA genes. *SNAPC1* has been linked to the tumourigenesis of breast cancers (Xie, *et al.*, 2002). *CLU* is a multifunctional gene which is involved in spermatogenesis and in the control of the immunological complement cascade (Scaltriti, *et al.*, 2004). It has recently been shown that *CLU* is downregulated in low- and high-grade human prostate cancers and it may be involved in supporting cell survival and the induction of programmed cell death (Scaltriti, *et al.*, 2004). *RBM3*, a glycine-rich RNA-binding protein induced by cold-stress, has recently been found to have apoptotic modulatory capabilities with a possible role as a tumour suppressor gene (Sutherland, *et al.*, 2005). Interestingly, there is evidence that *RBM3* can change the expression of some microRNAs and enhance global protein synthesis suggesting an involvement of a homeostatic mechanism that regulates global levels of protein synthesis under normal and cold-stress conditions (Dresios, *et al.*, 2005).

We found 124 upregulated and 42 downregulated genes in HPV-positive SCCHN as compared to HPV-negative SCCHN (supplementary data, Table I). Among this group of genes, we found a subset of genes (41 upregulated and 19 downregulated) that were also differentially expressed in HPV-positive SCCHN as compared to the normal oropharyngeal mucosa. It is likely that the differential expression of these 60 genes (Table 5) is a consequence of the presence of HPV-16. This group of genes included the upregulated genes *SYCP2, RFC5, ZNF238* and *DNMT1*, and the downregulated genes *KLK8* and *CRABP2. SYCP2* is part of the synaptonemal complex involved in forming lateral elements and cross bridges that contribute to pairing of sister chromatids during meiosis and probably are involved in the interaction between chromatin and nuclear envelope (Shakib*, et al.*, 2005). There is evidence that HPV infection in cervical cells promotes morphological changes in the nuclear membrane as well as condensation of chromatin attached to the inner part of the nuclear membrane (Bollmann*, et al.*, 2005). Thus, it is possible that *SYCP2* may be involved in chromatin-nuclear envelope interactions in HPV-positive SCCHN.

A recent study by Slebos and colleagues also reported up-regulation of the *SYCP2* and cyclin-dependent kinase inhibitor 2C (*CDKN2C*) in HPV-positive SCCHN as compared to HPV-negative SCCHN, similar to our data obtained using the two statistical methods (Slebos*, et al.*, 2006).  Since the data presented by Slebos and colleagues were based solely on the use of the SAM statistical method, we also compared the differentially expressed genes in our study

identified only by using this method (supplementary Table H). This analysis showed that the CDKN2A, NEFH and FLJ12973 genes were similarly affected in both the studies. This further strengthens the likelihood that the above genes play important roles in the carcinogenesis of HPV-related SCCHN. Based on the use of SAM statistical method alone, we found that the *SMARCA2* gene was up-regulated in HPV-positive SCCHN as compared to the HPV-negative SCCHN (Table H). This gene belongs to a family of proteins involved in transcriptional regulation by conformational changes of nucleosomes (Yamamichi, *et al.*, 2005). The closely-related and functionally similar *SMARCA3* gene was found to be up-regulated in the study of Slebos and colleagues (Slebos, *et al.*, 2006). Similarly, while replication factor C 5 (*RFC5*) was found to be overexpressed in HPV-positive SCCHN in our study (Table 5), Slebos and colleagues reported overexpression of *RFC4*. Theses genes encode part of the RFC protein complex required, in conjunction with *PCNA*, in chromosomal DNA replication by DNA polymerase delta and epsilon. *RFC5* protein not only interacts directly with PCNA but also with RAD24 which is involved in DNA damage checkpoint control in *Saccharomyces cerevisiae* (Kim, *et al.*, 2003;Naiki, *et al.*, 2000). These findings suggest a possible role of *RFC5* upon infection of the oropharyngeal tissue with HPVs that may promote a DNA damage response.

The transcription factor *ZNF238* (also known as *RP58*) is a component of a C2H2-type DNA-binding zinc finger protein that acts as a repressor of transcriptionally silent heterochromatin regions. Interestingly, the DNA

64

methyltransferase protein Dnmt3a directly interacts with *ZNF238* and promotes the repression of specific genes by histone deacetylation independently of its methyltransferase activity (Fuks*, et al.*, 2001). The overexpression of *ZNF238* could explain the downregulation of several genes in HPV-positive SCCHN as a consequence of chromatin condensation by histone deacetylation. *KLK8*, which is overexpressed in HPV-positive SCCHN, is a member of the human kallikrein gene family of serine proteases with a diverse physiologic function in many tissues. *KLK8* is implicated in terminal differentiation of keratinocytes and in the modulation of interaction between cells and fibronectin in extracellular matrix required for tumour growth and invasion. Several studies show different patterns of *KLK8* expression such as its downregulation in breast cancers, and in contrast, its overexpression in cervical cancers (Cane*, et al.*, 2004;Yousef*, et al.*, 2004). *CRABP2* is a transcription factor that functions as a cytoplasmic retinoic acid binding protein that is highly expressed in human skin. Similar to our results showing its downregulation in HPV-positive SCCHN, *CRABP2* is also downregulated in prostate cancer (Okuducu*, et al.*, 2005). Finally, in the unsupervised hierarchical cluster analysis, we were able to identify a group of genes in five different clusters with significant variation between HPV-positive SCCHN as compared to HPV-negative SCCHN as well as normal oropharyngeal epithelium (Fig.7). Some of the genes found to be differentially expressed in our statistical analysis were also found in these clusters such as *ZNF238, DNMT1, MCM3, AKR1C3, GRB10, TYK2* and *SART3*  (underlined in Fig. 7). The absence of other clustered genes in our supervised statistical analysis could be explained

by differences in expression levels below the 2-fold cut-off or by a value below the statistical significance that is considered by every test and overlapping processes. We also identified the localization of the various differentially expressed genes to particular chromosomes (Fig. 8). Two regions in chromosome 1 (p34-p36) and 12 (q21-q240) showed a significant number of differentially expressed genes. This may suggest common regulatory pathways involved in the regulation of some of these genes.

Our studies also identified genes whose expression is similarly affected in both HPV-positive and HPV-negative SCCHN as compared to the normal oropharyngeal tissue. For example, *STAT1* is overexpressed while *PCAF* is downregulated in both types of SCCHN. These genes may represent common pathways that are altered in oropharyngeal carcinogenesis. *STAT1* is a transcription factor that participates in cellular events such as IFN signaling, development of the mammary gland and embryogenesis. It is considered a tumour suppressor gene since its activation is associated with growth arrest, but studies in head and neck cancers showed an overexpression of *STAT1* in well or moderately differentiated tumours in vivo (Arany*, et al.*, 2003;Calo*, et al.*, 2003). Our findings contrast with the recent publication of Xi and colleagues showing the downregulation of *STAT1* in SCCHN through the hypermethylation of its promoter (Xi*, et al.*, 2006).

In summary, our data demonstrates specific changes in cellular gene expression profiles in HPV-positive SCCHN which also express the viral oncogenes. We have identified several genes such as those involved in nuclear structure and meiosis (*SYCP2*), cell differentiation (*CRABP2*), DNA repair (*RFC5*), transcription regulation (*ZNF238*) and epidermis development (*KLK8*) that were differentially expressed specifically in HPV-positive SCCHN as compared to both the HPV-negative SCCHN and normal oropharyngeal mucosa that could be used as potential biomarkers for the development of HPV-associated SCCHN.

**Chapter 3**


**GENE EXPRESSION PROFILES OF HUMAN PAPILLOMAVIRUS-POSITIVE**

**CERVICAL CELL LINES USING MICROARRAY ANALYSIS**

## 3.1 INTRODUCTION

Cervical cancer is the second most common cause of death in women worldwide after breast cancer (Jin, *et al.*, 1999). Cervical carcinogenesis involves two main types of epithelial transformation. The transformation of basal cells in the squamous epithelium starts with a noninvasive low-grade squamous intraepithelial lesion (LG-SIL), leading to a high-grade squamous intraepithelial lesion (HG-SIL), and ultimately evolving into invasive squamous cell carcinoma (SCC) (Alazawi, *et al.*, 2002). Adenocarcinomas and adenosquamous carcinomas (AC) originate from glandular endocervical columnar cells (Burk, *et al.*, 2003). Molecular and clinical evidence suggests a tight link between high-risk HPV infection and cervical cancer (zur Hausen, 2002). High-risk HPV DNAs (types 16, 18, 31, 33, 34, 35, 39, 45, 51, 52, 56, 58, 59, 66, 68, and 70) have been found in 99.7% of squamous cell cancers of the cervix (Walboomers, *et al.*, 1999). HPV infection is also implicated in approximately 25% of cases of oral cancers (Gillison, *et al.*, 2001). HPV-16 is the most common type associated with cervical cancer in the United States and Europe (50%), followed by types 18, 31 and 45 (Walboomers, *et al.*, 1999). Certain high-risk types such as HPV-16 and HPV-18 show differences in their oncogenic potential and also in their apoptotic stimuli sensitivity (Aguilar-Lemarroy, *et al.*, 2001;Badaracco, *et al.*, 2002). Also, HPV-16 is more commonly associated with SCC, whereas HPV-18 is predominantly associated with AC (Burk, *et al.*, 2003).

Three HPV proteins, E5, E6 and E7, have been implicated in cellular transformation. The HPV-16 E5 protein can stimulate cell growth via upregulation of the activity of growth factor receptors (McMurray, *et al.*, 2001). Recent studies have shown that E5 inhibits the induction of apoptosis in response to DNA damage (Zhang, *et al.*, 2002). E5 expression is likely to be important in the early steps of transformation since this gene is normally disrupted during viral integration (zur Hausen, 2002). The E6 proteins of high- risk HPVs can modify cellular growth regulatory pathways creating an environment that facilitates viral replication in differentiated cells (Syrjanen, *et al.*, 1999). E6 promotes the degradation of the tumor suppressor p53 protein involving the ubiquitin-protein ligase, upregulates telomerase activity, and affects transcriptional regulation by binding to the transcriptional activator p300/CBP (Klingelhutz, *et al.*, 1996;Patel, *et al.*, 1999;Thomas, *et al.*, 1999). There is also evidence that E6 protein can regulate transcription from minimal promoters containing a TATA box (Sedman, *et al.*, 1991). The E7 protein of high-risk HPVs interacts with the retinoblastoma tumor suppressor protein (pRB) and promotes degradation of this cell cycle regulatory protein (Dyson, *et al.*, 1989). The E7 protein disrupts the transcriptionally inactive pRB-E2F complex resulting in the upregulation of E2-responsive genes involved in cell cycle progression and entry into the S-phase (Muller, *et al.*, 2001;Wu, *et al.*, 1993). The HPV E7 protein also binds directly to histone deacetylase 1 (HDAC-1) and represses cellular transcription (Brehm, *et al.*, 1999). E7 also binds to the AP-1 transcription factor, interacts directly with

the TATA-binding protein (TBP) and the TBP-associated factor-110 (TAF-110) to modify transcriptional regulation, and reactivates DNA synthesis through its interaction with cyclin E-cdk2 (Antinore*, et al.*, 1996;Mazzarelli*, et al.*, 1995;McIntyre*, et al.*, 1996).

The HPV genome is usually integrated into the cellular genome in high-grade lesions and in cervical carcinomas (Durst*, et al.*, 1985). This generally results in the disruption of the viral E2 gene which is a negative regulator of the E6 and E7 oncogene expression (zur Hausen, 1991). However, E6 and E7 expression is also known to occur in cells containing episomal viral DNA (Bechtold*, et al.*, 2003). In other cases, the viral genome usually contains mutations in the viral long control region (LCR) that results in the upregulation of E6/E7 expression (May*, et al.*, 1994). This is consistent with recent studies showing that the presence of HPV-16 in an episomal state promotes chromosomal instability in an organotypic raft culture model (Duensing*, et al.*, 2001). This early events in cellular transformation which involve low-level expression of E6 and E7 in basal cells is sufficient to deregulate mitotic machinery and to produce abnormal number of centrosomes (Duensing*, et al.*, 2002). Chromosomal abnormalities such as tetraploidy or aneuploidy resulting from reduced mitotic fidelity are observed in HPV-positive cervical cell lines and cervical cancers (Southern*, et al.*, 2001). It has been hypothesized that such genomic instability may promote integration of the HPV genome (Kessis*, et al.*,

71

1996). These events also mimic DNA damage and may activate DNA repair and recombination processes.

Recently, studies on cellular gene expression profiles have been greatly facilitated by the development of high-density cDNA and oligonucleotide microarrays (Schena*, et al.*, 1996;Zhu*, et al.*, 1998). A few microarray studies have been done to analyze the global gene expression patterns of epithelial cells infected with HPV, such as HPV-transfected human keratinocytes, HPV-positive cervical carcinoma cell lines, snap-frozen HPV-positive cervical tumor biopsies, and early stage primary HPV-positive cervical cell lines (Ahn*, et al.*, 2004;Chang*, et al.*, 2000;Chen*, et al.*, 2003;Duffy*, et al.*, 2003;Lee*, et al.*, 2004;Santin*, et al.*, 2005). Due to the use of different types of microarray platforms and statistical analyses, there are considerable differences in the results of such studies. In the present study, we have used a high-density oligonucleotide microarray from Affymetrix (U133A) and two different types of statistical analyses to investigate the global gene expression profiles in HPV-16 as well as HPV-18 positive cervical cell lines containing either integrated or episomal forms of HPV DNA. Our results show dysregulation of genes involved in several cellular pathways such as cell cycle regulation, tumor suppression, cell adhesion, gene silencing by methylation and immune response that may play important roles in cervical carcinogenesis.

## 3.2    MATERIAL AND METHODS

*Cervical cell lines characteristics*

Cervical cancer cell lines CaSki and SiHa (containing HPV-16) as well as HeLa (HPV-18) and C-33A (HPV-negative) have been previously described (Meissner, 1999). Three clonal populations of the HPV-16 cervical cell line W12 (generated from a low-grade squamous intraepithelial lesion), 20863 (episomal HPV-16), 20861 and 201402 (integrated HPV-16) were obtained from the laboratories of Drs. Margaret Stanley and Paul Lambert (Medical Research Council, UK and University of Wisconsin, USA, respectively) (Jeon*, et al.*, 1995;Stanley*, et al.*, 1989). The L-18 cell line containing episomal HPV-18 was created by transfecting normal human foreskin keratinocytes with recircularized HPV-18 DNA and was obtained from Dr. Laimonis A. Laimins (Northwestern University, USA) (Frattini*, et al.*, 1997). The characteristics of various cell lines are described in Table 7. Cervical carcinoma cell lines were maintained in Dulbecco's modified Eagle medium (SiHa, HeLa and C-33A) or RPMI 1640 (CaSki) supplemented with 10% fetal bovine serum (FBS) at 37°C in the presence of 5% $CO_2$. Cell lines 201402, 20863, 20861 and L-18 were maintained in E medium with 10% FBS in the presence of murine 3T3 J2 fibroblast feeder cells previously treated with mitomycin C (Stanley*, et al.*, 1989). Before DNA or RNA isolation, fibroblast feeders were removed with versene (phosphate-buffered saline with 0.5 mM EDTA, Gibco BRL).

*DNA extraction and Southern blot analysis of the physical state of HPV-16 and HPV-18 DNA*

DNA isolation from cell lines 20863 (episomal HPV-16) and L-18 (episomal HPV-18) was performed by Hirt's method to enrich extrachromosomal DNA (Hirt, 1967). This procedure was also used with the cell line 20861 (integrated HPV-16) to maintain the same DNA extraction conditions. After DNA extraction, Southern blot analysis was performed as follows: 8 $\mu$g of DNA from each cell line was digested with a panel of restriction endonucleases (*Bgl*II, *Bam*HI, *Nco*I, and *Kpn*I for HPV-16 and *Bgl*II, *Eco*RV, *Nco*I and *Bam*HI for HPV-18), where the first enzyme does not cleave the HPV genome, the second enzyme cuts the DNA once and the third enzyme cuts the DNA twice. After electrophoresis through 0.7% agarose gels, the DNA was denatured and transferred to Gene Screen hybridization transfer membrane (PerkinElmer, Boston, MA). DNA was UV cross-linked to the membrane, and hybridization was performed at 68°C overnight using a $^{32}$P-labeled pUC19-HPV-16 or pUC19-HPV-18 plasmid as the probe. After hybridization, membranes were washed and subjected to autoradiography at -70°C (Maniatis*, et al.*, 1982).

# TABLE 7

**Cell lines characteristics**

| Cell line | Source | HPV genotype | Genomic physical state | Viral copy number |
|---|---|---|---|---|
| CaSki | Cervical carcinoma | 16 | integrated in tandem repeats | 60-600 |
| SiHa | Cervical carcinoma | 16 | integrated | 1-2 |
| 201402 | LG-SIL | 16 | integrated | 5 |
| 20863 | LG-SIL | 16 | episomal | 100 |
| 20861 | LG-SIL | 16 | integrated | 30 |
| HeLa | Cervical adenocarcinoma | 18 | integrated | 10-50 |
| L-18 | transfected keratinocytes | 18 | episomal | 100-200 |
| C-33A | Cervical carcinoma | - | - | - |

LG-SIL, low-grade squamous intraepithelial lesion

*RNA extraction*

Total RNA was isolated from monolayer cell cultures at approximately 80% confluence using ULTRASPEC RNA isolation system (Biotecx Lab, Inc. Houston, TX) according to the manufacturer's instructions. The RNA pellet was dried under a vacuum and resuspended in 30-50 µl of DEPC treated water.

*Northern blot analysis*

Northern blot analysis was carried out to identify transcripts encoded by HPVs in various cell lines. Ten micrograms of total RNA was separated on 1.2% denaturing formaldehyde-agarose gels and the RNA transferred to a Gene Screen Hybridization Transfer Membrane (PerkinElmer, Boston, MA). RNA was cross-linked to the membrane using UV light and hybridized to $^{32}$P-labeled, PCR-generated probes corresponding to the E2, E6 or E7 genes of HPV-16 or HPV-18 as well as probes for the genes LAT1 and NAP1L3. Hybridization was carried out overnight at 42°C (Maniatis*, et al.*, 1982). The membranes were subjected to autoradiography at -70°C.

*cRNA synthesis and microarray analysis*

Microarray analysis was carried out using the high-density oligonucleotide U133A GeneChip® expression microarray (Affymetrix, Santa Clara, CA) that

contains 22,215 transcript sets representing 14,820 human genes. RNA isolated from normal adult cervix from a pool of healthy women (Stratagene, La Jolla, CA) was used as a control in this study. For microarray analysis, total RNA isolated as described above was further purified by using the RNeasy Total RNA Isolation Kit (QIAGEN, Valencia, CA) to remove contaminating DNA. This kit is based on a silica-gel column that specifically binds single-stranded RNA of more than 200 nucleotides. Twenty micrograms of total RNA from each sample were used to generate double-stranded [ds] cDNA using the Superscript II Choice system (GIBCO-BRL, Rockville, MD). First-strand synthesis was carried out using a T7-$(dT)_{24}$ primer (Sigma-GenoSys, The Woodlands, TX) and Superscript II reverse transcriptase. This primer includes the promoter sequence for the T7 RNA polymerase. The second-strand synthesis was performed using RNase H and DNA polymerase I. The resulting ds cDNA was used to synthesize cRNA using biotin-labeled ribonucleotides and T7 RNA polymerase utilizing the BioArray HighYield RNA transcript labeling kit (ENZO Life Sciences, Inc. Farmingdale, NY). The cRNA was fragmented to 35-200 bases length and hybridized to the Human Genome U133A GeneChip® expression microarray in the Affymetrix Hybridization Oven (Affymetrix, Santa Clara, CA). The GeneChips were washed, and then stained with streptavidin-phycoerythrin (SAPE) to generate fluorescent signals from biotin-labeled cRNA. The GeneChips were scanned using an Affymetrix GeneArray® scanner (Affymetrix, Santa Clara, CA). Each probed array was scanned twice to calculate an average of two images, define the probe cells and compute intensity for each cell.

*Microarray data analysis*

The .CEL image files obtained from the Affymetrix Microarray Suite 5.0 were converted to .DCP files using the program dChip version 1.2 (www.dchip.org). The dChip program was used to generate normalized and non-normalized intensity signal data. Two different normalizations and statistical tests were used to analyze all the microarray data in order to identify differentially expressed genes that were robust enough to appear in both analysis: Non-normalized intensity signal data from dChip were normalized (log-base2 transformation, median within arrays, Global Mean Adjustment among arrays) and analyzed with a pooled variance t-Test at level alpha = 0.05 using the Gene Expression Data Analysis Tool (GEDA) program developed in the University of Pittsburgh (James Lyons-Weiler, University of Pittsburgh Cancer Institute, http://bioinformatics.upmc.edu) (Patel*, et al.*, 2004). In the second analysis, normalized intensity signal data from dChip were used directly in the Significance Analysis of Microarray (SAM) program version 1.21 (www.stat.stanford.edu/~tibs/SAM/) to identify differentially expressed genes. Genes with an expression ratio of at least 2-fold difference relative to the controls were considered differentially expressed in both analyses (in the SAM-Test for episomal versus integrated HPV-containing cell lines we used a 1.9-fold cut-off). After analyzing the data with the two statistical tests, we overlapped the lists of

differentially expressed genes in order to generate a list of differentially expressed genes that were expected to be more robust.

*Quantitative real-time RT-PCR analysis*

We validated the microarray data by quantitative real-time RT-PCR (QRT-PCR) analysis of a few representative genes using the QuantiTect SYBR Green one-step RT-PCR kit (QIAGEN, Valencia, CA).  Before cDNA synthesis, RNA was treated with amplification grade DNase I (Invitrogen, Rockville, MD) for 15 min at room temperature to avoid DNA contamination. Approximately 400 ng of DNase I-treated total RNA from cell lines or the normal cervix was mixed with 25 $\mu$l of 2 X QuantiTect SYBR Green RT-PCR Master Mix, 0.5 $\mu$M of specific forward and reverse primers and RNase-free water in a final volume of 50 $\mu$l. The amount of fluorescence emitted by SYBR Green I was quantified by the ABI Prism 7700 system software (Applied Biosystems, Foster City, CA). The following genes and primers were used in these studies:

transcription factor AP-2 alpha (TFAP2A), forward primer: 5'-ATTTGGGACCACCTGGTATT -3', reverse primer: 5'-TGATACATTGACAGGCATGG- 3'; carboxypeptidase A4 (CPA4), forward primer: 5'- CTTCCCTCCTCTTTGCCTTT -3', reverse primer: 5'-CAGGATACCCCAGCACAAAT -3'; neurofilament, heavy polypeptide 200kDa (NEFH), forward primer: 5'-TGCCTTCTGTTTTCCAAAGG-3', reverse primer: 5'-TGAACTCCAGTGGAAAGTGC-3'; solute carrier family 7 member 5 (LAT1),

forward primer: 5'-ACCCACCACAACAAGCAAGT-3', reverse primer: 5'-CTGAGCATGACCACTGGAAA-3'; sprouty homolog 1, antagonist of FGF signaling (SPRY1), forward primer: 5'- GACCTTGCCCATCTTCACTC-3', reverse primer: 5'-ATCCAAGTGCCCTCAATGAC-3'; inhibitor of DNA binding 4, dominant negative helix-loop helix protein (ID4), forward primer: 5'-TGCTTGCTACCAAAGGACAA-3', reverse primer: 5'-ATCCCTGCTTGAAGGATTTG-3'; immunoglobulin superfamily, member 4 (IGSF4), forward primer: 5'- AGCAGACGCAGACACAGCTA -3', reverse primer: 5'- TGGACACCTCATTGAAACAAA -3'; nucleosome assembly protein 1-like 3 (NAP1L3), forward primer: 5'- TTCCAGCCCATAAGACTAAGG -3', reverse primer: 5'- TTTCAAGCAGTCAAACTTCAAAA -3'; The following RT-PCR cycle parameters were used: reverse transcription at 50°C for 30 min, hot-start DNA polymerase activation 95°C for 15 min, 40 cycles of denaturation at 94°C for 15 sec each, annealing at 58°C for 30 sec and extension at 72°C for 30 sec. Each reaction was run in triplicate in a 96-well plate. Relative expression of the target gene was calculated using the 2 delta CT method described previously:(Relative expression = $2^{-\Delta CT}$; where $\Delta CT$ = $C_{T \ (Target \ gene)}$ − $C_{T \ (endogenous \ control \ gene)}$) where glyceraldehyde-3-phosphate dehydrogenase (GAPDH) was used as the endogenous control gene (Livak, *et al.*, 2001).

*In situ hybridization using tissue arrays*

We detected the expression of NAP1L3 and LAT1 by in situ hybridization in tissue-arrays containing ~50 different cervical tumor samples and 4 normal cervix controls per slide (4 mm x 1 mm diameter, 4 $\mu$m thick per spot) (IMGENEX, Cat #: IMH-368). Individual information for all the tissue samples in the arrays is available on the web page: http://www.imgenex.com/tissue_array_tds.php?id=929. Briefly, we dewaxed the paraffin with xylene three times, 10 min each. Slides were hydrated sequentially in 100%, 95%, 80%, 70%, and 60% ethanol and water unitl the sections were clear. Tissue-arrays were then permeabilized with 0.2 M HCl for 20 min at RT, followed by 0.3% triton X-100 in PBS for 15 min at RT and 17 $\mu$g/ml proteinase K in 0.1 MTris, pH 7.5, and 50 mM ethylenediaminetetraacetic acid (EDTA) for 30 min at 37°C. Tissues were rinsed in 0.2% glycine, pH 7.5, at RT and postfixed with 4% paraformaldehyde for 5 min and rinsed with 1X PBS. To decrease nonspecific background, sections were acetylated with 0.25% acetic anhydride in 0.1 M triethanolamine for 10 min. Tissues were prehybridized in 50% deionized formamide, 2 X SSC, 10% dextran sulfate, 250 $\mu$g/ml herring sperm DNA, and 0.5% sodium dodecyl sulfate (SDS) for 2 hr at 37°C (using a coverslip to reduce evaporation) and hybridized in the same solution with ~2 ng/ml of biotinylated probe overnight. The following day, slides were immersed in 4X SSC and coverslip removed. Tissue-arrays were washed twice at RT in 2X SSC for 15 min each, followed by a quick rinse with ELF mRNA Wash Buffer (ELF[®] 97 mRNA In

Situ Hybridization kit # 2, Cat # E-6605, Molecular Probes, Inc.). Blocking was done with ELF-97 mRNA Blocking Buffer for 30 min, followed by incubation with 10 μg/ml streptavidin-alkaline phosphatase in the same buffer for 15 min at RT. Samples were washed with ELF-97 mRNA Wash Buffer, equilibrated in ELF-97 mRNA Developing Buffer for 5 min, and incubated with 1:10 dilution of ELF-97 substrate with a 1:1000 dilution of substrate additives 1 and 2. The probes for LAT1 and NAP1L3 were single-stranded oligonucleotides (Reverse primers from the QRT-PCR method) labeled at the 3'-end with biotin-ddUTP according to the manufacturer's instructions (DIG Oligonucleotide 3'-End Labeling Kit (Cat # 3353575, Roche Applied Science).

## 3.3    RESULTS

*Southern blot analysis of the physical state of HPV-16 and HPV-18 DNA*

Prior to carrying out microarray analysis, we confirmed the physical state of the HPV DNA in various cell lines by Southern blot analysis.  In Fig. 10, predictable bands of ~7.9 kb representing linearized full length episomes were obtained using HPV-16 single cutter restriction enzymes *Bam*HI or *Nco*I and HPV-18 single cutters *Eco*RV or *Nco*I in the episomal 20863 (passage 35) and L-18 (passage 17) cell lines, respectively. Digestion of DNA from the 20861 cell line (passage 16) with the same single cutter enzymes showed different sized bands, consistent with this cell line containing integrated HPV DNA. Similarly,

HPV-16 and HPV-18 double cutters *Kpn*I and *Bam*HI generated two predicted

bands in 20863 and L-18 cell lines, but not in the integrated cell line 20861. The

HPV-16 and HPV-18 noncutter *Bgl*II enzyme showed two bands in the episomal

cell lines, corresponding to the supercoiled and open-circular forms of the DNA

(Fig. 10). By Southern blot analysis, we also confirmed that the HPV DNA in

CaSki, SiHa and HeLa cell lines was integrated (not shown).


*Expression of the HPV-16 and HPV-18 early genes in various cell lines*


Northern blot analysis was carried out to confirm the expression of HPV

early transcripts in all the cell lines and the control cervical sample. As expected

and consistent with previous data, the E6 and E7 viral oncogene transcripts (with

different splicing variants) were present in all the cell lines containing integrated

HPV sequences but not in the HPV-negative cell line C-33A or the normal cervix

(Fig. 11) (Baker*, et al.*, 1987). Interestingly, the 20863 and L-18 cell lines that

contain episomal HPV DNAs also expressed significant levels of the E6 and E7

transcripts (Fig. 11). E2 transcripts were present in cell lines 20863 and L-18

containing episomal HPVs DNA as well as in the HPV-16 head to tail tandem

integrated cell line, CaSki (Fig. 11). Interestingly, we found robust levels of a

larger than expected E2 transcript in the 20861 cell line containing integrated

HPV-16 DNA (Fig. 11). This larger E2 transcript presumably results from the

disruption of the E2 gene and its fusion to a cellular transcript. As expected, the

SiHa, 201402 and HeLa cell lines in which the E2 gene is disrupted did not show

# Fig. 10.   PHYSICAL STATE OF HPV DNA IN CELL LINES
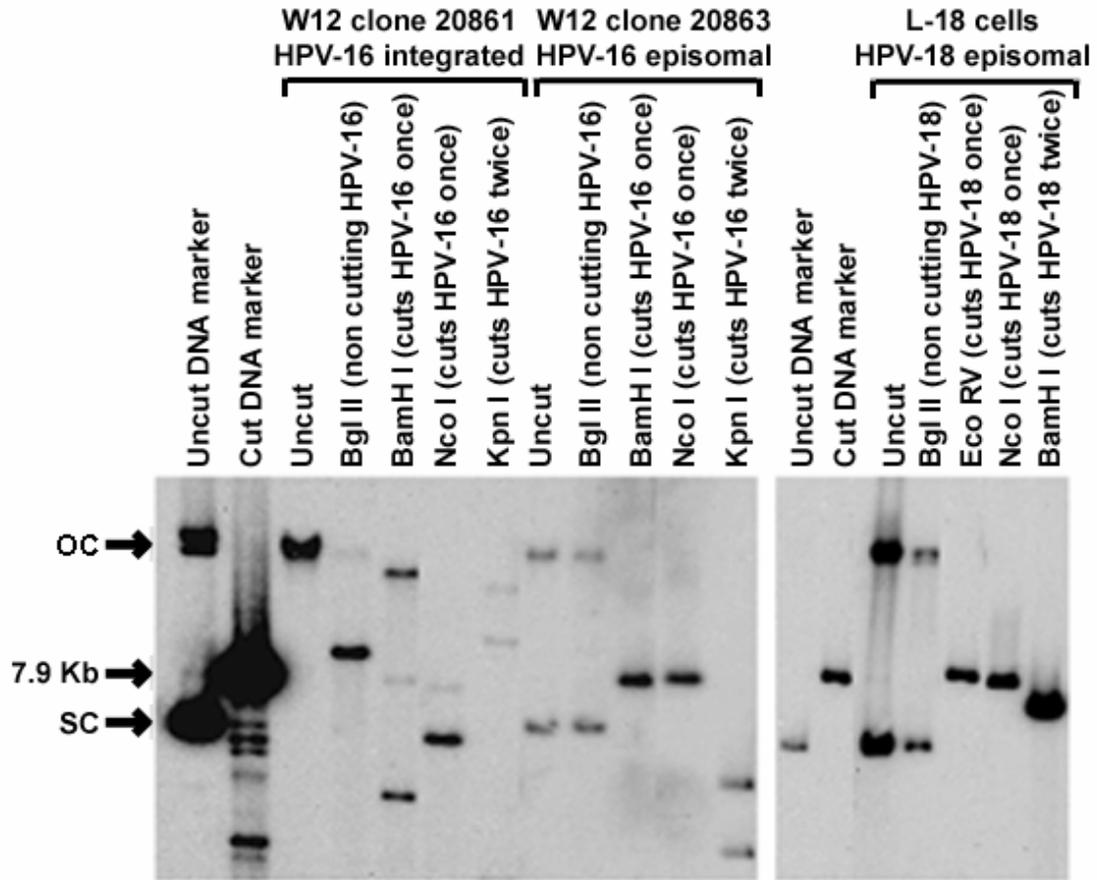


Physical state of HPV DNA in cell lines, as determined by Southern blot analysis. Uncut and cut DNAs from a 7.9 kb plasmid were used as size markers. SC, supercoiled DNA; OC, open-circular DNA.

any E2 transcript (Fig. 11). The above results confirmed the appropriate expression of the HPV early genes in various cell lines, and RNA isolated from these cell lines was used for microarray analysis of cellular gene expression profiles.

*Microarray analysis of genes differentially expressed in HPV-positive cell lines compared to the normal cervix*

In order to identify differentially expressed genes in HPV-16 positive (CaSki, SiHa, 201402, 20861, and 20863) and HPV-18 (HeLa and L-18) cervical cell lines compared to the normal cervix tissue, we used the Affymetrix oligonucleotide U133A GeneChip® that contains 22,215 transcript sets that represent 14,820 human genes. Normalization and differential gene expression analyses were carried out as described in material and methods. The microarray experiments were performed in duplicate (with the exception of the 20861 cell line) and only genes whose expression level was increased or decreased by more than 2-fold were selected. A total of 2,315 and 1,908 genes showed up- and down-regulation, respectively, using the pooled variance t-Test analysis when all the HPV-positive cervical cell lines were compared to the normal cervix (supplementary data, Table L). On the other hand, 544 and 1,634 genes were found to be up- and down-regulated, respectively, using the SAM-Test analysis (supplementary data, Table M). When the results of the above two statistical

# Fig. 11.   NORTHERN BLOT ANALYSIS OF THE HPV E6, E7 AND E2 GENES IN HPV-POSITIVE CELL LINES
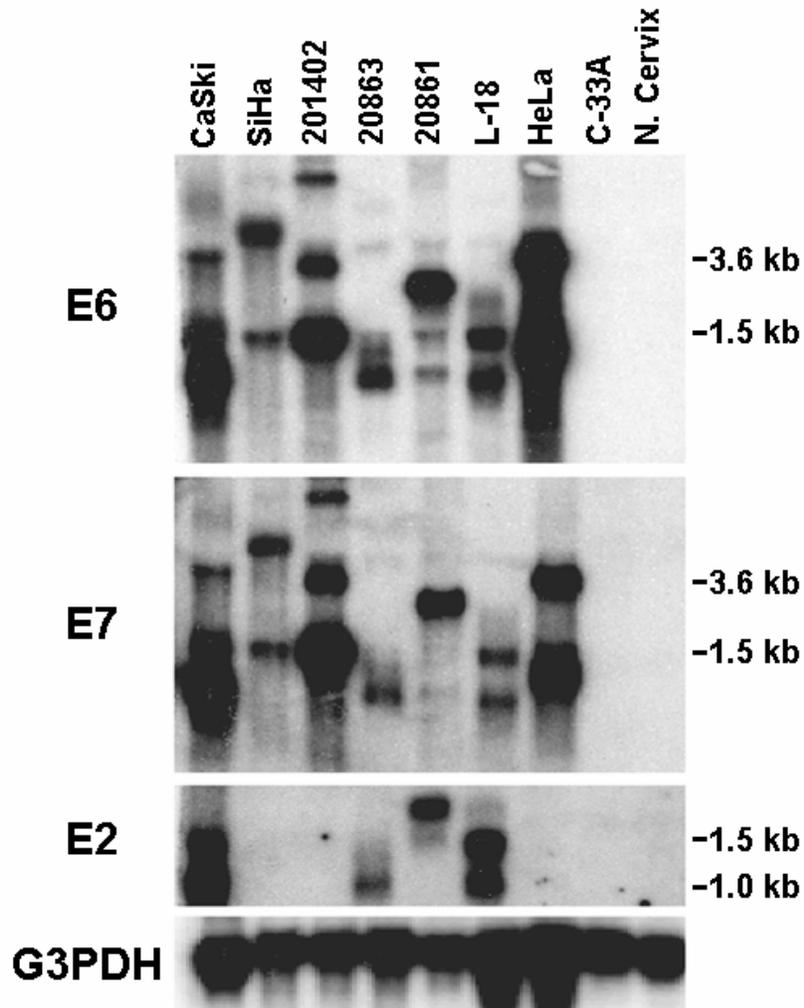


Hybridization probes are indicated on the far left. The gene glyceraldehyde-3-phosphate dehydrogenase (G3PDH) was used as the endogenous control gene. Approximate sizes of the various bands are shown on the right.

tests were overlapped, 272 genes were found to be upregulated and 605 downregulated in HPV-positive cervical cell lines as compared to the normal cervix (Fig. 12). A representative list of differentially expressed genes is shown in Table 8 (a complete list of genes is shown in the supplementary data, Table N). Our analysis identified genes belonging to various cellular pathways such as cell cycle, DNA replication, repair and recombination, signal transduction, regulation of transcription, cell adhesion and immune response whose expression was altered in HPV-positive cells. Our experiments also identified several genes that were previously implicated in carcinogenesis in different tissues but not in cervical cancer (Table 8) (Imai*, et al.*, 2005;Walter*, et al.*, 2003). Some genes upregulated in HPV-positive cell lines were aurora kinase B (AURKB), flap structure-specific endonuclease 1 (FEN1), member of RAS oncogene family RAB32 (RAB32), carboxypeptidase A4 (CPA4), 8-oxoguanine DNA glycosylase (OGG1) and Rab9 effector p40 (RAB9P40), while the downregulated genes included claudin 10 (CLDN10), Cullin 5 (CUL5), inhibitor of DNA binding 4, dominant negative helix-loop-helix protein (ID4), deleted in malignant brain tumors 1 (DMBT1), methyl-CpG binding domain protein 4 (MBD4) and nidogen (NID) (Table 8). Our results also identified several differentially expressed genes that were also found in previous studies, suggesting a concordance and validation of our results. Some such upregulated genes included cyclin B1 (CCNB1), cyclin-dependent kinase inhibitor 2A (CDKN2A), minichromosome maintenance deficient 7 (MCM7), transcription factor AP-2 alpha (TFAP2A),

**Fig. 12.   Venn diagram showing the number of differentially expressed genes that overlap between the two statistical analyses (pooled t-test and SAM-test) of the microarray data.**



ALL HPV POSITIVE CELL LINES
(CaSki, SiHa, 201402, 20863,20861,
HeLa, L-18)

VS

NORMAL CERVIX

Upregulated genes 272
Downregulated genes 605
(after pooled t-Test and SAM-Test
overlapping)

Overlap data from
both analyses
Upreg. genes 20
Downreg. genes 37

ALL HPV POSITIVE CELL LINES
(CaSki, SiHa, 201402, 20863, 20861,
HeLa, L-18)

VS

C-33A
(HPV NEGATIVE CELL LINE)

Upregulated genes 305
Downregulated genes 232
(after pooled t-Test and SAM-Test
overlapping)

**The two principal sets indicate differentially expressed genes between all HPV-positive cell lines and either normal cervix tissue or the HPV-negative cell line C-33A. The overlap region shows the number of differentially expressed genes in all HPV-positive cell lines as compared to both the controls.**

while the list of downregulated genes included immunoglobulin superfamily, member 4 (IGSF4), fibronectin (FN1), and interferon alpha-inducible protein 27 (IFI27) (Table 8) (Beger*, et al.*, 2001;Li*, et al.*, 2005;Santin*, et al.*, 2005;Steenbergen*, et al.*, 2004;Tan*, et al.*, 2003).

*Differential gene expression profiles in HPV-positive cell lines compared to the HPV-negative cell line C-33A*

We compared the gene expression profiles of all HPV-positive cervical cell lines with that of the HPV-negative cell line C-33A in order to identify cellular genes whose expression may be specifically altered due to the presence of HPV DNA (and not due to cellular immortalization). Using the pooled variance t-Test analysis, we found that 900 genes were upregulated 33A (supplementary data, Table O). and 1,342 downregulated in HPV-positive cell lines as compared to C-The SAM test identified 641 upregulated and 561 downregulated genes in the HPV-positive cell lines (supplementary data, Table P). A comparison of the differentially expressed genes revealed that 305 upregulated and 232 downregulated genes were common in t-Test and SAM analyses (supplementary data, Table Q). A representative list of such genes is shown in Table 9. Example of novel genes found to be differentially expressed in HPV-positive cervical cell lines compared to the HPV-negative cervical cell line C-33A as a control are: genes involved in metastasis (actinin, alpha 4 [ACTN4]), antiestrogen resistance

# Table 8. DIFFERENTIALLY EXPRESSED GENES IN HPV-POSITIVE CELL LINES COMPARED TO THE NORMAL CERVIX

| Gene Title | Gene Symbol | pooled t-Test | SAM-Test |
|---|---|---|---|
| **Upregulated genes** | | | |
| **Apoptosis** | | | |
| baculoviral IAP repeat-containing 5 (survivin) | BIRC5 | 6.33 | 3.38 |
| BCL2-antagonist of cell death | BAD | 2.43 | 2.12 |
| **Cell cycle** | | | |
| aurora kinase B | AURKB | 6.51 | 3.84 |
| cyclin B1 | CCNB1 | 5.88 | 4.52 |
| cyclin-dependent kinase inhibitor 2A | CDKN2A | 3.95 | 2.02 |
| DNA replication factor | CDT1 | 4.24 | 5.30 |
| MCM7 | MCM7 | 4.36 | 2.64 |
| **Cell proliferation** | | | |
| MAD1 mitotic arrest deficient-like 1 (yeast) | MAD1L1 | 2.50 | 2.15 |
| SHC (Src homology 2 domain containing) transforming protein 1 | SHC1 | 3.55 | 2.27 |
| ubiquitin-conjugating enzyme E2C | UBE2C | 4.68 | 2.40 |
| WD40 protein Ciao1 | CIAO1 | 3.60 | 2.32 |
| **Cell transport** | | | |
| adaptor-related protein complex 1, sigma 1 subunit | AP1S1 | 4.38 | 2.08 |
| metaxin 1 | MTX1 | 3.42 | 2.26 |
| Rab9 effector p40 | RAB9P40 | 2.49 | 2.11 |
| **Development and cell differentiation** | | | |
| pescadillo homolog 1, containing BRCT domain (zebrafish) | PES1 | 3.10 | 2.32 |
| **DNA repair** | | | |
| 8-oxoguanine DNA glycosylase | OGG1 | 2.31 | 2.04 |
| CHK1 checkpoint homolog (S. pombe) | CHEK1 | 3.46 | 2.56 |
| flap structure-specific endonuclease 1 | FEN1 | 5.32 | 3.27 |
| H2A histone family, member X | H2AFX | 3.76 | 2.33 |
| **Nuclear organization and biogenesis** | | | |
| carboxypeptidase A4 | CPA4 | 2.92 | 3.14 |
| **Regulation of transcription** | | | |
| Cbp/p300-interacting transactivator | CITED2 | 3.30 | 2.50 |
| cofactor of BRCA1 | COBRA1 | 3.90 | 2.17 |
| transcription factor AP-2 alpha | TFAP2A | 4.30 | 2.86 |
| **Signal transduction** | | | |

| | | | |
|---|---|---|---|
| HMT1 hnRNP methyltransferase-like 2 | HRMT1L2 | 4.71 | 2.24 |
| RAB32, member RAS oncogene family | RAB32 | 3.89 | 2.43 |
| Ran GTPase activating protein 1 | RANGAP1 | 5.93 | 2.53 |
| v-Ha-ras | HRAS | 2.62 | 2.67 |
| **Structural function** | | | |
| neurofilament, heavy polypeptide 200kDa | NEFH | 4.47 | 12.72 |
| tubulin beta MGC4083 | MGC4083 | 4.05 | 2.46 |
| ZW10 interactor | ZWINT | 3.52 | 2.08 |
| **Downregulated genes** | | | |
| **Cell adhesion** | | | |
| claudin 10 | CLDN10 | -15.61 | 0.175 |
| fibronectin 1 | FN1 | -5.31 | 0.467 |
| mucin 4, tracheobronchial | MUC4 | -7.78 | 0.273 |
| nidogen (entactin) | NID | -3.99 | 0.097 |
| **Cell cycle** | | | |
| cullin 5 | CUL5 | -2.14 | 0.479 |
| cylindromatosis (turban tumor syndrome) | CYLD | -4.09 | 0.484 |
| microtubule-actin crosslinking factor 1 | MACF1 | -5.15 | 0.444 |
| p300/CBP-associated factor | PCAF | -3.03 | 0.383 |
| Wilms tumor 1 | WT1 | -3.06 | 0.220 |
| **Cell differentation** | | | |
| collagen, type I, alpha 1 | COL1A1 | -7.51 | 0.066 |
| deleted in malignant brain tumors 1 | DMBT1 | -7.86 | 0.261 |
| mal, T-cell differentiation protein | MAL | -7.38 | 0.101 |
| Notch homolog 2 (Drosophila) | NOTCH2 | -2.32 | 0.385 |
| **DNA repair** | | | |
| methyl-CpG binding domain protein 4 | MBD4 | -2.18 | 0.426 |
| mutS homolog 3 (E. coli) | MSH3 | -3.35 | 0.300 |
| REV3-like, catalytic subunit of DNA polymerase zeta (yeast) | REV3L | -3.27 | 0.250 |
| **Immune response** | | | |
| immunoglobulin superfamily, member 4 | IGSF4 | -9.16 | 0.410 |
| interferon, alpha-inducible protein 27 | IFI27 | -4.21 | 0.076 |
| interleukin 1 receptor, type II | IL1R2 | -4.46 | 0.392 |
| major histocompatibility complex, class II, DQ beta 1 | HLA-DQB1 | -6.26 | 0.149 |
| **Protein modification** | | | |
| cathepsin B | CTSB | -2.95 | 0.479 |
| P21 (CDKN1A)-activated kinase 3 | PAK3 | -2.96 | 0.363 |
| **Regulation of transcription** | | | |
| activating transcription factor 7 interacting protein | ATF7IP | -2.89 | 0.429 |

| | | | |
|---|---|---|---|
| Inhibitor of DNA binding 4 | ID4 | -3.94 | 0.122 |
| SWI/SNF related, matrix associated, actin dependent regulator of chromatin, subfamily a, member 2 | SMARCA2 | -3.44 | 0.392 |
| zinc finger protein 198 | ZNF198 | -2.83 | 0.353 |
| **RNA binding** | | | |
| Dicer1, Dcr-1 homolog (Drosophila) | DICER1 | -3.17 | 0.307 |
| splicing factor, arginine/serine-rich 7, 35kDa | SFRS7 | -2.45 | 0.377 |
| zinc finger protein 638 | ZNF638 | -2.60 | 0.443 |
| **Structural function** | | | |
| SMC5 structural maintenance of chromosomes 5 like 1 (yeast) | SMC5L1 | -2.38 | 0.411 |

(breast cancer anti-estrogen resistance 3 [BCAR3]), mitogenic signaling (caveolin 1 [CAV1], caveolin 2 [CAV2]), angiogenesis (endothelial PAS domain protein 1 [EPAS1]) and double-strand DNA break repair via nonhomologous end-joining (MRE11, meiotic recombination 11 homolog A [MRE11A]).

Down-regulated genes found in these analyses were those involved in transcriptional regulation (CBP/p300-interacting transactivator, with Glu/Asp-rich carboxy-terminal domain, 2 [CITED2]), apoptosis (DNA fragmentation factor, 40kDa, beta polypeptide [DFFB]), cellular morphogenesis (myosin, heavy polypeptide 10, non-muscle [MYH10]), centrosome formation (pericentriolar material 1 [PCM1]) and mismatch repair (postmeiotic segregation increased 1 [PMS1]) (Table 9).

*Genes differentially expressed in HPV-positive cell lines compared to both C-33A cells and the normal cervix*

Results described above identified genes that are differentially expressed in HPV-positive cell lines as compared to the normal cervix or the HPV-negative cell line C-33A utilizing two different statistical analyses. To identify cellular genes whose expression might be specifically altered due to the presence of HPV sequences, we compared the data obtained for HPV-positive cell lines with those of both the normal cervix and the C-33A cell line.  This analysis revealed that a total of 20 genes were upregulated and 37 downregulated in the HPV-

## Table 9.  DIFFERENTIALLY EXPRESSED GENES IN HPV-POSITIVE CELL LINES COMPARED TO HPV-NEGATIVE C-33A

| Gene Title | Gene Symbol | pooled t-Test | SAM-Test |
|---|---|---|---|
| **Upregulated genes** | | | |
| **Apoptosis** | | | |
| lectin, galactoside-binding, soluble, 1 (galectin 1) | LGALS1 | 8.908 | 9.67513 |
| **Cell adhesion** | | | |
| laminin, beta 1 | LAMB1 | 4.694 | 6.82606 |
| tenascin C (hexabrachion) | TNC | 2.728 | 3.08323 |
| **Cell cycle** | | | |
| breast cancer anti-estrogen resistance 3 | BCAR3 | 4.631 | 3.37527 |
| cyclin-dependent kinase inhibitor 1A (p21, Cip1) | CDKN1A | 6.212 | 10.70517 |
| polo-like kinase 2 (Drosophila) | PLK2 | 7.372 | 20.19033 |
| **Cell proliferation and angiogenesis** | | | |
| endothelial PAS domain protein 1 | EPAS1 | 9.146 | 15.6005 |
| insulin-like growth factor binding protein 7 | IGFBP7 | 10.765 | 10.3496 |
| jagged 2 | JAG2 | 3.195 | 3.39671 |
| S100 calcium binding protein A11 (calgizzarin) | S100A11 | 15.295 | 54.11594 |
| **Development and cell differentiation** | | | |
| mesoderm specific transcript homolog (mouse) | MEST | 8.928 | 14.46237 |
| **DNA repair** | | | |
| growth arrest and DNA-damage-inducible, beta | GADD45B | 5.008 | 3.73185 |
| MRE11 meiotic recombination 11 homolog A (S. cerevisiae) | MRE11A | 2.119 | 2.43135 |
| **Immune response** | | | |
| interferon, gamma-inducible protein 30 | IFI30 | 5.345 | 3.23202 |
| major histocompatibility complex, class I, E | HLA-E | 2.023 | 2.0984 |
| **Nuclear organization and biogenesis** | | | |
| carboxypeptidase A4 | CPA4 | 2.839 | 3.27686 |
| neurofilament, heavy polypeptide 200kDa | NEFH | 2.078 | 3.29482 |
| **Regulation of transcription** | | | |
| CCAAT/enhancer binding protein (C/EBP), beta | CEBPB | 3.737 | 2.44182 |
| HIV-1 Tat interactive protein 2, 30kDa | HTATIP2 | 10.801 | 3.63548 |
| v-fos FBJ murine osteosarcoma viral oncogene homolog | FOS | 3.409 | 3.11517 |
| v-jun sarcoma virus 17 oncogene homolog (avian) | JUN | 2.313 | 2.15402 |
| **Signal transduction** | | | |
| mitogen-activated protein kinase kinase kinase 5 | MAP3K5 | 3.399 | 2.30119 |

| plasminogen activator, urokinase receptor | PLAUR | 2.312 | 3.59171 |
|---|---|---|---|
| **Structural function** | | | |
| actinin, alpha 4 | ACTN4 | 3.032 | 2.69455 |
| caveolin 1, caveolae protein, 22kDa | CAV1 | 5.406 | 16.81007 |
| caveolin 2 | CAV2 | 7.068 | 12.5958 |
| CDC42 effector protein (Rho GTPase binding) 4 | CDC42EP4 | 7.203 | 2.68908 |
| **Transport** | | | |
| aquaporin 3 | AQP3 | 2.675 | 17.36881 |
| Ras and Rab interactor 2 | RIN2 | 5.538 | 15.32698 |
| Sortilin-related receptor, L(DLR class) A repeats containing | SORL1 | 3.558 | 5.30152 |

| **Downregulated genes** | | | |
|---|---|---|---|
| **Apoptosis** | | | |
| BCL2/adenovirus E1B 19kDa interacting protein 3 | BNIP3 | -2.31 | 0.47945 |
| caspase 2, apoptosis-related cysteine protease | CASP2 | -2.896 | 0.41099 |
| DNA fragmentation factor, 40kDa, beta polypeptide (caspase-activated DNase) | DFFB | -3.427 | 0.46688 |
| programmed cell death 8 (apoptosis-inducing factor) | PDCD8 | -6.32 | 0.27251 |
| **Cell cycle** | | | |
| cell division cycle 34 | CDC34 | -4.917 | 0.44664 |
| cyclin-dependent kinase 5 | CDK5 | -3.499 | 0.49566 |
| **Development and cell differentiation** | | | |
| homeo box B2 | HOXB2 | -4.604 | 0.23615 |
| homeo box C6 | HOXC6 | -3.214 | 0.35205 |
| myosin, heavy polypeptide 10, non-muscle | MYH10 | -2.339 | 0.39779 |
| tumor protein D52 | TPD52 | -2.826 | 0.43865 |
| **DNA repair** | | | |
| RuvB-like 2 (E. coli) | RUVBL2 | -2.692 | 0.44269 |
| RAD51 homolog C (S. cerevisiae) | RAD51C | -2.333 | 0.45786 |
| RAD51-like 1 (S. cerevisiae) | RAD51L1 | -8.702 | 0.32049 |
| postmeiotic segregation increased 1 (S. cerevisiae) | PMS1 | -2.701 | 0.48804 |
| **Immune response** | | | |
| regulatory factor X-associated ankyrin-containing protein | RFXANK | -2.829 | 0.4833 |
| G antigens 8, 4, 5, 7, 2, 6, 7B | GAGE | -2.759 | 0.44522 |
| transcription factor 8 (represses interleukin 2 expression) | TCF8 | -4.298 | 0.30337 |
| **Metabolism** | | | |
| aldehyde dehydrogenase 7 family, member A1 | ALDH7A1 | -2.848 | 0.22177 |
| **Nuclear organization and biogenesis** | | | |

| | | | |
|---|---|---|---|
| chromobox homolog 1 (HP1 beta homolog Drosophila) | CBX1 | -2.387 | 0.48223 |
| HMT1 hnRNP methyltransferase-like 2 (S. cerevisiae) | HRMT1L2 | -3.007 | 0.46137 |
| nucleosome assembly protein 1 like 3 | NAP1L3 | -6.355 | 0.40808 |
| **Regulation of transcription** | | | |
| Cbp/p300-interacting transactivator, with Glu/Asp rich carboxy-terminal domain, 2 | CITED2 | -5.39 | 0.26298 |
| chromobox homolog 6 | CBX6 | -3.153 | 0.42101 |
| nuclear receptor subfamily 2, group F, member 1 | NR2F1 | -2.843 | 0.39402 |
| pericentriolar material 1 | PCM1 | -3.163 | 0.41163 |
| TAF12 RNA polymerase II, TATA box binding protein (TBP)-associated factor, 20kDa | TAF12 | -6.651 | 0.47371 |
| **Signal transduction** | | | |
| gap junction protein, alpha 1, 43kDa (connexin 43) | GJA1 | -2.278 | 0.14202 |
| sprouty homolog 1, antagonist of FGF signaling (Drosophila) | SPRY1 | -13.228 | 0.13179 |
| **Transport** | | | |
| ATPase, H+ transporting, lysosomal 14kDa, V1 subunit F | ATP6V1F | -2.619 | 0.48982 |
| heat shock 60kDa protein 1 (chaperonin) | HSPD1 | -2.043 | 0.46505 |

**Table 10.** DIFFERENTIALLY EXPRESSED GENES IN HPV-POSITIVE CELL LINES COMPARED TO BOTH THE NORMAL CERVIX AND C-33A

| Gene Title | Gene Symbol | pooled t-Test normal cervix | SAM-Test normal cervix | pooled t-Test C-33A | SAM-Test C-33A |
|---|---|---|---|---|---|
| **Upregulated genes** | | | | | |
| **Apoptosis** | | | | | |
| serine/threonine kinase 17a (apoptosis-inducing) | STK17A | 2.797 | 2.5915 | 3.341 | 3.45054 |
| **Cell cycle** | | | | | |
| breast cancer anti-estrogen resistance 3 | BCAR3 | 3.5 | 2.44826 | 4.631 | 3.37527 |
| **Development and cell differentiation** | | | | | |
| neurofilament, heavy polypeptide 200kDa | NEFH | 4.474 | 12.72799 | 2.078 | 3.29482 |
| transcription factor AP-2 alpha | TFAP2A | 4.302 | 2.86491 | 7.049 | 8.32077 |
| tuftelin 1 | TUFT1 | 2.817 | 2.39515 | 3.642 | 3.3442 |
| **Functionally unclassified genes** | | | | | |
| DNA-damage-inducible transcript 4 (RTP801, REDD1) | DDIT4 | 2.952 | 3.27046 | 2.909 | 4.15816 |
| kinesin family member C3 | KIFC3 | 3.653 | 2.08519 | 5.539 | 3.21941 |
| **Immune response** | | | | | |
| prostaglandin E synthase | PTGES | 2.79 | 3.68109 | 4.031 | 7.93537 |
| protein C receptor, endotelial (EPCR) | PROCR | 2.735 | 4.11815 | 5.63 | 15.0201 |
| **Ion binding** | | | | | |
| metallothionein 1G | MT1G | 2.715 | 2.00639 | 2.639 | 2.89159 |
| metallothionein 1H | MT1H | 2.515 | 2.33668 | 4.368 | 7.2314 |
| metallothionein 1X | MT1X | 2.546 | 2.30442 | 3.022 | 4.0148 |
| **Nuclear organization and biogenesis** | | | | | |
| carboxypeptidase A4 | CPA4 | 2.924 | 3.14366 | 2.839 | 3.27686 |
| fascin homolog 1, actin-bundling protein | FSCN1 | 5.105 | 2.45535 | 4.055 | 2.48198 |
| **Signal transduction** | | | | | |
| dual specificity phosphatase 2 | DUSP2 | 4.471 | 2.42672 | 5.089 | 2.94673 |

| | | | | | |
|---|---|---|---|---|---|
| (PAC1) | | | | | |
| RAB32, member RAS oncogene family | RAB32 | 3.896 | 2.43675 | 5.047 | 3.47442 |
| **Transport** | | | | | |
| solute carrier family 7 (LAT1), member 5 | SLC7A5 | 7.454 | 6.68393 | 4.837 | 3.26239 |

| | | | | | |
|---|---|---|---|---|---|
| **Downregulated genes** | | | | | |
| **Cell adhesion** | | | | | |
| junctional adhesion molecule 3 | JAM3 | -4.777 | 0.19145 | -25.731 | 0.13795 |
| multimerin 1 | MMRN1 | -3.545 | 0.21082 | -11.016 | 0.06491 |
| nidogen (enactin) | NID | -5.643 | 0.48854 | -2.369 | 0.30813 |
| slit homolog 2 (Drosophila) | SLIT2 | -3.182 | 0.49024 | -9.427 | 0.30273 |
| syndecan 2 | SDC2 | -3.394 | 0.15779 | -2.78 | 0.31783 |
| **Cellular defense against toxic compounds** | | | | | |
| glutathione S-transferase A4 | GSTA4 | -2.358 | 0.45777 | -7.943 | 0.22224 |
| **Development and cell differentiation** | | | | | |
| homeo box B2 | HOXB2 | -2.548 | 0.40722 | -4.604 | 0.23615 |
| inhibitor of DNA binding 2, dominant negative helix-loop helix protein | ID2 | -2.031 | 0.40894 | -4.729 | 0.18137 |
| sprouty homolog 1, antagonist of FGF signaling (Drosophila) | SPRY1 | -7.766 | 0.32769 | -13.228 | 0.13179 |
| suppressor of cytokine signaling 2 | SOCS2 | -2.01 | 0.32378 | -2.355 | 0.46024 |
| **Electron transport** | | | | | |
| aldehyde dehydrogenase 1 family, member A1 | ALDH1A1 | -5.032 | 0.11719 | -5.089 | 0.14196 |
| **Functionally unclassified genes** | | | | | |
| annexin A6 | ANXA6 | -2.102 | 0.41035 | -2.475 | 0.43148 |
| **Nuclear organization and biogenesis** | | | | | |
| B lymphoma Mo-MLV insertion region (mouse) | BMI1 | -2.196 | 0.39872 | -3.819 | 0.26948 |
| nucleosome assembly protein 1 like 3 | NAP1L3 | -4.614 | 0.18303 | -6.355 | 0.40808 |
| **Nucleotide biosynthesis** | | | | | |
| Ectonucleotide pyrophosphatase 4 | ENPP4 | -9.869 | 0.22956 | -6.935 | 0.18048 |
| **Protein modification, catabolism and translation** | | | | | |
| matrix metalloproteinase 1 | MMP1 | -3.316 | 0.29692 | -3.082 | 0.37765 |

| | | | | | |
|---|---|---|---|---|---|
| (interstitial collagenase) matrix metalloproteinase 10 (stromelysin 2) | MMP10 | -4.233 | 0.38217 | -4.77 | 0.37865 |
| **Regulation of transcription** | | | | | |
| inhibitor of DNA binding 4, dominant negative helix-loop helix protein | ID4 | -3.943 | 0.12224 | -8.888 | 0.07752 |
| zinc finger protein 91 (HPF7, HTF10) | ZNF91 | -7.163 | 0.41156 | -5.224 | 0.48333 |
| **RNA binding** | | | | | |
| CUG triplet repeat, RNA binding protein 2 | CUGBP2 | -3.745 | 0.30752 | -7.223 | 0.10279 |
| decapping enzyme hDcp2 | DCP2 | -2.115 | 0.46469 | -2.104 | 0.43709 |
| **Structural function** | | | | | |
| Microtubule-associated protein 1B | MAP1B | -2.51 | 0.20551 | -3.547 | 0.19213 |

positive cell lines as compared to both the normal cervix and C-33A (Fig.12 and Table 10). Upregulated genes included those involved in cell stress by DNA damage (DNA-damage-inducible transcript 4 [DDIT4]), amino acid transporter involved in tumor progression (solute carrier family 7, member 5 [SLC7A5/LAT1]), immune response (prostaglandin E synthase [PTGES]) and development (tuftelin 1 [TUFT1]). Genes involved in development (homeo box B2 [HOXB2]), growth factor regulation and development (sprouty homolog 1, antagonist of FGF [SPRY1]), and cell adhesion (junctional adhesion molecule 3 [JAM3]) were found to be downregulated (Table 10).

*Gene expression analysis in cell lines containing integrated versus episomal HPV DNA*

Since the integration of HPV DNA into the cellular genome is usually associated with the progression of HPV-associated disease, we also analyzed gene expression differences between cell lines containing integrated or episomal HPV DNA. Due to the paucity of cell lines containing episomal HPV DNA, we grouped the 20863 and L-18 cell lines containing episomal HPV 16 and 18 DNA, respectively, into a single group. Similarly, cell lines containing integrated HPV 16 and 18 DNAs were grouped together. A total of 411 and 805 genes were found to be up- and down-regulated, respectively, using the pooled variance t-Test analysis in cell lines containing integrated HPV DNA (CaSki, SiHa, 201402, 20861 and HeLa) compared to cell lines containing episomal HPV (20863 and L-

18) (supplementary data, Table R). Also, 309 and 172 genes were up- and down-regulated, respectively, using the SAM-Test analysis (supplementary data, Table S). When the differentially expressed genes from both the analyses were compared, 52 and 47 genes showed up- and down-regulation, respectively, in cell lines with integrated HPV DNA (Table 11). The upregulated genes included those involved in regulation of mitosis (centromere protein F, 350/400ka (mitosin) [CENPF]), cell growth and anti-adhesion (podocalyxin-like [PODXL]) and chromosome organization (SMC4 structural maintenance of chromosomes 4-like 1 [SMC4L1]). Genes downregulated in cell lines containing integrated HPV DNA included those involved in defense response (GATA binding protein 3 [GATA3]), maintenance of chromatin architecture (high mobility group AT-hook 2 [HMGA2]) and epidermis development (laminin, gamma 2 [LAMC2]). Two upregulated genes identified in our study have previously been reported by Alazawi *et al* (Alazawi*, et al.*, 2002). comparing the episomal HPV cell line 20863 and its subsequent passages lacking feeder cells (which is expected to result in HPV DNA integration). These genes are involved in vesicle coat assembly (bone marrow stromal cell antigen 2 [BST2]) and DNA topological changes (topoisomerase (DNA) II alpha 170kDa [TOP2A]) (Table 11).

*Gene expression analysis in HPV-18 containing cell lines compared to those containing HPV-16*

We also analyzed gene expression differences between cell lines containing HPV-16 and HPV-18 DNA. All HPV-16 positive cell lines (CaSki, SiHa, 201402, 20863 and 20861) containing either integrated or episomal viral DNA were grouped together. Similarly, HeLa and L-18 cell lines containing HPV-18 DNA were also grouped. A total of 443 and 330 genes showed up- and down-regulation, respectively, in HPV-18 cell lines compared to HPV-16 cell lines using the pooled variance t-Test analysis (supplementary data, Table T). On the other hand, 80 and 3 genes showed up- and down-regulation, respectively, in HPV-18 cell lines using the SAM-Test analysis (supplementary data, table U). When the genes identified by the two different statistical methods were compared, 36 upregulated and 3 downregulated genes were found to be common (Table 12). Some of the up-regulated genes in HPV-18 positive cell lines included those involved in cytoskeleton organization (keratin 8 [KRT8], keratin 18 [KRT18]), cell adhesion (mesothelin [MSLN]), migration and proliferation (vimentin [VIM]) and calcium ion binding (S100 calcium binding protein A4 [S100A4]). Genes involved in cell differentiation (N-myc downstream regulated gene 1 [NDRG1]) and regulation of transcription (transcription factor AP-2 gamma [TFAP2C]) were found to be downregulated in HPV-18 containing cells compared to the HPV-16 positive cell lines (Table 12).

*Validation of microarray data by Real time Quantitative RT-PCR and Northern blot analyses*

We carried out QRT-PCR analysis to study the expression of a few representative genes in order to validate the differential gene expression profiles obtained by microarray analysis. Four upregulated genes (transcription factor AP-2 alpha [TFAP2A], carboxypeptidase A4 [CPA4], neurofilament, heavy polypeptide 200kDa [NEFH], and solute carrier family 7 member 5 [LAT1]), and four downregulated genes (sprouty homolog 1, antagonist of FGF signaling [SPRY1], inhibitor of DNA binding 4, dominant negative helix-loop helix protein [ID4], immunoglobulin superfamily, member 4 [IGSF4], and nucleosome assembly protein 1-like 3 [NAP1L3]) were selected for validation (genes differentially expressed between HPV-positive cell lines compared to normal cervix or C-33A) (Fig. 13). These data showed that the results of QRT-PCR data were consistent with those of microarray analysis, although there were variations in the fold-difference determined by the microarray and QRT-PCR analysis. We also use northen blot analysis for one upregulated gene (LAT1) and one downregulated gene (NAP1L3) in order to confirmed the results obtained by the QRT-PCR. In the northern blot of fig. 13, we found high expression of LAT1 in all the HPV-positive cervical cell line and less expression in the HPV-negative cell line C-33A as well as no detectable expression of LAT1 in normal cervical tissue. In contrast, the northern blot for NAP1L3 shows no detectable expression in all the HPV-positive cell lines, a very faint band in C-33A and a detectable expression in normal cervical tissue (Fig. 13). Together, our QRT-PCR data and northern blot verified some of the differentially expressed genes found in our microarray experiments.

**Table 11.  DIFFERENTIALLY EXPRESSED GENES IN HPV-INTEGRATED COMPARED TO HPV-EPISOMAL CELL LINES**

| Gene Title | Gene Symbol | pooled t-Test | SAM-Test |
|---|---|---|---|
| **Upregulated genes** | | | |
| **Cell cycle** | | | |
| activator of S phase kinase | ASK | 2.986 | 2.05485 |
| cyclin-dependent kinase inhibitor 3 | CDKN3 | 2.135 | 2.60898 |
| **Cell proliferation** | | | |
| aldo-keto reductase family 1, member C3 | AKR1C3 | 2.442 | 3.80378 |
| bone marrow stromal cell antigen 2 | BST2 | 2.342 | 18.48414 |
| disabled homolog 2, mitogen-responsive phosphoprotein | DAB2 | 2.185 | 2.47187 |
| **DNA repair** | | | |
| RecQ protein-like (DNA helicase Q1-like) | RECQL | 2.039 | 2.12691 |
| **DNA replication** | | | |
| CDC7 cell division cycle 7 (S. cerevisiae) | CDC7 | 2.7 | 2.15410 |
| centromere protein F, 350/400ka (mitosin) | CENPF | 2.513 | 2.28573 |
| topoisomerase (DNA) II alpha 170kDa | TOP2A | 2.166 | 2.13678 |
| topoisomerase (DNA) II binding protein 1 | TOPBP1 | 2.267 | 2.01674 |
| **Development and cell differentiation** | | | |
| podocalyxin-like | PODXL | 2.622 | 4.63347 |
| **Nuclear organization and biogenesis** | | | |
| H1 histone family, member X | H1FX | 2.249 | 1.91310 |
| SMC4 structural maintenance of chromosomes 4 like 1 | SMC4L1 | 2.001 | 1.97850 |
| **Downregulated genes** | | | |
| **Cell adhesion** | | | |
| chondroitin sulfate proteoglycan 2 (versican) | CSPG2 | -2.21 | 0.46375 |
| collagen, type XVI, alpha 1 | COL16A1 | -3.238 | 0.41036 |
| fibronectin 1 | FN1 | -2.952 | 0.25825 |
| laminin, gamma 2 | LAMC2 | -2.991 | 0.35576 |
| **Cell proliferation** | | | |
| amphiregulin | AREG | -2.147 | 0.29429 |
| **Development and cell differentiation** | | | |
| kallikrein 7 (chymotryptic, stratum corneum) | KLK7 | -2.712 | 0.27634 |
| keratin 13 | KRT13 | -2.336 | 0.12282 |
| **Immune response** | | | |
| GATA binding protein 3 | GATA3 | -2.093 | 0.45008 |
| **Nuclear organization and biogenesis** | | | |
| high mobility group AT-hook 2 | HMGA2 | -2.982 | 0.22794 |
| villin-like | VILL | -2.029 | 0.48133 |

**Table 12    DIFFERENTIALLY EXPRESSED GENES IN HPV-18**

**CONTAINING CELL LINES COMPARED TO HPV-16 POSITIVE**

**CELL LINES**

| Gene Title | Gene Symbol | pooled t-Test | SAM-Test |
|---|---|---|---|
| **Upregulated genes** | | | |
| **Cell adhesion** | | | |
| collagen, type XV, alpha 1 | COL15A1 | 2.445 | 8.02698 |
| mesothelin | MSLN | 9.127 | 4.00520 |
| **Development and cell differentiation** | | | |
| paired-like homeodomain transcription factor 2 | PITX2 | 4.052 | 6.33972 |
| **Electron transport** | | | |
| cytochrome c oxidase subunit VIIa polypeptide 1 | COX7A1 | 2.235 | 4.95734 |
| **Functionally unclassified genes** | | | |
| preferentially expressed antigen in melanoma | PRAME | 3.045 | 4.43197 |
| **Immune response** | | | |
| D component of complement (adipsin) | DF | 2.104 | 3.26767 |
| melanoma associated gene | D2S448 | 3.457 | 3.00100 |
| **Ion binding** | | | |
| S100 calcium binding protein A4 (metastasin) | S100A4 | 2.356 | 2.76983 |
| SNRPN upstream reading frame | SNURF | 11.682 | 5.41101 |
| **Protein modification, catabolism and translation** | | | |
| eukaryotic translation initiation factor 1A, Y linked | EIF1AY | 2.034 | 4.82245 |
| **Regulation of transcription** | | | |
| nuclear receptor subfamily 2, group F, member 1 | NR2F1 | 2.545 | 7.90107 |
| **Signal transduction** | | | |
| Arg/Abl-interacting protein ArgBP2 | ARGBP2 | 3.555 | 2.10753 |
| chemokine (C-X-C motif) receptor 4 | CXCR4 | 2.99 | 7.57239 |
| dual specificity phosphatase 4 | DUSP4 | 3.021 | 2.00880 |
| **Structural function** | | | |
| keratin 18 | KRT18 | 3.007 | 2.09911 |
| keratin 8 | KRT8 | 3.723 | 2.93771 |
| microfibrillar associated protein 5 | MFAP5 | 4.329 | 6.58325 |
| vimentin | VIM | 2.158 | 2.57079 |
| Wiskott-Aldrich syndrome protein interacting protein | WASPIP | 3.036 | 3.42596 |
| **Downregulated genes** | | | |

| | | | |
|---|---|---|---|
| **Development and cell differentiation** | | | |
| N-myc downstream regulated gene 1 | NDRG1 | -6.132 | 0.29486 |
| **Functionally unclassified genes** | | | |
| synaptogyrin 3 | SYNGR3 | -5.394 | 0.41355 |
| **Regulation of transcription** | | | |
| transcription factor AP-2 gamma | TFAP2C | -3.737 | 0.46549 |

*In situ hybridization of NAP1L3 and LAT1 in tissue arrays*

In order to verify that the differentially expressed genes LAT-1 and NAP1L3 are not affected in HPV-positive cervical cell line as a consequence of *in vitro* immortalization or specific tissue culture variations, we carried out *in situ* hybridization for the mRNAs of these genes in tissue-arrays containing ~50 different cervical tumor samples and 4 normal cervix controls per slide. When we used the NAP1L3 probe in the cervical carcinoma tissue-array, we found that most of these carcinomas were negative for NAP1L3 mRNA (Fig. 14*B* and 14*C* and Table 13). Interestingly, all the four normal cervical samples were positive for NAP1L3 transcripts (yellow fluorescent spots) (Fig. 14*A*), supporting the data obtained with the HPV-positive cervical cell lines (Fig. 13). On the other hand, when we hybridized the LAT1 probe to the cervical carcinoma tissue-arrays, most of the samples in the array were positive for LAT1 mRNA (Fig.14*E* and 14*F*). The normal cervical tissues did not give a positive signal for LAT1 transcripts (Fig. 14*D* and Table 13). Together, our data showed that NAP1L3 and LAT1 are also affected in cervical carcinoma tumor tissues in a manner similar to that observed with the cell lines, suggesting a possible role of these genes in the carcinogenesis of the cervix.

# Fig. 13. VALIDATION OF MICROARRAY DATA IN HPV-POSITIVE CELL LINES BY QUANTITATIVE RT-PCR AND NORTHERN BLOT
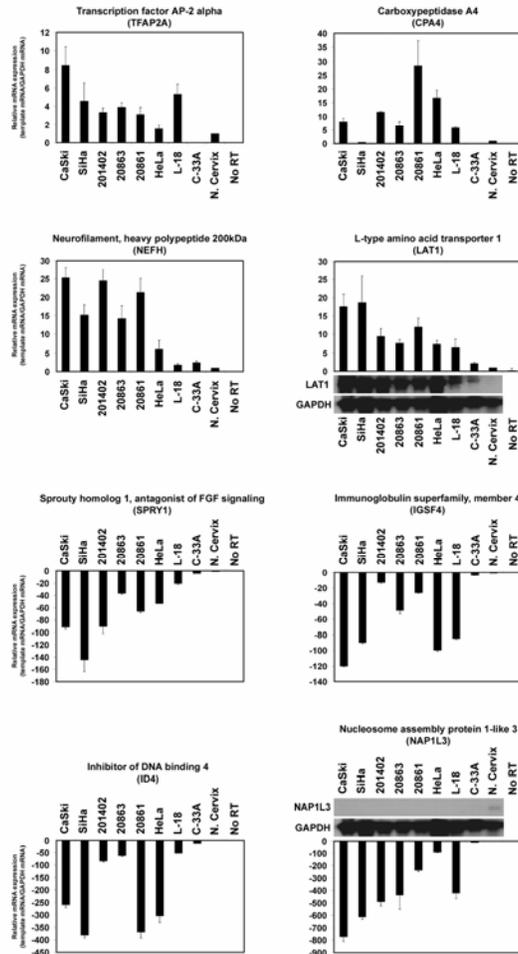


**Four upregulated genes (TFAP2A, CPA4, NEFH and LAT1), and four downregulated genes (SPRY1, IGSF4, ID4 and NAP1L3) between HPV-positive cell lines compared to the normal cervix tissue or the HPV-negative cell line C-33A were selected for validation by QRT-PCR. All reactions were performed in triplicates. Relative expression of the target gene was calculated using the 2 delta CT method, where GAPDH was used as the endogenous control gene. For northern blot we used the probes for LAT1 and NAP1L3 in all the cell lines and normal cervix. GAPDH was used as a loading control.**

# Fig. 14. *In Situ* HYBRIDIZATION OF NAP1L3 AND LAT1 IN A CERVICAL CARCINOMA TISSUE-ARRAY
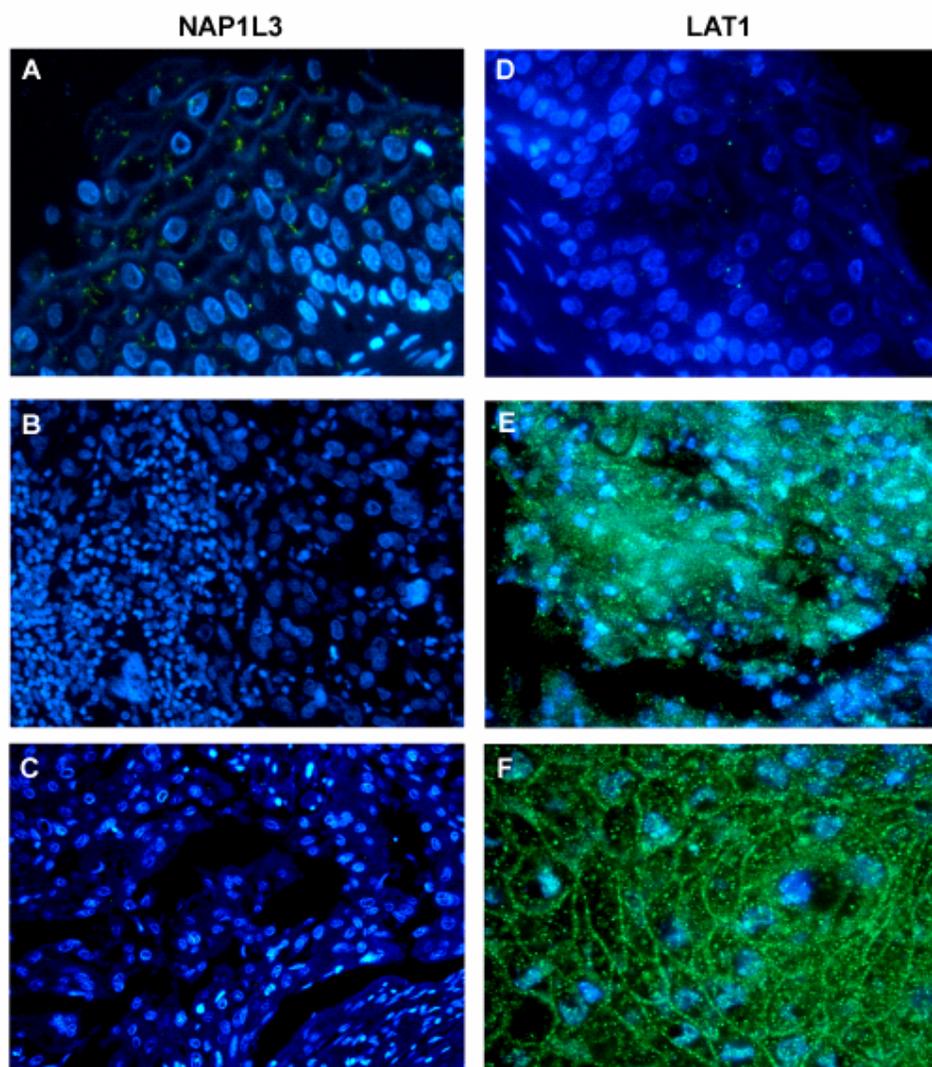


Cervical carcinoma tissue-arrays (containing ~50 different cervical tumor samples and 4 normal cervix controls per slide) were used for NAP1L3 (Fig. 14*A*-14*C*) and LAT1 (Fig. 14*D*-14*F*) mRNA detection by *in situ* hybridization. Positive hybridization signal is shown as yellow fluorescent spots. Fig. 14*A* and 14*D* are normal cervical tissues. Fig. 14*B*, 14*C*, 14*E* and 14*F* are cervical carcinoma tissues.

**Table 13.  In Situ HYBRIDIZATION RESULTS FOR LAT1 AND NAP1L3 IN DIFFERENT CERVICAL CARCINOMA TISSUE SAMPLES PRESENT IN THE TISSUE-ARRAY SLIDES**

| Tissue specimens | No. of samples | LAT1 Positive (%) | NAP1L3 Positive (%) |
|---|---|---|---|
| Squamous Cell Carcinoma *in situ* | 3 | 2 (66.6%) | 0 (0%) |
| Squamous Cell Carcinoma | 37 | 32 (86%) | 2 (5.4%) |
| Adenosquamous Carcinoma | 5 | 4 (80%) | 1 (20%) |
| Microinvasive Squamous Cell Carcinoma | 5 | 5 (100%) | 0 (0%) |
| Metastatic Carcinoma (lymph node) | 5 | 5 (100%) | 0 (0%) |
| Total Tissue-array Carcinoma Samples | 50 | 43 (86%) | 3 (6%) |
| Normal Cervix | 4 | 0 (0%) | 4 (100%) |

Signal was recorded as negative if <5% of the tumor or normal tissue samples were positive

## 3.4    DISCUSSION

Cervical cancer represents an important health problem worldwide. An understanding of the molecular interactions between HPV and cervical cells is important for a better understanding of the pathogenesis of HPV-associated cancers as well as for early diagnosis and possible new treatment approaches for diseases caused by this virus. In this study, we used high-density microarrays to analyze gene expression differences in several HPV-positive and HPV-negative cervical cell lines and related our findings to mechanisms or pathways altered in HPV-positive cells as a consequence of viral gene expression, integration, and other downstream effects.   A few investigators have previously studied gene expression profiles of HPV-containing cell lines using microarray analysis. These studies used a wide variety of microarray platforms, control samples, normalization approaches and statistical analyses that showed a range of differentially expressed genes (Alazawi*, et al.*, 2002;Chang*, et al.*, 2000;Chen*, et al.*, 2003;Duffy*, et al.*, 2003;Lee*, et al.*, 2004;Schena*, et al.*, 1996;Zhu*, et al.*, 1998). For example, the study by Chang et al. used HPV-31 transfected primary keratinocytes and a microarray with 7,075 human genes and expressed sequence tags (ESTs) (Chang*, et al.*, 2000). Another analysis used cervical cancer tissues and a cDNA microarray containing 30,732 Unigene clones and normal tissue as control (Chen*, et al.*, 2003). Ahn *et al.* used invasive cancer tissues and a 4,700 cDNA microarray and used the HaCat cells as a control (Ahn*, et al.*, 2004). Alazawi et al. carried out microarray analysis using an

episomal HPV-16 W12 clone and compared it to a W12 clone containing integrated HPV DNA utilizing a 12,600 human transcript high-density microarray (Alazawi*, et al.*, 2002).

A significant problem arises when comparing results of different microarray data from individual studies due to the use of a wide variety of statistical analyses that show different levels of false positive/negative rates (Tan*, et al.*, 2003). To our knowledge, our study is the first to use two different types of normalizations (GEDA and dChip normalization) and two statistical tests (pooled variance t-Test and SAM-Test) in order to overlap the results of each analysis and identify a more robust panel of differentially expressed genes and possibly reduce the number of false positives. Our study also utilized a high density oligonucleotide microarray platform (Affymetrix U133A GeneChip® expression microarray) that represents a larger number of human transcripts (22,215). Only one recent report utilized this expression array and compared the gene expression profiles in primary cell lines derived from cervical cancer tissues and the normal cervix (Santin*, et al.*, 2005). Many of the differentially expressed genes reported in this study were also found in our analyses.

Since episomal HPV DNA present in cell lines frequently integrates during in vitro propagation, we first confirmed by restriction enzyme and Southern blot analyses that the HPV-positive cell lines 20863 and L-18 contained episomal HPV-16 or HPV-18 DNA, respectively, after several passages as previously

reported (Alazawi, *et al.*, 2002). We also confirmed the expression of the viral E6 and E7 oncogenes in various HPV-positive cell lines by Northern blot analysis (Fig. 11). This analysis showed multiple transcripts corresponding to the E6 and E7 genes (Fig. 11), consistent with previous studies showing multiple splicing variants of these genes (Baker, *et al.*, 1987). As expected, cell lines containing episomal HPV DNA generally expressed lower levels of the E6/E7 genes as compared to those containing integrated HPV DNA (Fig. 11). This is likely to be due to the repression of the p97 promoter transcribing the E6/E7 genes by the E2 protein and/or the chromatin structure of the episomes (Bechtold, *et al.*, 2003;May, *et al.*, 1994). The E2 gene is frequently interrupted or deleted upon integration of the HPV DNA into the chromosome (zur Hausen, 2002). As expected, the E2 protein was expressed in cell lines containing episomal HPV DNA (Fig. 11). As shown earlier (Baker, *et al.*, 1987), the CaSki cell line containing integrated HPV-16 DNA also expressed high levels of the E2 transcript due to heat-to-tail tandem integration of multiple copies of HPV DNA that restores an intact E2 open reading frame (Fig. 11). The 20861 cell line that contains integrated HPV DNA expressed significant levels of an E2 transcript of larger than expected size (Fig. 11). This is presumably due to the generation of a fusion transcript resulting from the fusion of E2 and a cellular gene.

After confirming appropriate expression of the viral genes in various cell lines, we analyzed the cellular gene expression profiles in order to identify differentially expressed genes in HPV-positive cervical cell lines compared to the

normal cervix tissue. We used two statistical methods, pooled t-Test and SAM-Test, for data analysis and used a cut-off of two-fold to identify differentially expressed genes. A comparison of the differentially expressed genes using these two statistical methods showed that 272 upregulated and 605 downregulated genes were in common. Some upregulated genes identified in our study are known to be involved in various cellular pathways such as cell cycle (AURKB), DNA repair (FEN1,OGG1), Signal transduction (RAB32), Nuclear organization and biogenesis (CPA4), and cell transport (RAB9P40) (Table 8). Some examples of upregulated genes found in our analysis as well as in previous microarray studies are genes involved in cell cycle  (CCNB1,CDKN2A,MCM7), DNA repair (H2A histone family, member X [H2AFX]) and regulation of transcription (TFAP2A) (Table 8) (Beger*, et al.*, 2001;Li*, et al.*, 2005;Santin*, et al.*, 2005;Steenbergen*, et al.*, 2004;Tan*, et al.*, 2003). Other upregulated genes included those implicated in human carcinogenesis in previous studies (Santin*, et al.*, 2005;Walter*, et al.*, 2003). For example, high expression of the cell cycle regulator cyclin B1 has been seen in a wide variety of human tumors including cervical cancer, and is also associated with mutations or deletions of p53 (Santin*, et al.*, 2005). It is likely that binding and degradation of p53 by the HPV E6 protein results in overexpression of cyclin B1. Another molecule involved in cell cycle regulation that was found to be upregulated in our analysis was cyclin-dependent kinase inhibitor 2A (CDKN2A). As suggested previously by Santin *et al*. (Santin*, et al.*, 2005), overexpression of CDKN2A may result from the inactivation of pRB function by the HPV E7 protein, allowing the E2F1

transcription factor to inhibit cyclin D1-dependent kinase activity and consequently upregulate CDKN2A. Our studies also identified several genes that were previously not known to be upregulated in cervical cancer. For example, Rab9 effector p40 (RAB9P40) which is a small GTPase involved in lipid transport from the late endosomal-lysosomal system to the trans-Golgi network, was found to be upregulated in HPV-positive cell lines. The upregulation of RAB9P40 has been co-related with increased expression of the human telomerase reverse transcriptase (hTERT) in immortalized human cells (Walter*, et al.*, 2003). Telomerase activation by high risk HPV E6 proteins could result in upregulation of RAB9P40 (Klingelhutz*, et al.*, 1996).

Our microarray analysis identified several downregulated genes that were not reported in previous studies with HPV-positive cervical cell lines.  Such genes included those involved in cell adhesion (CLDN10,NID), cell cycle (CUL5), regulation of transcription (ID4), cell differentiation (DMBT1), and DNA repair (MBD4) (Table 8). DMBT1 (deleted in malignant brain tumor 1 gene) is a candidate tumor suppressor gene previously found to be downregulated in brain, esophageal, gastric, colorectal, lung and oral cancers (Imai*, et al.*, 2005). DMBT1 is known to trigger cellular differentiation when located in the extracellular matrix (Mollenhauer*, et al.*, 2002). Thus, downregulation of DMBT1 may promote epithelial tumorigenesis as suggested earlier. Our results also confirm the downregulation of several genes identified in previous studies. Some of these genes have been shown to be involved in immune response (IGSF4, IFI27), cell

adhesion (FN1), cell differentiation (collagen, type I, alpha 1 [COL1A1]) and protein modification (cathepsin B [CTSB]). IGSF4 (tumor suppressor gene immunoglobulin superfamily, member 4) earlier known as TSLC1, SgIGSF or SynCAM, is a novel immunoglobulin (Ig)-like intercellular adhesion molecule involved in spermatogenic adhesion, mast-cell adhesion, formation of functional synapses, and probable cell-to-cell interactions of epithelial cells (Watabe, *et al.*, 2003). The silencing of IGSF4 has been associated with aberrant promoter hypermethylation in several human cancers, including cervical cancers and in HPV-positive cell lines (Li, *et al.*, 2005;Steenbergen, *et al.*, 2004). Reduced expression of IGSF4 in HPV-positive cell lines may disrupt the adhesion of epithelial cells and contribute to tumor cell motility.

Using the HPV-negative cervical carcinoma cell line C-33A as a control, we found 305 upregulated and 232 downregulated genes in all the HPV-positive cell lines (Table 9) by combining the results from the two statistical analyses. Upregulated genes belong to different cellular pathways such as metastasis (ACTN4), antiestrogen resistance (BCAR3), mitogenic signaling (CAV1, CAV2), angiogenesis (EPAS1) and double-strand DNA break repair via nonhomologous end-joining (MRE11A). Downregulated genes found in our analyses included those involved in transcriptional regulation (CITED2), apoptosis (DFFB), cellular morphogenesis (MYH10), centrosome formation (PCM1) and mismatch repair (PMS1) (Table 9). One of the genes found to be upregulated for the first time in HPV-positive cell lines was actinin, alpha 4 (ACTN4). This protein is involved in

actin cytoskeleton remodeling and  is known to promote cell motility and metastasis of colorectal cancer cells (Honda*, et al.*, 2005). Another upregulated gene found in our analysis encodes carboxypeptidase A4 (CPA4) that is located at the putative prostate cancer-aggressiveness locus at 7q32 (Kayashima*, et al.*, 2003). CPA4 is expressed at extremely low levels in human adult tissues but is abundant in prostate cancer cell lines, suggesting that CPA4 may modulate the function of peptide hormones that play an essential role in the growth and differentiation of prostate epithelial cells (Kayashima*, et al.*, 2003). It may play a similar role in the growth of HPV-positive cervical epithelial cells.

When cellular gene expression profiles of all HPV-positive cervical cell lines were compared to those of both the normal cervix and the HPV-negative cell line C-33A by the above two statistical methods, a total of 20 genes were found to be upregulated and 37 were downregulated (Table 10). Upregulated genes included those involved in cell stress by DNA damage (DDIT4), amino acid transporter involved in tumor progression (SLC7A5/LAT1), immune response (PTGES), and development (TUFT1). Downregulated genes included those involved in development (HOXB2), growth factor regulation and development (SPRY1), and cell adhesion (JAM3) (Table 10). One of the upregulated genes found in our analysis as well as in previous studies was the antimicrobial humoral response prostaglandin E synthase (PTGES) (Santin*, et al.*, 2005). There is evidence that lung cancer cells secreting high levels of PTGES alter the cytokine balance in the tumor micro-environment and could help

in the immune evasion of neoplastic cells (Huang*, et al.*, 1998). A similar immune evasion could play an important role in HPV pathogenesis.

Integration of HPV DNA usually occurs during the progression of HPV infection, and the DNA is found to be mostly integrated in high-grade lesions and cervical cancers (zur Hausen, 2002). We analyzed the differences in cellular gene expression profiles between cervical carcinoma cell lines containing integrated HPV DNAs (CaSki, SiHa, 201402, 20861 and HeLa) and those containing episomal HPV DNA that mimic low-grade lesions (20863 and L-18). Due to the limited number of available cell lines containing episomal HPV DNA, the 20863 and L-18 cell lines containing HPV-16 and HPV-18 DNA, respectively, were combined into a group. Data analysis using the two statistical methods showed that 52 genes were upregulated while 47 were downregulated in cell lines containing integrated HPV DNA compared to those containing episomal DNA (Table 11). The upregulated genes in HPV-16 or HPV-18 integrated cell lines included those involved in the regulation of mitosis (CENPF), cell growth and anti-adhesion (PODXL), and chromosome organization (SMC4L1). The downregulated genes included those involved in defense response (GATA3), maintenance of chromatin architecture (HMGA2), and epidermis development (LAMC2). The CENPF gene that encodes the centromere protein F, 350/400ka (mitosin) was found to be upregulated. CENPF overexpression may de-regulate kinetochore assembly at the G2/M checkpoint and facilitate defective chromosome segregation, promoting cervical carcinogenesis (de la Guardia*, et*

*al.*, 2001). The upregulation of CENPF in cell lines containing integrated HPV DNA is consistent with E7 expression and chromosomal instability associated with HPV integration (Pett*, et al.*, 2004). Our data also showed that the PODXL (mucin-like surface sialoglycoprotein podocalyxin-like) gene is upregulated in cell lines containing integrated HPV DNA. PODXL is an integral membrane glycoprotein that has different cell adhesion properties in different tissues and is repressed by the p53 protein at the transcriptional level (Stanhope-Baker*, et al.*, 2004). Increased expression of E6 in cell lines with integrated HPV DNA causes p53 degradation and could result in an increase in PODXL expression (zur Hausen, 2002). Several genes that are downregulated in cell lines containing integrated HPV DNA were also identified (Table 11). The protein encoded by one such gene, GATA3 (essential transcription factor GATA binding protein 3), is part of the GATA family zinc-finger transcription factors that are involved in differentiation and development of different tissues in vertebrates (Lim*, et al.*, 2000). A recent study showed downregulation of GATA3 mRNA by differential display in HPV-immortalized keratinocytes, HG-SIL and cervical carcinomas suggesting that this is a late event in cervical tumorigenesis (Steenbergen*, et al.*, 2002). The increased expression of GATA3 may be co-related with integration of HPV DNA into the host genome and progression of cervical cancer. Two upregulated genes identified in our study have previously been reported by Alazawi et al. comparing the episomal HPV cell line 20863 and its subsequent passages lacking feeder cells (which is expected to result in the integration of

HPV DNA) (Alazawi, *et al.*, 2002). These genes are involved in vesicle coat assembly (BST2) and DNA topological changes (TOP2A).

Although both HPV types 16 and 18 are high-risk viruses, HPV-16 is commonly associated with SCC, whereas HPV-18 is predominantly associated with AC (Burk, *et al.*, 2003). In order to identify cellular pathways that may be specifically altered by the presence of HPV 16 or 18 DNA, we compared the cellular gene expression profiles of cell lines containing HPV 16 or 18 DNA. We found that 36 genes were upregulated while 3 were downregulated in HPV-18 positive cells based upon a comparison of differentially expressed genes using two different statistical methods (Table 12). Of note, the keratin 8 (KRT8) and keratin 18 (KRT18) genes involved in cytoskeleton organization were found to be upregulated in HPV-18 containing cells as compared to the HPV-16 positive cell lines. The relatively few genes that were found to be differentially expressed in this analysis may suggest the involvement of predominantly similar pathways during the progression of cervical carcinogenesis induced by the high-risk HPV-16 and HPV-18.

We carried out QRT-PCR and northern blot analyses in order to validate the differential gene expression profiles obtained by microarray analysis. All the eight differentially expressed genes used for QRT-PCR (TFAP2A, CPA4, NEFH, LAT1, SPRY1, ID4, IGSF4, and NAP1L3) showed data consistent with our microarray data (Fig. 13). We also used Northen blot analysis for one

upregulated gene (LAT1) and one downregulated gene (NAP1L3) in order to further confirm the results obtained with QRT-PCR. We found the same pattern of expression between this analysis and the QRT-PCR. *In situ* hybridization of NAP1L3 and LAT1 was done in tissue-arrays containing ~50 different cervical tumor samples and 4 normal cervix controls per slide in order to verify that these genes are not an artifact of *in vitro* immortalization or specific tissue culture variations. We found that most of the cervical carcinomas were negative for NAP1L3 mRNA, while most of the cervical carcinomas in the array were positive for LAT1 mRNA (Fig. 14 and Table 13). Together, our data showed that NAP1L3 and LAT1 are also affected in cervical carcinoma tumor tissues suggesting a possible role in the carcinogenesis of the cervix.

In summary, these results of global gene expression profile in HPV-positive cervical cell lines could lead to the identification of new cellular pathways affected during the progression of HPV disease.

**Chapter 4**



**HUMAN PAPILLOMAVIRUS TYPE 16 REDUCES THE EXPRESSION OF**

**MICRORNA-218 IN CERVICAL CARCINOMA CELLS**

## 4.1    INTRODUCTION

Cervical cancer is the second most common malignancy in women worldwide (Pisani, *et al.*, 2002). High-risk human papillomaviruses (HPVs) such as types 16 and 18 are causally involved in the carcinogenesis of the cervix (zur Hausen, 2002). HPVs are small double-stranded DNA viruses that contain two oncogenes E6 and E7 that are involved in cellular transformation (Hebner, *et al.*, 2006;Munger, *et al.*, 2002;zur Hausen, 2002). Most low-grade cervical lesions contain HPV DNA in an episomal state. However, in most cases of cervical carcinomas the HPV DNA is found integrated into the host chromosomes, increasing expression of E6 and E7 (Hebner, *et al.*, 2006;Munger, *et al.*, 2002;zur Hausen, 2002). The E6 protein promotes ubiquitination and proteasomal degradation of the tumor suppressor protein p53 (Band, *et al.*, 1993;Huibregtse, *et al.*, 1991;Lechner, *et al.*, 1992;Pim, *et al.*, 1994;Scheffner, *et al.*, 1990;Thomas, *et al.*, 1999;Werness, *et al.*, 1990) and PDZ domain-containing disc large protein (DLG) (Gardiol, *et al.*, 1999;Lee, *et al.*, 2004). E6 also dysregulates the cell cycle (Shai, *et al.*, 2007). The E7 protein binds to and inactivates the function of the pRB and related tumor suppressor proteins p107 and p130 (Davies, *et al.*, 1993;Dyson, *et al.*, 1989;Munger, *et al.*, 1989). E7 also interacts with additional cellular proteins such as TBP, histone H1 kinase and cyclin E (Davies, *et al.*, 1993;Massimi, *et al.*, 1996;McIntyre, *et al.*, 1996). In addition, E6/E7 expression promotes chromosomal instability, foreign DNA

integration and other mutagenic events in the cell (Duensing, *et al.*, 2000;Pett, *et al.*, 2004;Schaeffer, *et al.*, 2004).

Originally discovered in *C. elegans*, microRNAs (miRNAs) are small non-coding RNAs that are conserved across many species and may regulate thousands of mRNA targets (Lee, *et al.*, 1993;Lewis, *et al.*, 2005). MiRNAs are transcribed in the nucleus (Cai, *et al.*, 2004;Lee, *et al.*, 2004), and after processing (Hutvagner, *et al.*, 2001;Lee, *et al.*, 2003;Lee, *et al.*, 2002;Lund, *et al.*, 2004), they associate with the RISC complex and act as negative regulators of gene expression by binding to their complementary mRNA targets and either repressing translation or promoting mRNA degradation (Khvorova, *et al.*, 2003;Kim, 2005;Schwarz, *et al.*, 2003). Recently, changes in the expression of miRNAs have been shown to be associated with human cancers (Calin, *et al.*, 2006;Esquela-Kerscher, *et al.*, 2006;Lu, *et al.*, 2005). For example, miR-15 and miR-16 which negatively regulate the *Bcl-2* oncogene are usually underexpressed in chronic lymphocytic leukemia (CLL) (Calin, *et al.*, 2002;Cimmino, *et al.*, 2005). The *let-7* family of miRNAs regulates the *ras* oncogenes, and in lung cancer underexpression of these miRNAs results in overexpression of *Ras* (Johnson, *et al.*, 2005;Takamizawa, *et al.*, 2004). Also, the *C-myc* oncogene increases the expression of the miRNA cluster miR-17-92, and their overexpression has been observed in B cell lymphomas (O'Donnell, *et al.*, 2005;Ota, *et al.*, 2004).

Little is known about miRNA expression in cervical carcinogenesis. In this study, we demonstrate differential expression of several miRNAs in HPV-16 positive cervical cell lines and tissues compared to the normal cervical tissue and an HPV-negative cervical carcinoma cell line. We also demonstrate that miR-218 and the tumor suppressor gene *SLIT2,* whose intron encodes miR-218, are specifically downregulated in several HPV-16 positive cervical cell lines and tissues, and this effect is likely mediated by the HPV-16 E6 oncogene. Finally, our studies show that *LAMB3* is a possible target of miR-218 at a transcriptional level.


## 4.2    MATERIALS AND METHODS


*Cell Lines*


Cervical cancer cell lines CaSki, SiHa (HPV-16 positive) and C-33A (HPV-negative) and their growth conditions have been described (Meissner, 1999;Scheffner*, et al.*, 1991). Three clonal populations of the HPV-16 cervical cell line W12 (Stanley*, et al.*, 1989), 20863 (episomal HPV-16) and 20861 and 201402 (integrated HPV-16) were obtained from the laboratories of Drs. Margaret Stanley and Paul Lambert (Medical Research Council, UK and University of Wisconsin, USA, respectively). The cell lines U2OS-Neo, U2OS-E6 and U2OS-E7 (Duensing*, et al.*, 2000) were obtained from Dr. Stefan Duensing (University of Pittsburgh, USA).

*Cervical Tissues Characteristics*

Human cervical tissue samples were collected under an IRB approved protocol from patients with cervical preinvasive neoplasia undergoing leep excision or radical hysterectomy for invasive cervical cancer. Informed consent was obtained from all subjects. Parallel specimens of the same site were stained (Hemotoxylin and Eosin) to confirm the diagnosis. HPV-16 positivity was confirmed by RT-PCR. Total RNA from normal human cervix was obtained from a pool of 2 different donors (Stratagene).

*MicroRNA Microarray Analysis*

Total RNA was extracted using the Ultraspec™ RNA isolation system (BIOTECX) according to the manufacturer's instructions. Total RNA from cell lines and normal cervix was used to isolate small RNAs (<200 nt) which includes miRNAs using the RNeasy Mini Kit and the MinElute Cleanup Kit (Qiagen). *MirVana* miRNA Bioarray (Ambion) was used to analyze miRNA expression in cervical cell lines. The Bioarray consist of 662 probes (~22 nt long antisense oligonucleotides) that include known human miRNAs (328), theoretical human microarrays known as ambi-miRs (152) as well as unique miRNAs from mouse (114) and rat (Scheffner*, et al.*, 1991). Each array contains probes for all miRNAs in quadruplicate, and the signals obtained for each miRNA is represented as an

average of these four values and is subsequently used for statistical analysis. The enriched small RNA fractions obtained from 25 µg of total RNA was fluorescently labeled and hybridized to the Bioarrays according to the manufacturer's instructions (Ambion). Each experiment was done twice. The Bioarrays were scanned using GenePix 4000B Scanner and the median fluorescent intensity was obtained after subtracting the background using the GenePix Pro 6.0 software. To identify differential miRNA expression between samples, the median fluorescent intensities were log$_2$ transformed and normalized using the median within the array and the global mean adjustment among arrays using the GEDA program (http://bioinformatics.upmc.edu/Help/GEDADescription.html). After transformation and normalization, we used the Significance Analysis of Microarray (SAM) program version 1.21 (www-stat.stanford.edu/~tibs/SAM/) to obtain the differential expression profiles. For the unsupervised hierarchical clustering analysis, we used the Genesis 1.0 program (http://genome.tugraz.at). The fluorescent ratios of all human miRNAs in the Bioarray were used to obtain the average linkage clustering based on the Pearson correlation coefficient similarity metric.

*MicroRNA Northern blot analysis*

Enriched miRNA fractions obtained from 25 µg of total RNA were separated on 15% urea-containing polyacrylamide gels. RNA was then

transferred onto GeneScreen Plus membranes which were hybridized to ³²P-labeled oligonucleotide probes complementary to various cellular miRNAs. We hybridized a probe complementary to the housekeeping splicing-related small U6 RNA for loading control. Hybridization was carried out overnight at 50°C and the membranes were subjected to autoradiography at -80°C.

*MicroRNA and mRNA Real-time Quantitative RT-PCR Analysis*

DNase I-treated total RNA (10 ng) was subjected to qRT-PCR analysis using the TaqMan® miRNA Reverse Transcription Kit and miRNA Assays (Applied Biosystems), and the Real-Time thermocycler iQ5 (BioRad). The small nucleolar RNU43 was used as the housekeeping small RNA reference gene. For qRT-PCR analysis of *SLIT2* and *LAMB3* mRNAs, the following primers were used: (*SLIT2*, forward 5'-CTGTGAATGCAGCAGTGGAT-3' and reverse 5'-TTGTTTGGCAAGCAGCATAG-3' (116-bp product); *LAMB3*, forward 5'-GGGAGACCATGGAGATGATG-3' and reverse 5'-ACACGCTTCTCCAGTCCTGT-3' (112-bp product). Also, 500 ng of total RNA from the cervical samples was amplified using the one step QuantiTect SYBR Green RT-PCR Master Mix (Qiagen). The primer sequences for the control β-actin gene were: forward 5'- TGCGCAGAAAACAAGATGAG-3' and reverse 5'-CACCTTCACCGTTCCAGTTT-3' (114-bp product); primers for the control housekeeping glyceraldehyde-3-phosphate dehydrogenase gene (G3PDH) have

128

been described previously (Martinez, *et al.*, 2007). Relative expression of RNAs was calculated using the 2 delta CT method (Livak, *et al.*, 2001).

*HPV-16 E6 and E7 siRNA and Transfection Assays*

Double-stranded small interfering RNAs (siRNAs) against HPV-16 E6 (siRNA 209) and E7 (E7 siRNA) (Jiang, *et al.*, 2002;Tang, *et al.*, 2006) were obtained from Dharmacon. BLOCK-iT$^{TM}$ Fluorescent double-stranded oligo (with no human homologous sequence) was used as a negative control as well as to measure the transfection efficiency (Invitrogen). MiR-218 was expressed in cell lines by transfecting with a Pre-miR$^{TM}$ miR-218 precursor molecule (Ambion). Cell lines were seeded (1.5 X 10$^5$) into 6-well plates and after 24 hr transfected (125 nM per well of HPV siRNAs or 100 nM of Pre-miR$^{TM}$ miR-218 precursor Molecule) using Lipofectamine 2000 (Invitrogen) according to the manufacturer's instructions. Cells were harvested after 72 hr and RNA extractions were carried out as described earlier.

## 4.3    RESULTS

*MiRNA Expression in the Cervix*

The basal expression of miRNAs in the normal cervix was identified using miRNA microarrays. We obtained the median fluorescent intensity of each

miRNA, and those having values of zero in 50% of their corresponding spots (4 spots per array and two repetitions) were considered to be non-expressing. This analysis showed that approximately 220 known human miRNAs out of 328 represented on the array were expressed in the normal cervix (supplementary data, table V). The miRNAs that were most highly expressed in the cervix were miR-145, miR-26a, miR-99a, let-7a, miR-143, let-7b, let-7c, miR-125b, miR-126, and miR-195 in that order.

*Differential Expression of MicroRNAs in Cervical Cell Lines Compared to the Normal Cervix*

We investigated the miRNA expression profile in normal cervical tissue and cervical carcinoma cell lines SiHa and CaSki containing integrated HPV-16 DNA (Meissner, 1999). We also used three clonal derivatives of the W12 cell line derived from a low-grade CIN I lesion (Stanley*, et al.*, 1989): 20863 contains episomal HPV-16 DNA, while 20861 and 201402 contain integrated HPV-16 DNA and show changes in gene expression that mimic neoplastic progression (Alazawi*, et al.*, 2002). The 20863 cell line mimics early stages of HPV infection and is the only well-characterized cell line containing stable HPV-16 episomal DNA. SAM analysis showed that 24 miRNAs, including miR-126, miR-143, miR-145 and miR-195 were underexpressed in all the integrated HPV-16 cervical cell lines compared to the normal cervix (Table 14). Only three miRNAs, miR-182, miR-183 and miR-210, were found to be overexpressed in the integrated HPV-16

cell lines (Table 1). Six miRNAs were underexpressed in the 20863 cell line containing episomal HPV-16 DNA compared to the normal cervix (Table 14). A direct comparison of the miRNA expression profiles of integrated vs. episomal HPV-16 cell lines did not reveal any significant differences (data not shown). This, at least in part, may result from the use of a 2-fold cut-off value and a q value of zero in statistical analysis.

*Differential Expression of MicroRNAs in HPV-16 Positive Cervical Cell Lines Compared to the HPV-Negative Cell Line C-33A*

We compared the expression profile of cell lines containing HPV-16 DNA with that of the only available, well-characterized HPV-negative cervical carcinoma cell line C-33A (Scheffner*, et al.*, 1991). Nine miRNAs in cell lines containing integrated HPV-16 DNA and 10 miRNAs in the episomal HPV-16 cell line 20863 were found to be expressed at much higher levels compared to the C-33A cells (Table 15). Interestingly, miR-218 was the only miRNA that was underexpressed in cell lines containing integrated HPV-16 DNA as compared to both the HPV-negative cell line C-33A and the normal cervix (Tables 14 and 15). This suggested that miR-218 may be specifically affected in the presence of HPV-16.

**TABLE 14.** **miRNAs DIFFERENTIALLY EXPRESSED IN HPV-16 POSITIVE CELL LINES COMPARED TO NORMAL CERVICAL TISSUE**

| miRNA | HPV16 Integrated Fold Change | HPV16 episomal Fold Change |
|---|---|---|
| **Overexpressed** | | |
| hsa_miR_210 | 7.3 | |
| hsa_miR_182 | 6.4 | |
| hsa_miR_183 | 5.1 | |
| **Underexpressed** | | |
| hsa_miR_126 | -14.5 | |
| hsa_miR_145 | -14.1 | -35.5 |
| hsa_miR_451 | -11.3 | |
| ambi_miR_7029 | -9.5 | |
| hsa_miR_195 | -8.3 | -18.2 |
| hsa_miR_143 | -8 | |
| hsa_miR_199b | -7.9 | |
| hsa_miR_133a | -7.6 | -10.1 |
| hsa_miR_368 | -7.6 | -16.6 |
| hsa_miR_1 | -7.2 | |
| hsa_miR_495 | -6.3 | |
| hsa_miR_497 | -6.2 | |
| hsa_miR_133b | -6 | |
| hsa_miR_223 | -5.4 | |
| hsa_miR_146a | -5.1 | |
| hsa_miR_218 | -4.8 | |
| hsa_miR_126_AS | -4.7 | |
| hsa_miR_150 | -4.4 | |
| hsa_miR_376a | -4.2 | |
| hsa_miR_214 | -4.1 | -12.1 |
| hsa_miR_487b | -4.1 | |
| hsa_miR_10b | -3.9 | |
| ambi_miR_5021 | -3.6 | |
| ambi_miR_7070 | -3.5 | |
| hsa_miR_199a | | -14.7 |

**TABLE 15. miRNAs DIFFERENTIALLY EXPRESSED IN HPV-16 POSITIVE CELL LINES COMPARED TO HPV NEGATIVE CELL LINE C-33A**

| miRNA | HPV16 Integrated Fold Change | HPV16 episomal Fold Change |
|---|---|---|
| **Overexpressed** | | |
| hsa_miR_200c | 27.9 | 36 |
| hsa_miR_203 | 23.4 | 14.7 |
| hsa_miR_193b | 21.2 | 10.6 |
| hsa_miR_34a | 10.4 | |
| hsa_miR_31 | 8.4 | 10.9 |
| hsa_miR_210 | 5.7 | |
| hsa_miR_27a | 5.4 | |
| hsa_miR_503 | 5.4 | |
| hsa_miR_27b | 4.9 | |
| hsa_miR_205 | | 50.1 |
| hsa_miR_141 | | 16.8 |
| ambi_miR_13258 | | 8 |
| hsa_miR_200b | | 7 |
| hsa_miR_200a | | 5.3 |
| hsa_miR_224 | | 5.1 |
| **Underexpressed** | | |
| hsa_miR_218 | -7.4 | |

*Differential Expression of MicroRNAs in the HPV-Negative Cell Line C-33A Compared to the Normal Cervix*

The miRNA expression profile of the HPV-negative cell line C-33A showed that 4 miRNAs (miR-143, miR-145, miR-200c and miR-203) were underexpressed as compared to the normal cervix (Supporting Information, Table 16). In general, the downregulation of miRNAs in C-33A was more severe than that observed in the HPV-positive cell lines (Table 14 and 16).

*Validation of the MiRNA Microarray Expression Data by Quantitative Real-Time RT-PCR and Northern Blot Analyses*

The miRNA microarray results were validated by qRT-PCR and Northern blot analyses of 8 representative miRNAs whose expression was either most affected, affected in other types of cancers, or one that appeared to be HPV-specific (miR-218) (Fig. 15*A* and 15*B*). In general, the results of these analyses were consistent with the microarray data, although the fold-changes observed by qRT-PCR analyses were much greater than those obtained with the microarrays. Also, there were differences in individual miRNA levels between some HPV-16 positive cell lines. For example, the expression pattern of miR-200c and miR-205 in SiHa was different than the other HPV-positive cell lines, but similar to that of C-33A (Fig. 15*A* and

**TABLE 16. MiRNAs DIFFERENTIALLY EXPRESSED IN HPV NEGATIVE CELL LINE C-33A COMPARED TO NORMAL CERVICAL TISSUE**

| miRNA | Fold Change |
|---|---|
| **Underexpressed** | |
| hsa_miR_145 | -29.7 |
| hsa_miR_143 | -14.9 |
| hsa_miR_203 | -13.7 |
| hsa_miR_200c | -10 |

# Fig. 15.  CONFIRMATION OF miRNA MICROARRAY EXPRESSION DATA IN VARIOUS CERVICAL CELL LINES AND NORMAL CERVICAL TISSUE
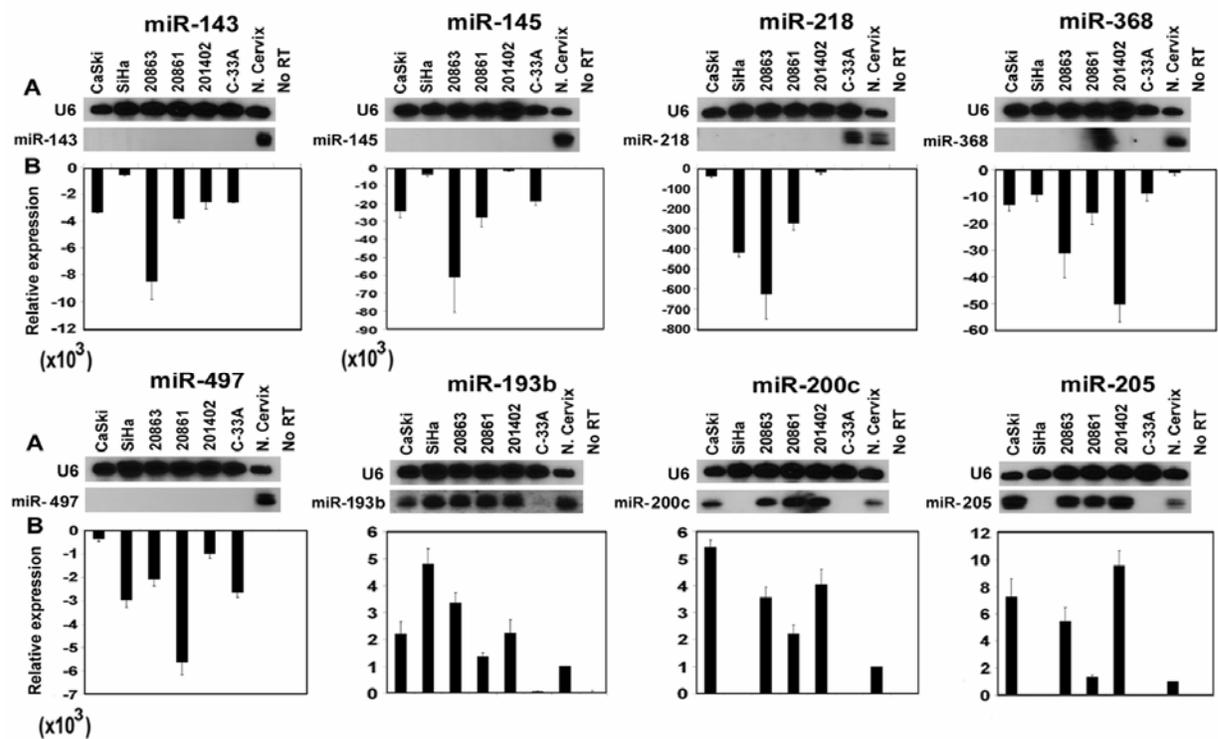


(A) Northern blot analysis. The housekeeping splicing-related small U6 RNA was used as a loading control. (B) Real-time qRT-PCR analysis. All reactions were performed in triplicates and the error bar represents the standard deviation. RNU43 served as the endogenous control for miRNAs.

15*B*). The "band" seen in the Northern blot for miR-368 in the 20861 sample is a gel artifact and does not represent a miRNA signal.

*Clustering Analysis of MiRNA Expression Patterns in Various Cell Lines*

Unsupervised hierarchical clustering analysis revealed two major miRNA expression clusters (Fig. 16). Cluster 1 contained six miRNAs which were expressed to much higher levels in the normal cervix and the HPV-positive cell lines compared to C-33A. The only exception was the HPV-16 positive cell line SiHa which showed a pattern similar to that of C-33A and had low-level expression of the miRNAs in cluster 1. The expression data for miR-200c and miR-205 obtained by Northern blot and qRT-PCR analyses for all the above samples (Fig. 15*B*) were also consistent with those found in cluster 1 (Fig. 16). Cluster 2 contained 28 miRNAs that, in general, showed lower expression in all the cervical cell lines compared to the normal cervix (Fig. 16).

*Expression Patterns of MiRNAs in HPV-16 Positive Cervical Lesions and Cervical Cancer Tissues*

We also analyzed the expression profile of six miRNAs in three HPV-16 positive cervical intraepithelial neoplasia grade III (CIN III) tissues and five HPV-16 positive cervical cancer tissues (CaCx) by qRT-PCR analysis. Due to the

# Fig. 16. HIERARCHICAL CLUSTERING ANALYSIS OF miRNAs IN CERVICAL CELL LINES AND NORMAL CERVICAL TISSUE
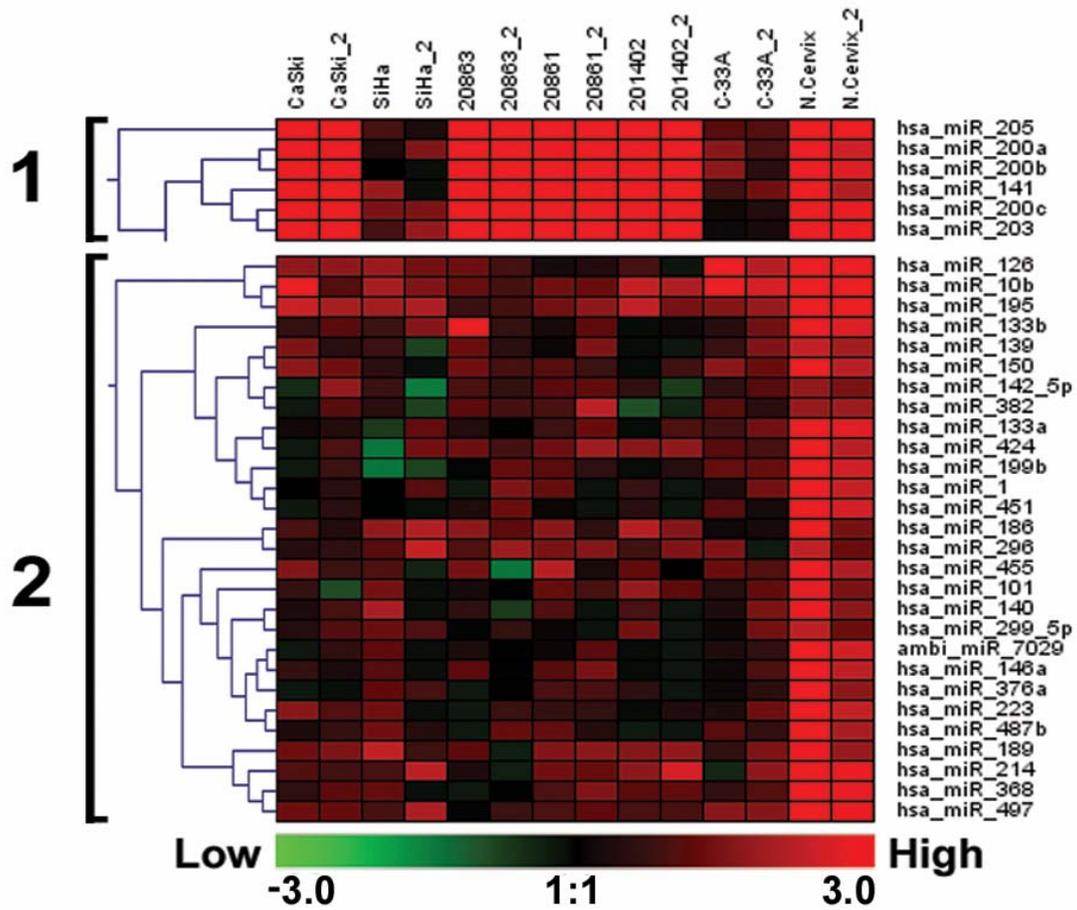


Hierarchical clustering analysis of miRNAs in cervical cell lines and normal cervical tissue. Red and green indicates high or low expression, respectively.

limited amount of RNA available from the tissues, miRNA microarray analysis could not be done. The expression pattern of the above miRNAs in the tissues was generally consistent with that of the HPV-positive cell lines (Fig. 17). Although the relative levels of various miRNAs varied between the individual samples, miRNAs that were underexpressed in the HPV-16 positive cell lines were generally also underexpressed in the HPV-positive tissues compared to the normal cervical tissue (Fig. 17). The CIN III samples generally showed a more limited underexpression of the miRNAs compared to the CaCx tissues, although this correlation was not absolute. The results obtained with the tissues provide further validation of the data obtained with the cervical cell lines. Importantly, miR-218 was found to be underexpressed in all the CIN III and CaCx samples compared to the normal cervix (Fig. 17). Since miR-218 is encoded by an intron of the *SLIT2* tumor suppressor gene (Griffiths-Jones, 2004;Griffiths-Jones*, et al.*, 2006), we tested whether their expression is correlated. Previous analysis of global gene expression (manuscript in preparation) has also shown that *SLIT2* was underexpressed in the HPV-positive cell lines. The qRT-PCR results showed that *SLIT2* expression paralleled that of miR-218, and both of these were underexpressed in the CIN III and CaCx tissues (Fig. 17). These results suggest that the expression of miR-218 and *SLIT2* is similarly affected in HPV-positive lesions and CaCx tissues.

# Fig. 17. EXPRESSION OF miRNAs AND THE *SLIT2* GENE IN CERVICAL TISSUES
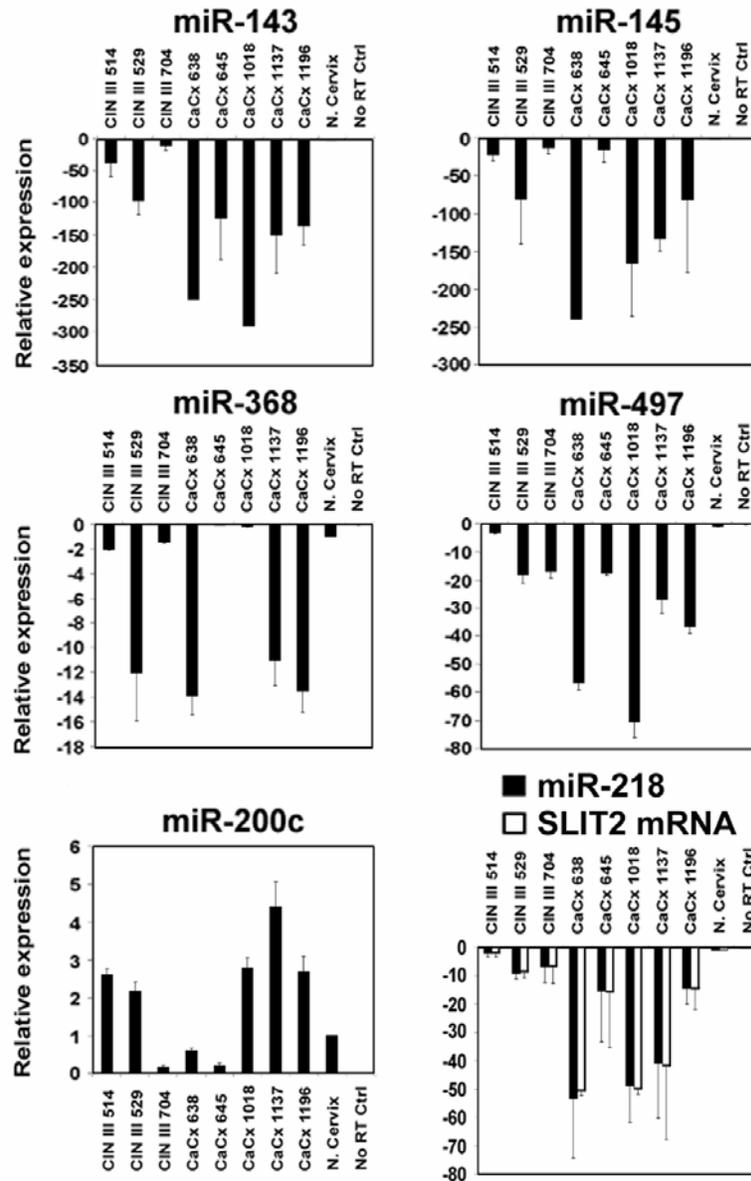


MicroRNAs and *SLIT2* qRT-PCR analysis of three cervical intraepithelial neoplasias type III (CIN III), five cervical carcinomas (CaCx) and the normal cervix. All reactions were performed in triplicates. G3PDH served as the endogenous control for *SLIT2*.

*HPV-16 E6 Oncogene Downregulates MiR-218*

We wished to determine whether E6 and/or E7 expression is directly correlated with reduced expression of miR-218. For this, we utilized the osteosarcoma cell line U2OS (Duensing, *et al.*, 2000) either expressing the HPV-16 E6 or E7 gene, or the control neomycin resistance gene. The qRT-PCR results showed that both miR-218 and *SLIT2* were underexpressed in the E6-expressing U2OS cell line compared to the E7-expressing U2OS and the control U2OS-Neo cell line (Fig. 18*A*). In another approach, the 20861 cell line containing integrated HPV-16 was transfected with HPV-16 E6/E7 siRNAs. Since E6 and E7 are derived from alternative splicing of the same RNA, a specific siRNA for E6 alone could not be used. The E6/E7 siRNAs reduced expression of these genes while increasing the expression of both miR-218 and the *SLIT2* gene in 20861 cells (Fig. 18*B*). These results indicate that the HPV-16 E6 gene is involved in the downregulation of miR-218 and the *SLIT2* gene in HPV-16 positive cell lines.

*Laminin 5 β3 is a Possible Transcriptional Target of MiR-218*

To identify possible miR-218 targets, we compared computationally predicted targets in the miRBase Registry (Griffiths-Jones, 2004;Griffiths-Jones, *et al.*, 2006) with our gene expression data (manuscript in preparation), and then used the program rna22 [http://cbcsrv.watson.ibm.com/rna22.html]. The resulting

141

**Fig. 18. HPV-16 E6 oncogene reduces the expression of miR-218**
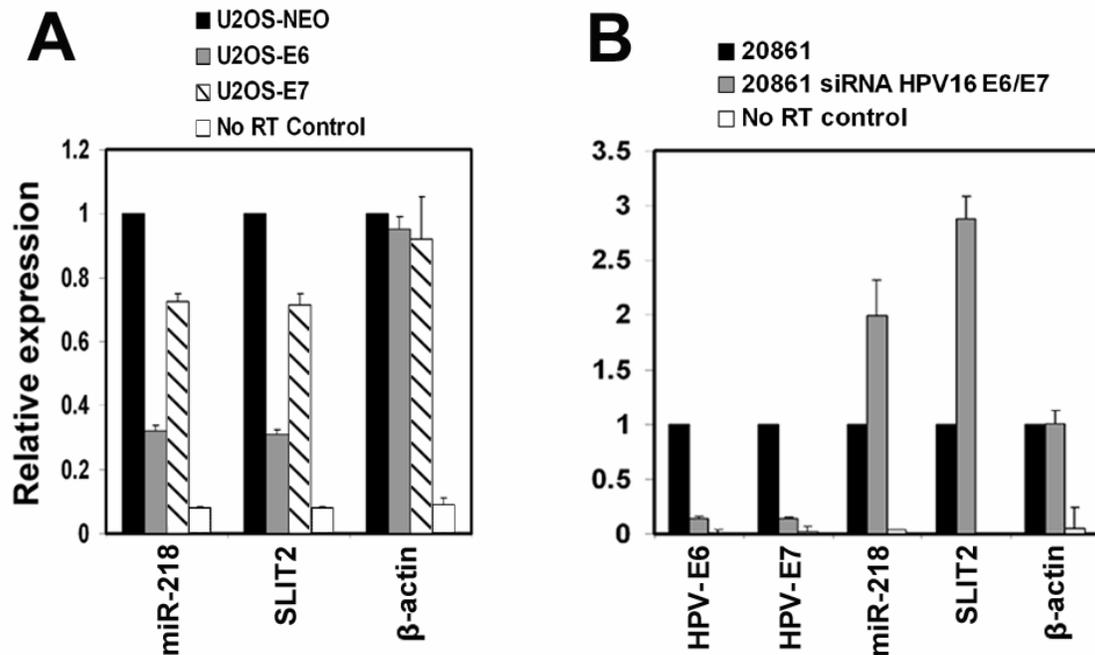


(*A*) qRT-PCR analysis of miR-218, SLIT2, and β-actin in U2OS-Neo, U2OS-E6, and U2OS-E7. (*B*) Expression of HPV-16 E6 and E7, miR-218, *SLIT2*, and β-actin (housekeeping control) in the 20861 cell line with or without RNAi against HPV-16 E6/E7.

6 possible targets were analyzed by qRT-PCR in SiHa and 20861 cell lines transfected with an artificial miR-218 precursor molecule. After confirming the expression of mature miR-218 (data not shown), we found that only the *LAMB3* transcript was significantly underexpressed in miR-218 expressing cells (Fig. 19*A*). We also found that *LAMB3* was underexpressed in the 20861 cell line in the presence of the E6/E7 siRNAs compared to a control oligo (Fig. 19*B*). Furthermore, U2OS-E6 cells showed an increase in the levels of *LAMB3* mRNA as compared to the U2OS-Neo cells (Fig. 19*B*). Taken together, these results demonstrate that miR-218 reduces *LAMB3* expression at the transcriptional level.

## 4.4    DISCUSSION

The microarray data showed that 24 miRNAs were underexpressed and 3 overexpressed in integrated HPV-16 cell lines compared to the normal cervix (Table 14). We also validated these data by Northern blot and qRT-PCR analyses for five miRNAs that were underexpressed (miR-143, miR-145, miR-218, miR-368 and miR-497) and three that were overexpressed (miR-193b, miR-200c and miR-205) in the cell lines compared to the normal cervix (Fig. 15). The probable targets of the above eight miRNAs are shown in supplementary data, table W. Our studies identified 10 underexpressed miRNAs in integrated HPV-16 cell lines that are not known to be altered in any cancers. These included 7 known human miRNAs,

**Fig. 19. EXPRESSION OF LAMB3 IS AFFECTED IN THE**

**PRESENCE OF miR-218**
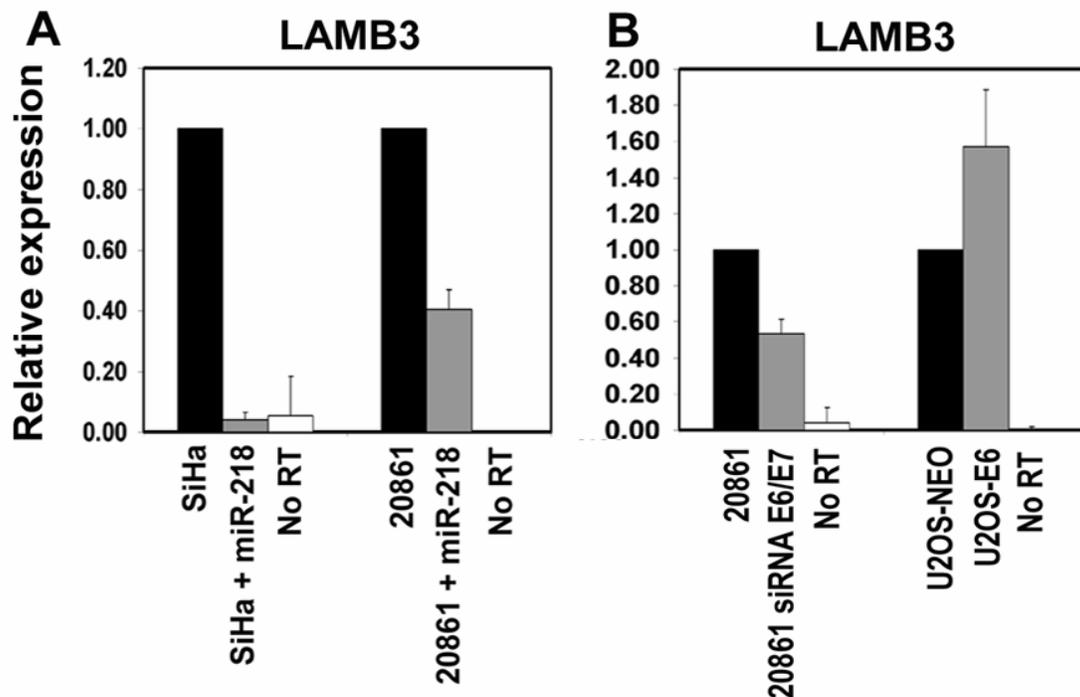


(*A*) qRT-PCR analysis of *LAMB3* in SiHa and 20861 cell lines transfected with a precursor of miR-218. (*B*) Expression of *LAMB3* in the 20861 cell line with or without RNAi against HPV-16 E6/E7, and in the U2OS-Neo, U2OS-E6, and U2OS-E7 cell lines. G3PDH served as the endogenous control for *LAMB3*.

miR-1, miR-133a, miR-126-AS, miR-376a, miR-451, miR-487b, miR-495, and 3 predicted miRNAs, ambi-miR-5021, ambi-miR-7029, and ambi-miR-7070. Many of the differentially expressed miRNAs in integrated HPV-16 cell lines have been previously associated with various types of cancers. For example, miR-143 and miR-145 are underexpressed in colon and breast cancers (Iorio*, et al.*, 2005;Michael*, et al.*, 2003;Volinia*, et al.*, 2006) and they are also found in chromosomal fragile sites in many cancers (Calin*, et al.*, 2004). Since miR-143 and miR-145 are underexpressed in all cervical cell lines including C-33A (Supporting Information, Table 4), our results suggest that they may be important in cervical carcinogenesis independent of HPV infection. MiR-368 is underexpressed in colon cancer cell lines (Bandres*, et al.*, 2006), miR-497 is altered in papillary thyroid carcinomas (Pallante*, et al.*, 2006), and miR-193b is underexpressed in colon cancer cell lines with a Dicer mutation (Cummins*, et al.*, 2006). Interestingly, miR-218 was the only miRNA found to be underexpressed in integrated HPV-16 cell lines compared to both the normal cervix and C-33A cells, suggesting that it may be a specific cellular target of HPV-16. MiR-218 is also underexpressed in several cancers and the DNA encoding miR-218 is also deleted in ovarian, breast, and melanoma cancers (Volinia*, et al.*, 2006;Zhang*, et al.*, 2006). Of the miRNAs overexpressed in integrated HPV-16 cell lines, miR-210 is overexpressed in many epithelial cancers (Volinia*, et al.*, 2006), while miR-182 and miR-183 are overexpressed in colon cancer cell lines (Bandres*, et al.*, 2006). We also found that 6 miRNAs have reduced expression in the episomal

HPV-16 cell line 20863 compared to the normal cervix (Table 14). These miRNAs may represent early targets of HPV-16 infection.

Hierarchical clustering analysis revealed miRNAs that exhibit similar expression patterns in various cell lines. Cluster 1 includes miR-200a and miR-200b which are part of the same miRNA family located on chromosome 1p36.33, while miR-200c and miR-141 are part of another family located on chromosome 12p13.31 (Griffiths-Jones, 2004;Griffiths-Jones, *et al.*, 2006). This suggests that these miRNAs may be subject to common transcriptional regulation. Cluster 2 contained 28 miRNAs that in general had lower expression in all the cervical cell lines compared to the normal cervix (Fig. 16). Two of these miRNAs, miR-368 and miR-497, were also found to be underexpressed in most of CIN III and CaCx tissues (Fig. 17). Interestingly, 11 miRNAs in cluster 2 (miR-126, miR-195, miR-150, miR-142-5p, miR-133a, miR-1, miR-296, miR-101, miR-146a, miR-223 and miR-214) are also underexpressed in leukemia and lymphoma cell lines compared to normal B-cells (Lu, *et al.*, 2005) suggesting that reduced expression of these miRNAs could be a common mechanism during immortalization and/or transformation.

Analysis of 6 miRNAs (miR-143, miR-145, miR-200c, miR-218, miR-368 and miR-497) in CIN III and CaCx tissues by qRT-PCR showed expression patterns that were generally similar to those of the HPV-positive cell lines (Fig. 15 and 17). Importantly, miR-218 was also underexpressed in CIN III and CaCx

tissues (Fig. 17). These results suggest that miR-218 underexpression is likely linked to the process of HPV-associated carcinogenesis *in vivo*. The parallel expression of miR-218 and the *SLIT2* gene (which encodes miR-218) in all the tumor tissues (Fig. 17) suggests that they may be coordinately regulated. Previous studies have shown that the *SLIT2* tumor suppressor gene is frequently inactivated in lung and breast cancers (which have reduced miR-218 levels) (Dallol*, et al.*, 2002;Volinia*, et al.*, 2006). Whether regulation of *SLIT2* plays a role in the pathogenesis of HPV-associated malignancies is currently unknown.

Our results showed that both miR-218 and *SLIT2* were underexpressed in the U2OS-E6 cell line (Fig. 18*A*), suggesting that E6 may contribute to their downregulation. This possibility is also supported by our observations that expression of E6/E7 siRNAs in 20861 cells increases miR-218 and *SLIT2* mRNA levels (Fig. 18*B*). Interestingly, *SLIT2* is a possible target of miR-200c (overexpressed in our HPV-positive cell lines and tissues) (supplementary data, table W). One study has shown that the absence of the p53 gene increases miR-200c expression (Xi*, et al.*, 2006). Thus, it is possible that E6-dependent degradation of the p53 protein results in miR-200c overexpression which in turn reduces the levels of miR-218 and *SLIT2* mRNA.

Our studies showed that introduction of miR-218 or E6/E7 siRNAs into SiHa and/or 20861 cell lines reduces Laminin 5 β3 gene (*LAMB3*) expression (Fig. 19*A* and 19*B*). Also, E6 expression in U2OS cells increases *LAMB3*

expression (Fig. 19*B*) presumably by reducing miR-218 expression. The above studies suggest that miR-218 may regulate *LAMB3* expression at the transcriptional level. LAMB3 protein is part of the polymeric cell surface receptor laminin 5 that is expressed in the basal lamina of the epithelium and is overexpressed in cervical cancers (Kohlberger*, et al.*, 2003;Skyldberg*, et al.*, 1999). LAMB3 increases cell migration and tumorigenicity in SCID mice, and in collaboration with its ligand α6β4-integrin promotes tumorigenesis in human keratinocytes (Calaluce*, et al.*, 2004;Dajee*, et al.*, 2003). A recent study suggests that secreted laminin 5 can be used by HPV as a transient receptor to aid the virus in the infection of basal cells that express α6β4-integrin (Culp*, et al.*, 2006). Thus, downregulation of miR-218 by E6 and the consequent overexpression of *LAMB3* may promote viral infection of the surrounding tissue and contribute to eventual tumorigenesis.

**Chapter 5**

**GENERAL DISCUSSION**

Cervical and oropharyngeal cancers represent an important health problem worldwide. A better understanding of the interactions between HPV and these tumor cells is important for the identification of the molecular mechanisms involved in HPV-associated diseases as well as for the identification of diagnostic and prognostic markers and approaches towards treatment. In these studies, we used high-density microarrays (Affymetrix U133A GeneChip® expression microarray containing 22,215 human transcripts) to analyze gene expression differences in several HPV-positive and HPV-negative cervical cell lines and SCCHN tissues in order to identify the pathways altered in HPV-positive cells as a consequence of viral gene expression, integration, and other downstream events.

Infection with high-risk HPVs has been implicated in a subset of squamous cell carcinoma of the head and neck (SCCHN). Part of our study was to compare the cellular gene expression profiles of HPV16-positive and HPV-negative oropharyngeal carcinomas with those of the normal oral epithelium. Although the number of specimens used in our study was small, a unique feature of this study was that we have exclusively utilized tissues derived from the oropharyngeal site. While HPV infection can be identified throughout tumors of the head and neck, a disproportionately high incidence is observed in the oropharynx (approximately 50%). Thus, utilizing tumors from this region reduces sample heterogeneity and eliminates variability due to different tumor sub-sites which has been observed in previous studies. After confirming the expression of

the HPV-16 E6 and E7 genes in the HPV-positive SCCHN, we proceeded with the global gene expression analysis. We found that 228 genes were upregulated and 169 downregulated in HPV-positive SCCHN samples as compared to the normal oral tissues (Table 2). The upregulated genes included those involved in DNA replication and cell cycle regulation such as *CDKN2A* (*p16INK4a*), *MCM2, PCNA, TOP2A* and *RBL1* that have been previously identified to be affected in HPV-infected cervical and oral carcinomas (Li*, et al.*, 2004;Ren*, et al.*, 2002;Santin*, et al.*, 2005). Several studies have shown either direct or indirect regulation of these genes by the tumour suppressor protein p53 (*RBL1, PCNA, TOP2A*) or by E2F transcription factors (*CDKN2A* and *MCM2*) (Ren*, et al.*, 2002;Wei*, et al.*, 2006). One possible explanation for the transcriptional deregulation of this group of genes may be the degradation of p53 and pRB (that bind and repress the activity of the family of E2F transcription factors) by the viral oncogenes E6 and E7, respectively (Kannan*, et al.*, 2001;Ren*, et al.*, 2002;Wei*, et al.*, 2006;Zhao*, et al.*, 2000). These data support the notion that HPV-positive SCCHN has a specific transcription profile partly dependent on E6 and E7 expression. When we compared the HPV-negative SCCHN against normal oropharyngeal squamous mucosa, we found 79 upregulated and 83 downregulated genes (Table 3). We found upregulated genes such as *AKR1C3* and *SNAPC1* and downregulated genes such as *CLU* and *RBM3* that were previously not known to be differentially expressed in SCCHN. *AKR1C3* is a hydroxysteroid dehydrogenases involved in the regulation of local concentration of androgens and estrogens in hormone-dependent tissues such as prostate,

breast and endometrium (Penning, *et al.*, 2006). *AKR1C3* has been found to be overexpressed at both the mRNA and protein levels in prostate cancer specifically in the stromal cells (Penning, *et al.*, 2006). *SNAPC1* is a subunit of the TBP-TAF complex involved in the transcription of both RNA polymerase II and III dependent small nuclear RNA genes. *SNAPC1* has been linked to the tumorigenesis of breast cancers (Xie, *et al.*, 2002). *CLU* is a multifunctional gene which is involved in spermatogenesis and in the control of the immunological complement cascade (Scaltriti, *et al.*, 2004). It has recently been shown that *CLU* is downregulated in low- and high-grade human prostate cancers and it may be involved in supporting cell survival and the induction of programmed cell death (Scaltriti, *et al.*, 2004). *RBM3*, a glycine-rich RNA-binding protein induced by cold-stress, has recently been found to have apoptotic modulatory capabilities with a possible role as a tumor suppressor gene (Sutherland, *et al.*, 2005). Interestingly, there is evidence that *RBM3* can change the expression of some microRNAs and enhance global protein synthesis suggesting the involvement of a homeostatic mechanism that regulates global levels of protein synthesis under normal and cold-stress conditions (Dresios, *et al.*, 2005). Interesting data were also obtained when we compared the HPV-positive SCCHN against HPV-negative SCCHN. We found that 124 genes were upregulated and 42 genes were downregulated in HPV-positive SCCHN as compared to HPV-negative SCCHN (Table 4). Among this group of genes, we found a subset of genes (41 upregulated and 19 downregulated) that were also differentially expressed in HPV-positive SCCHN as compared to the normal oropharyngeal mucosa (Table

5). It is likely that the differential expression of these 60 genes is a consequence of the presence of HPV-16. This group of genes included the upregulated genes *SYCP2, RFC5, ZNF238* and *DNMT1*, and the downregulated genes *KLK8* and *CRABP2. SYCP2* is part of the synaptonemal complex involved in forming lateral elements and cross bridges that contribute to pairing of sister chromatids during meiosis and probably are involved in the interaction between chromatin and nuclear envelope (Shakib*, et al.*, 2005). There is evidence that HPV infection in cervical cells promotes morphological changes in the nuclear membrane as well as condensation of chromatin attached to the inner part of the nuclear membrane (Bollmann*, et al.*, 2005). Thus, it is possible that *SYCP2* may be involved in chromatin-nuclear envelope interactions in HPV-positive SCCHN. A recent study by Slebos and colleagues also reported up-regulation of the *SYCP2* in HPV-positive SCCHN as compared to HPV-negative SCCHN, supporting the idea that this gene may play important roles in the carcinogenesis of HPV-related SCCHN (Slebos*, et al.*, 2006). The transcription factor *ZNF238* (also known as *RP58*) is a component of a C2H2-type DNA-binding zinc finger protein that acts as a repressor of transcriptionally silent heterochromatin regions. Interestingly, the DNA methyltransferase protein Dnmt3a directly interacts with *ZNF238* and promotes the repression of specific genes by histone deacetylation independently of its methyltransferase activity (Fuks*, et al.*, 2001). The overexpression of *ZNF238* could explain the downregulation of several genes in HPV-positive SCCHN as a consequence of chromatin condensation by histone deacetylation. *CRABP2* is a transcription factor that functions as a cytoplasmic

retinoic acid binding protein that is highly expressed in human skin. Similar to our results showing its downregulation in HPV-positive SCCHN, *CRABP2* is also downregulated in prostate cancer (Okuducu*, et al.*, 2005). In order to understand why some of these genes are differentially expressed in the HPV-positive SCCHN, we used an unsupervised hierarchical clustering analysis. We were able to identify a group of genes in five different clusters with significant variation between HPV-positive SCCHN as compared to HPV-negative SCCHN as well as normal oropharyngeal epithelium (Fig. 7). Some of the genes found to be differentially expressed in our microarray analysis were also found in these clusters such as *ZNF238, DNMT1, MCM3, AKR1C3, GRB10, TYK2* and *SART3* (underlined in Fig. 7). After detecting these five different clusters, we identified the localization of these genes to particular chromosomes (Fig. 8). Two regions in chromosome 1 (p34-p36) and 12 (q21-q240) showed a significant number of differentially expressed genes, particularly upregulated genes in HPV-positive SCCHN. Interestingly, there are two previous publications showing chromosomal aberrations and amplification in chromosome 1 and chromosome 12q24 in HPV-positive cervical intraepithelial neoplasias and head and neck carcinomas, respectively. This may suggest that some of these chromosomal regions are important in the carcinogenesis of HPV-positive SCCHN. We also attempted to identify possible transcription factor binding sites in the promoter regions of various gene clusters that could be involved in regulating gene expression. After obtaining the promoter sequences of all the genes from cluster 4 and 5 (genes mostly downregulated in HPV-positive SCCHN), we found that all these genes

154

shared the binding sequence for the transcription factor GATA3 (data not shown). Interestingly, a recent study showed downregulation of GATA3 mRNA by differential display in HPV-immortalized keratinocytes, HG-SIL and cervical carcinomas suggesting that this is a late event in cervical tumorigenesis (Steenbergen, *et al.*, 2002). This data suggests that the absence of GATA3 could be responsible for the down-regulation of these cluster of genes in HPV-positive SCCHN. In summary, our data demonstrates specific changes in cellular gene expression profiles in HPV-positive SCCHN which also express the viral oncogenes.

In another study, we analyzed the global gene expression profile of HPV-positive cervical cell lines in comparison with one HPV-negative cervical cell line named C-33A and normal cervical tissue. After confirming appropriate expression of the viral genes and the physical state of the HPV DNA in various cell lines, we analyzed the cellular gene expression profiles of these cervical cell lines. Our results showed that 272 genes were upregulated while 605 were downregulated in all the HPV-positive cell lines compared to the normal cervix tissue. A representative list of differentially expressed genes is shown in Table 8. We then compared the global gene expression between all the HPV-positive cervical cell lines and the HPV-negative cervical carcinoma cell line C-33A. We found that 305 genes were upregulated and 232 genes were downregulated in all the HPV-positive cell lines compared to C-33A (Table 9). In order to identify cellular genes whose expression might be specifically altered due to the

presence of HPV sequences, we compared the data obtained for HPV-positive cell lines with those of both the normal cervix and the C-33A cell line. This analysis revealed that a total of 20 genes were upregulated and 37 downregulated in the HPV-positive cell lines as compared to both the normal cervix and C-33A (Table 10). Some of the differentially expressed genes found in our analysis have been previously described in other cervical cancer microarray studies such as TFAP2A (involved in transcription regulation) and IGSF4 (may involved in cell-cell adhesion). Other genes such as CPA4, NEFH, LAT1, SPRY1, ID4 and NAP1L3 have not been previously been shown to de differentially expressed in cervical cancer cell lines. Nuclear organization and biogenesis CPA4 gene is expressed at extremely low levels in human adult tissues but is abundant in prostate cancer cell lines, suggesting that it may modulate the function of peptide hormones that play an essential role in the growth and differentiation of prostate epithelial cells (Kayashima, *et al.*, 2003). It may play a similar role in the growth of HPV-positive cervical epithelial cells. The L-type amino acid transporter 1 (LAT1, also known as SLC7A5) found to be upregulated in all the HPV-positive cell lines compared to both the normal cervix and HPV-negative C-33A cell line (Table 10 and Fig. 13), is an amino acid transporter essential for growth and proliferation. The y+ system L-amino acid transporters are responsible for the $Na^+$-independent transport of neutral amino acids, including several essential amino acids. LAT1 has been found to be overexpressed in astrocytic tumors suggesting that some tumors including cervical cancer may require more nutrient transport for growth and cell

proliferation. Another gene found to be downregulated in our study is the nucleosome assembly protein 1-like 3 (NAP1L3) gene. NAP1L3 is part of the nuclesome assembly complex that probably plays a role in gene expression during brain development. Interestingly, NAP1L3 has been found to be down-regulated in uterus carcinomas suggesting that the disruption of this gene could be frequent in tumors of the female reproductive epithelium. To verify that LAT-1 and NAP1L3 are not affected in HPV-positive cervical cell line as a consequence of *in vitro* immortalization or growth in tissue culture, we carried out *in situ* hybridization to detect LAT1 and NAP1L3 mRNA in tissue-arrays containing ~50 different cervical tumor samples and 4 normal cervix controls per slide. Our data showed that similar to the cervical carcinoma cell lines, these genes were also differentially expressed in cervical carcinoma tissues suggesting a possible role in the carcinogenesis of the cervix. Because the integration of HPV DNA usually occurs during the progression of HPV infection and the DNA is found to be mostly integrated in high-grade lesions and cervical cancers (zur Hausen, 2002), we also analyzed the differences in cellular gene expression profiles between cervical carcinoma cell lines containing integrated HPV DNAs (CaSki, SiHa, 201402, 20861 and HeLa) and cell lines containing episomal HPV DNA that mimic low-grade lesions (20863 and L-18). Our data showed that 52 genes were upregulated while 47 were downregulated in cell lines containing integrated HPV DNA (Table 11). The upregulated genes in HPV-16 or HPV-18 integrated cell lines included those involved in the regulation of mitosis (CENPF) and cell growth and anti-adhesion (PODXL). The downregulated genes included those involved

in host defense response like GATA3. The CENPF gene that encodes the centromere protein F, 350/400ka (mitosin) may de-regulate kinetochore assembly at the G2/M checkpoint and facilitate defective chromosome segregation, promoting cervical carcinogenesis (de la Guardia, *et al.*, 2001). The upregulation of CENPF in cell lines containing integrated HPV DNA is consistent with E7 expression and chromosomal instability associated with HPV integration (Pett, *et al.*, 2004). Our data also showed that the PODXL (mucin-like surface sialoglycoprotein podocalyxin-like) gene is upregulated in cell lines containing integrated HPV DNA. PODXL is an integral membrane glycoprotein that has different cell adhesion properties in different tissues and is repressed by the p53 protein at the transcriptional level (Stanhope-Baker, *et al.*, 2004). Increased expression of E6 in cell lines with integrated HPV DNA causes p53 degradation and could result in an increase in PODXL expression (zur Hausen, 2002). One gene found downregulated in HPV-integrated cell lines compared to HPV-episomal cel lines was GATA3 (essential transcription factor GATA binding protein 3). GATA3 is part of the GATA family of zinc-finger transcription factors that are involved in differentiation and development of different tissues in vertebrates (Lim, *et al.*, 2000). A recent study showed downregulation of GATA3 mRNA by differential display in HPV-immortalized keratinocytes, HG-SIL and cervical carcinomas suggesting that this is a late event in cervical tumorigenesis (Steenbergen, *et al.*, 2002). The increased expression of GATA3 may be co-related with the integration of HPV DNA into the host genome and progression of cervical cancer. Interestingly, two upregulated genes identified in our study BST2

(involved in vesicle coat assembly) and TOP2A (involved in DNA topological changes) have previously been reported by Alazawi et al. comparing the episomal HPV cell line 20863 and its subsequent passages lacking feeder cells (which is expected to result in the integration of HPV DNA) (Alazawi*, et al.*, 2002). These genes could be important in the process of HPV integration into the host genome and consequently in cervical carcinogenesis. Although both HPV types 16 and 18 are high-risk viruses, HPV-16 is commonly associated with SCC, whereas HPV-18 is predominantly associated with adenocarcinoma (Burk*, et al.*, 2003). In order to identify differences in host gene expression when they are infected by one of these two high-risk HPVs, we compared the gene expression profiles in cell lines containing HPV-18 with those in cell lines containing HPV-16 DNA. We found that 36 genes were upregulated while 3 were downregulated in HPV-18 positive cells compared to the HPV-16 positive cells (Table 12). Of note, the keratin 8 (KRT8) and keratin 18 (KRT18) genes involved in cytoskeleton organization were found to be upregulated in the presence of HPV-18 compared to cells containing HPV-16. The relatively few genes that were found to be differentially expressed in this analysis may suggest the involvement of predominantly similar pathways during the progression of cervical carcinogenesis induced by the high-risk HPV-16 and HPV-18. In summary, our data demonstrate specific changes in cellular gene expression profiles in HPV-positive cervical cell lines that could be used as a basis for identifying molecular mechanisms involved in the pathogenesis of HPV-associated cervical cancers.

After identifying the global gene expression profiles in HPV-positive cervical cell lines and HPV-positive SCCHN, we compared the above two sets of data to identify genes that may be differentially expressed in both of these types of HPV-positive cancers, and therefore may provide a molecular signature of HPV infection. When we compared the data obtained with "HPV-positive cervical cell lines vs normal cervical tissue" against "HPV-positive SCCHN vs normal oral tissue", we found that 20 upregulated genes and 27 downregulated genes were shared between these cancers (Table 17). Several of the upregulated genes such as CKS1B, CDC6, CDKN2A, FEN1, MAD2L1, MCM7 and RNASEH2A are involved in cell cycle and DNA replication, while many of the downregulated genes are involved in immune response (CCL14, CCL15, CRISP3, LTF and SPINK5), cell differentiation (DMBT1, SCEL and MAL), and cell cycle arrest (PCAF). Future studies of these particular genes will be important to validate this data and understand the influence of HPV presence in the regulation of host genes that could be involved in the carcinogenesis of these epitheliums.

We were also interested in identifying the mechanisms by which HPVs may affect cellular gene expression. Recently, the discovery of microRNAs which are small non-coding RNAs that can regulate gene expression at a posttranscriptional level have revolutionized our understanding of gene regulation in eukaryotic cells. Therefore, we undertook a study to identify human microRNAs which may be targeted by HPVs and whose expression may be

**Table 17.** DIFFERENTIALLY EXPRESSED GENES FOUND IN COMMON BETWEEN "THE HPV-POSITIVE CERVICAL CELL LINES VS NORMAL CERVICAL TISSUE DATA" AND "THE HPV-POSITIVE SCCHN VS NORMAL ORAL TISSUE DATA"

| Gene Title | Gene Symbol | GO Biological Process Description |
|---|---|---|
| **Upregulated genes** | | |
| CDC28 protein kinase regulatory subunit 1B | CKS1B | cell cycle /// cytokinesis |
| CDC6 cell division cycle 6 homolog (S. cerevisiae) | CDC6 | DNA replication checkpoint /// cell cycle /// cytokinesis /// mitosis |
| cofactor of BRCA1 | COBRA1 | regulation of transcription, DNA-dependent |
| cyclin B2 | CCNB2 | cytokinesis /// mitosis /// regulation of cell cycle |
| cyclin-dependent kinase inhibitor 2A | CDKN2A | cell cycle checkpoint /// negative regulation of cell proliferation |
| dual specificity phosphatase 14 | DUSP14 | protein amino acid dephosphorylation |
| flap structure-specific endonuclease 1 | FEN1 | DNA replication /// UV protection /// double-strand break repair |
| frizzled homolog 2 (Drosophila) | FZD2 | Wnt receptor signaling pathway /// development /// establishment of tissue polarity |
| glutathione S-transferase omega 1 | GSTO1 | metabolism |
| MAD2 mitotic arrest deficient-like 1 (yeast) | MAD2L1 | cell cycle /// mitosis /// mitotic checkpoint |
| maternal embryonic leucine zipper kinase | MELK | protein amino acid phosphorylation |
| MCM7 minichromosome maintenance deficient 7 | MCM7 | DNA replication initiation /// cell cycle /// regulation of transcription |
| Opa-interacting protein 5 | OIP5 | cell communication |
| pescadillo homolog 1, containing BRCT domain (zebrafish) | PES1 | morphogenesis |
| proteasome (prosome, macropain) subunit, beta type, 5 | PSMB5 | ubiquitin-dependent protein catabolism |
| ribonuclease H2, large subunit | RNASEH2A | DNA replication /// RNA catabolism |
| S-phase kinase-associated protein 2 (p45) | SKP2 | G1/S transition of mitotic cell cycle /// cell proliferation /// ubiquitin cycle |
| thioesterase superfamily member 2 | THEM2 | --- |
| thymidine kinase 1, soluble | TK1 | DNA metabolism |
| ubiquitin-conjugating enzyme E2C | UBE2C | cell cycle /// cytokinesis /// mitosis /// positive regulation of cell proliferation |
| **Downregulated genes** | | |
| aldehyde dehydrogenase 1 family, member A1 | ALDH1A1 | aldehyde metabolism |
| amylase, alpha 2A; pancreatic | AMY2A | carbohydrate metabolism |
| chemokine (C-C motif) ligand 14 /// chemokine (C-C motif) ligand 15 | CCL14 /// CCL15 | immune response /// cell-cell signaling /// chemotaxis |
| chromosome 6 open reading frame 29 | C6orf29 | --- |
| cysteine-rich secretory protein 3 | CRISP3 | cell-cell adhesion /// innate immune response /// spermatogenesis |
| deleted in malignant brain tumors 1 | DMBT1 | epithelial cell differentiation /// innate immune response |
| flavin containing monooxygenase 3 | FMO3 | electron transport |

| | | |
|---|---|---|
| glycine amidinotransferase | GATM | biosynthesis /// creatine biosynthesis |
| glycoprotein M6A | GPM6A | --- |
| homeodomain-only protein /// homeodomain-only protein | HOP | regulation of transcription, DNA-dependent |
| kallikrein 13 | KLK13 | proteolysis and peptidolysis |
| lactotransferrin | LTF | humoral immune response /// iron ion homeostasis /// iron ion transport |
| LIM domain only 2 (rhombotin-like 1) | LMO2 | cell growth and/or maintenance /// development |
| loss of heterozygosity, 11, chromosomal region 2, gene A | LOH11CR2A | --- |
| mal, T-cell differentiation protein | MAL | apical protein localization /// cell differentiation /// induction of apoptosis |
| multimerin 1 | MMRN1 | blood coagulation /// cell adhesion |
| p300/CBP-associated factor | PCAF | cell cycle arrest /// chromatin remodeling /// protein amino acid acetylation |
| paternally expressed 3 | PEG3 | --- |
| prominin 1 | PROM1 | visual perception |
| sciellin | SCEL | epidermis development |
| selenium binding protein 1 | SELENBP1 | --- |
| serine protease inhibitor, Kazal type 4 | SPINK4 | --- |
| serine protease inhibitor, Kazal type 5 | SPINK5 | anti-inflammatory response /// antimicrobial humoral response (sensu Vertebrata) |
| transcobalamin I | TCN1 | cobalamin transport /// cobalt ion transport /// transport |
| tryptase, alpha | TPS1 | defense response /// proteolysis and peptidolysis |
| zinc finger protein 254 /// zinc finger protein 539 | ZNF254 /// ZNF539 | negative regulation of transcription from Pol II promoter |
| zinc finger protein 91 (HPF7, HTF10) | ZNF91 | carbohydrate metabolism /// regulation of transcription, DNA-dependent |

affected by the viral oncogenes.  First, by using miRNA microarrays our studies demonstrate for the first time that approximately 220 miRNAs are expressed in the normal cervix (Supplementary data, table V). The ten most highly expressed miRNAs in the cervix were miR-145, mir-26a, miR-99a, let-7a, miR-143, let-7b, let-7c, miR-125b, miR-126, and miR-195. These miRNAs may be important for the maintenance of a certain phenotype in the cervical tissue.  After this analysis, we compared the miRNA profiles in five HPV-16 positive cervical cell lines (four HPV-16 integrated and one HPV-16 episomal cell lines) with that of the normal cervical tissue. The microarray data showed that 24 human miRNAs were underexpressed and 3 overexpressed in integrated HPV-16 cell lines compared to the normal cervix (Table 14). Many of the differentially expressed miRNAs in integrated HPV-16 cell lines have previously been found to be associated with various types of cancers.  MiR-143 and miR-145 are underexpressed in colon and breast cancers (Iorio*, et al.*, 2005;Michael*, et al.*, 2003;Volinia*, et al.*, 2006). Also, several miRNAs are located in chromosomal fragile sites, such as miR-195, and 199b (Calin*, et al.*, 2004).  Of the miRNAs that were found to be overexpressed in integrated HPV-16 cell lines, miR-182 and miR-210 are expressed at low levels and miR-183 is not expressed in the normal cervix (Supplementary data, Table V).  MiR-210 is also known to be overexpressed in breast, colon, lung, pancreatic, prostate, and stomach cancers (Volinia*, et al.*, 2006), while miR-182 and miR-183 are overexpressed in colon cancer cell lines (Bandres*, et al.*, 2006).  We also found that 6 miRNAs, miR-145, miR-195, miR-368, miR-199a, miR-214, and miR-133a have reduced expression in the

163

episomal HPV-16 cervical cell line 20863 as compared to the normal cervix (Table 1). Changes in miRNA expression in the episomal cell line may represent early targets of HPV-16 infection. We also compared the miRNA expression profiles of HPV-16 positive cervical cell lines against the HPV-negative cervical carcinoma cell line C-33A. This cell line contains a number of mutations including those in p53 that most likely contribute to its carcinogenic phenotype (Scheffner, *et al.*, 1991).   We performed this comparison in order to filter the miRNAs that may be differentially expressed due to the presence of HPV-16. We found that 9 miRNAs exhibit higher expression and one miRNA, miR-218, exhibits lower expression in the integrated HPV-16 cell lines as compared to the C-33A cell line (Table 2). Importantly, miR-218 is underexpressed in HPV-16 positive cell lines as compared to both the normal cervix and C-33A cells.   MiR-218 is also underexpressed in breast, colon, lung, pancreatic, prostate, and stomach cancers (Volinia*, et al.*, 2006). A comparison of miRNA expression in C-33A cells with that of the normal cervix showed that 4 miRNAs, miR-145, miR-143, miR-203, and miR-200c were significantly underexpressed in C-33A cells (Table 16). Since miR-143 and miR-145 were also underexpressed in all the HPV-positive cell lines, our results suggest that these miRNAs may be important in cervical carcinogenesis independent of HPV infection. Hierarchical clustering analysis was done in order to identify specific miRNA expression patterns in the cervical cell lines and normal cervix (Fig. 16).   Cluster 1 contained six miRNAs which were expressed to much lower levels in the HPV-negative C-33A cell line as compared to the normal cervix as well as the HPV-positive cell lines.   The only

exception was the HPV-16 positive cervical carcinoma cell line SiHa which showed a pattern similar to that of C-33A and had low-level expression of the miRNAs in cluster 1. Of the miRNAs in cluster 1, miR-200a and miR-200b are part of the same miRNA family localized on chromosome 1 p36.33, while miR-200c and miR-141 are part of another miRNA family localized on chromosome 12 p13.31 (47, 48). This pattern suggests that miRNAs of the above two families may be subject to a common transcriptional regulation even though they are located in different chromosomal regions. Cluster 2 contained 28 miRNAs that in general had lower expression in all the cervical cell lines as compared to the normal cervix (Fig. 16). Interestingly, 11 miRNAs in cluster 2 (miR-126, miR-195, miR-150, miR-142_5p, miR-133a, miR-1, miR-296, miR-101, miR-146a, miR-223 and miR-214) were also found to be overexpressed in the normal B-cells as compared to leukemia and lymphoma cell lines (23) suggesting that reduced expression of these miRNAs could be a common mechanism during immortalization and/or transformation. We also utilized three CIN III tissues and five cervical cancer tissue samples in our studies to rule out the possibility that deregulation of miRNAs observed in the cell lines is simply a result of their *in vitro* growth in culture. We performed qRT-PCR analysis of five miRNAs, miR-143, miR-145, miR-218, miR-368 and miR-497 that were underexpressed in cervical cell lines, and miR-200c that was overexpressed in the cell lines. The results obtained with the CIN III and cervical cancer tissues (Fig. 17) were generally consistent with the data obtained with the cell lines (Fig. 15). Importantly, miR-218 which is underexpressed in integrated HPV-16 cell lines

165

compared to both the normal cervix and C-33A, was also underexpressed in the CIN III and cervical cancer tissues (Fig. 17). These results suggest that changes in miR-218 expression are likely linked to the process of HPV-associated cervical carcinogenesis and not to other abnormalities that may accumulate in immortalized cells *in vitro*. It is known that miR-218 is encoded by an intron of the SLIT2 tumor suppressor gene (Dallol*, et al.*, 2002). qRT-PCR analysis of the tissue samples showed that SLIT2 expression was also reduced in all the tumor tissues (Fig. 3). Thus, the expression of miR-218 parallels that of SLIT2 suggesting that miR-218 production may be regulated through the promoter of the tumor suppressor gene SLIT2. Since the HPV E6 and E7 oncogenes play important roles in viral-associated carcinogenesis, we tested whether their expression affects the cellular levels of miR-218 and SLIT2 mRNA. For this, we utilized the osteosarcoma cell line U2OS that has previously been used successfully to study the downstream effects of E6 and E7 (Duensing*, et al.*, 2000). Utilizing the U2OS-Neo cell line, or its derivatives expressing the HPV-16 E6 or E7 gene, we found that both miR-218 and SLIT2 were underexpressed in the E6-expressing U2OS cell line as compared to the control U2OS-Neo by qRT-PCR (Fig. 18*A*). The levels of miR-218 or SLIT2 were not significantly affected in E7-expressing U2OS cells. These results suggest that the HPV-16 E6 gene may be involved in the downregulation of miR-218 and SLIT2 gene. We further studied the effect of E6 on miR-218 and SLIT2 expression by using siRNA against the HPV-16 E6/E7 genes. Transfection of the integrated HPV-16 cell line 20861 with an siRNA directed against E6/E7 reduced their expression

significantly and resulted in a concomitant increase in the levels of miR-218 and the SLIT2 mRNA as shown by qRT-PCR (Fig. 18*B*). These results suggest that miR-218 and SLIT2 expression may either be linked to one or more pathways known to be affected by E6 expression (for example, the p53 pathways) or an unknown pathway that targets miR-218. Interestingly, one of the possible mRNA targets of miR-200c (generally overexpressed in HPV-positive cell lines and tissues in our analysis) is the SLIT2 gene. A recent study has also shown that the p53 tumor suppressor gene downregulates miR-220c expression (Xi*, et al.*, 2006). Thus, it is possible that E6-dependent degradation of the p53 protein results in miR-200c overexpression which in turn reduces the levels of miR-218 and SLIT2 mRNA. Future studies should identify the precise mechanism by which E6 affects the levels of these RNAs.

Finally, to investigate the potential involvement of miR-218 in the carcinogenesis of the cervix, we attempted to identify possible mRNA targets of miRNA-218. After using two different programs to identify possible targets of miRNAs, we found that six mRNA that showed overexpression in our previous study using high-density microarrays could be possible targets of miR-218 at the transcriptional level. Out of these six possible targets, only the expression of the Laminin 5 β3 gene (LAMB3) was reduced when the SiHa and 20861 cell lines containing integrated HPV-16 DNA were transfected with an artificial miR-218 precursor (Fig. 19*A*). We also found that the 20861 cell line transfected with the E6/E7 siRNA expressed lower levels of LAMB3 (Fig. 19*B*), further suggesting

that miR-218 may target LAMB3. Also, the expression of E6 in U2OS-Neo cells resulted in an increase in LAMB3 expression (Fig. 19*B*). This effect on LAMB3 was presumably a consequence of reduced expression of miR-218 in the presence of E6. The above studies also suggest that miR-218 may regulate the expression of its target gene LAMB3 at the transcriptional level. LAMB3 is part of the polymeric transmembrane cell surface receptor named laminin 5 (consisting of 3 different chains, α, β and γ) that is expressed in the basal lamina of the epithelium, and it is overexpressed in cervical cancers (Kohlberger*, et al.*, 2003;Skyldberg*, et al.*, 1999). LAMB3 has been found to increase cell migration and tumorigenicity in SCID mice, and its expression along with that of one of its ligands α6β4-integrin (a component of the hemidesmosome structure of basal epithelial cells) has been shown to be necessary for tumorigenesis in human keratinocytes with mutations in laminin 5 and β4-integrin genes (Calaluce*, et al.*, 2004;Dajee*, et al.*, 2003). A recent study has shown that secreted laminin 5 can be used by HPV as a transient receptor to help the virus in the infection of basal cells that express α6β4-integrin (Culp*, et al.*, 2006). Thus, downregulation of miR-218 by the HPV E6 oncogene and the consequent overexpression of LAMB3 may promote viral infection of the surrounding tissue and contribute to eventual tumorigenesis. In summary, our study demonstrate specific changes in miRNA expression profiles in HPV-positive cervical cell lines and tumors that some of these could be a consequence of the expression of the HPV viral oncogenes.

## 5.1    FUTURE DIRECTIONS

The results of studies described in this thesis raise several important issues that can be pursued in the future to get a better understanding of the mechanisms by which HPVs may alter the expression of cellular genes, including those encoding miRNAs.

One of the next steps after the study of global gene expression in HPV-positive SCCHN could be the use of tissue-arrays in order to get further validation of the differentially expressed genes. These experiments may reveal how frequently genes such as SYCP2 and KLK8 are affected in HPV-positive SCCHN, and if there is any correlation with tumor behavior (invasion and metastasis) and/or patient survival. These findings may identify important molecular markers and provide better patient diagnosis and prognosis. Another important experiment could involve the manipulation of gene expression of some affected cellular genes as well as HPV genes in order to understand their importance in the carcinogenesis of the Head and Neck tissues. These experiments may involve knocking-down the expression of the upregulated genes (such as SYCP2) or the viral genes such as E6 and E7 in HPV-positive SCCHN by using specific siRNAs in SCCHN cell lines containing or lacking HPV DNA. To study the role of cellular genes downregulated in HPV-positive SCCHN such as KLK8, one could transiently or stably overexpress such genes in HPV-positive SCCHN cell lines. Identification of phenotypical changes (cell death,

growth suppression, etc.) in the above cell lines may provide clues about the role of the particular cellular gene in the development of HPV-associated SCCHN.

In the project dealing with the global expression profiles in HPV-positive cervical cell lines, the next series of experiments should involve the verification of differentially gene expression in cervical tumor tissues by microarray, qRT-PCR or tissue-array analyses. These studies will help identify genes whose expression may be affected due to cellular growth in vitro. Another experiment could involve the manipulation of expression of the upregulated LAT-1 and the downregulated NAP1L3 genes in HPV-positive cervical cell lines in order to understand their importance in cervical carcinogenesis (by using LAT1 siRNA or an expression vector containing NAP1L3). Another important study would be to identify cellular pathways affected by the expression of HPV oncogenes. This can be done by expressing E6 and E7 siRNAs in HPV-positive cell lines and then hybridizing the cellular mRNAs to microarrays containing probes for specific cellular pathways such as cell cycle, DNA repair, or apoptosis. Similarly, the gene expression profiles of cervical keratinocytes transfected with vectors expressing the HPV E6 and/or E7 genes can be compared to those of non-transfectedcells using the above pathway arrays.  Such studies may provide interesting information on the mechanisms by which HPVs promote cervical carcinogenesis.

Finally, a number of exciting studies can be done to identify the specific roles of cellular and viral miRNAs in HPV pathogenesis. An important study would be to identify changes in cellular miRNA expression in HPV-positive

SCCHN tumor tissues and cell lines. This information will be important in order to identify miRNAs that are expressed in a tissue-specific manner (cervix vs oropharyngeal tissue) as well as possible miRNAs that are affected by the presence of the HPV oncogenes in both types of epitheliums. Another experiment could involve the identification of additional potential mRNA targets of some of the cervical cancer related miRNAs by comparing their expression to those of cellular mRNAs. An inverse correlation between the expression levels of a particular miRNA and its predicted mRNA target would provide a basis for a more in-depth future analysis of miRNA-dependent gene regulation in HPV-positive cancers. Similarly, the protein expression levels of the predicted miRNA targets can be measured and correlated with the particular miRNA levels in HPV-positive cells. The levels of specific miRNA can also be experimentally manipulated in cell lines by using a miRNA expressing plasmid or a miRNA-siRNA, and the effect of such changes on predicted mRNA/protein target determined. Another interesting project would involve studies on Dicer expression in some of the HPV-positive cervical and oropharyngeal cell lines and tissues. Dicer is a ribonuclease involved in the maturation of miRNAs. Most of the differentially expressed miRNAs that we found in the cervical cell lines and tumor tissues were underexpressed. One possible explanation of these findings is the underexpression of the enzymes involved in the maturation of miRNAs such as Drosha and Dicer. Interestingly, our gene expression profiling studies showed that Dicer 1 was downregulated in the HPV-positive cervcal cell lines compared to the control cervical tissue (Table 8). By qRT-PCR, western blotting

171

and sequencing we could further study Dicer 1 expression and identify the basis for its downregulation, if any, for example due to a mutation. Finally, another interesting study would be the identification of possible miRNAs encoded by HPVs. Several viruses have recently been shown to encode miRNAs. There is no evidence yet that HPVs encode miRNAs, but this may be due to the lack of sensitive technologies. Also, unlike other viruses HPVs can not be grown to high titers in tissue culture which may hamper efforts to identify HPV miRNAs.  Amy Gardiner, a graduate student in the laboratory of Dr. Saleem Khan has developed a microarray containing the complete genome of HPV-16 in order to detect possible miRNAs expressed by HPVs.

**Chapter 6**


**BIBLIOGRAPHY**

**Aguilar-Lemarroy A, Kirchhoff S, Whitaker N, Gariglio P, zur Hausen H, Krammer PH & Rosl F**. 2001. Differential sensitivity of human papillomavirus type 16(+) and type 18(+) cervical carcinoma cells to CD95-mediated apoptosis. *Int J Cancer*. **93**:823-831.

**Ahn WS, Bae SM, Lee JM, Namkoong SE, Han SJ, Cho YL, Nam GH, Seo JS, Kim CK & Kim YW**. 2004. Searching for pathogenic gene functions to cervical cancer. *Gynecol Oncol*. **93**:41-48.

**Alazawi W, Pett M, Arch B, Scott L, Freeman T, Stanley MA & Coleman N**. 2002. Changes in cervical keratinocyte gene expression associated with integration of human papillomavirus 16. *Cancer Res*. **62**:6959-6965.

**Andersen B, Hariri A, Pittelkow MR & Rosenfeld MG**. 1997. Characterization of Skn-1a/i POU domain factors and linkage to papillomavirus gene expression. *J Biol Chem*. **272**:15905-15913.

**Antinore MJ, Birrer MJ, Patel D, Nader L & McCance DJ**. 1996. The human papillomavirus type 16 E7 gene product interacts with and trans-activates the AP1 family of transcription factors. *Embo J*. **15**:1950-1960.

**Arany I, Chen SH, Megyesi JK, Adler-Storthz K, Chen Z, Rajaraman S, Ember IA, Tyring SK & Brysk MM**. 2003. Differentiation-dependent expression of signal transducers and activators of transcription (STATs) might modify responses to growth factors in the cancers of the head and neck. *Cancer Lett*. **199**:83-89.

**Badaracco G, Venuti A, Sedati A & Marcante ML**. 2002. HPV16 and HPV18 in genital tumors: Significantly different levels of viral integration and correlation to tumor invasiveness. *J Med Virol*. **67**:574-582.

**Baker CC, Phelps WC, Lindgren V, Braun MJ, Gonda MA & Howley PM**. 1987. Structural and transcriptional analysis of human papillomavirus type 16 sequences in cervical carcinoma cell lines. *J Virol*. **61**:962-971.

**Baker TS, Newcomb WW, Olson NH, Cowsert LM, Olson C & Brown JC**. 1991. Structures of bovine and human papillomaviruses. Analysis by cryoelectron microscopy and three-dimensional image reconstruction. *Biophys J*. **60**:1445-1456.

**Band V, Dalal S, Delmolino L & Androphy EJ**. 1993. Enhanced degradation of p53 protein in HPV-6 and BPV-1 E6-immortalized human mammary epithelial cells. *Embo J*. **12**:1847-1852.

**Bandres E, Cubedo E, Agirre X, Malumbres R, Zarate R, Ramirez N, Abajo A, Navarro A, Moreno I, Monzo M & Garcia-Foncillas J**. 2006. Identification by Real-time PCR of 13 mature microRNAs differentially expressed in colorectal cancer and non-tumoral tissues. *Mol Cancer*. **5**:29.

**Bechtold V, Beard P & Raj K**. 2003. Human papillomavirus type 16 E2 protein has no effect on transcription from episomal viral DNA. *J Virol*. **77**:2021-2028.

**Beger M, Butz K, Denk C, Williams T, Hurst HC & Hoppe-Seyler F**. 2001. Expression pattern of AP-2 transcription factors in cervical cancer cells and analysis of their influence on human papillomavirus oncogene transcription. *J Mol Med*. **79**:314-320.

**Bollmann M, Bankfalvi A, Trosic A, Speich N, Schmittt C & Bollmann R**. 2005. Can we detect cervical human papillomavirus (HPV) infection by cytomorphology alone? Diagnostic value of non-classic cytological signs of HPV effect in minimally abnormal Pap tests. *Cytopathology*. **16**:13-21.

**Brazma A, Hingamp P, Quackenbush J, Sherlock G, Spellman P, Stoeckert C, Aach J, Ansorge W, Ball CA, Causton HC, Gaasterland T, Glenisson P, Holstege FC, Kim IF, Markowitz V, Matese JC, Parkinson H, Robinson A, Sarkans U, Schulze-Kremer S, Stewart J, Taylor R, Vilo J & Vingron M**. 2001. Minimum information about a microarray experiment (MIAME)-toward standards for microarray data. *Nat Genet*. **29**:365-371.

**Brehm A, Nielsen SJ, Miska EA, McCance DJ, Reid JL, Bannister AJ & Kouzarides T**. 1999. The E7 oncoprotein associates with Mi2 and histone deacetylase activity to promote cell growth. *Embo J*. **18**:2449-2458.

**Brown PO & Botstein D**. 1999. Exploring the new world of the genome with DNA microarrays. *Nat Genet*. **21**:33-37.

**Burd EM**. 2003. Human papillomavirus and cervical cancer. *Clin Microbiol Rev*. **16**:1-17.

**Burk RD, Terai M, Gravitt PE, Brinton LA, Kurman RJ, Barnes WA, Greenberg MD, Hadjimichael OC, Fu L, McGowan L, Mortel R, Schwartz PE & Hildesheim A**. 2003. Distribution of human papillomavirus types 16 and 18 variants in squamous cell carcinomas and adenocarcinomas of the cervix. *Cancer Res*. **63**:7215-7220.

**Burkhardt A, Willingham M, Gay C, Jeang KT & Schlegel R**. 1989. The E5 oncoprotein of bovine papillomavirus is oriented asymmetrically in Golgi and plasma membranes. *Virology*. **170**:334-339.

**Butte A**. 2002. The use and analysis of microarray data. *Nat Rev Drug Discov*. **1**:951-960.

**Cai X, Hagedorn CH & Cullen BR**. 2004. Human microRNAs are processed from capped, polyadenylated transcripts that can also function as mRNAs. *Rna*. **10**:1957-1966.

**Calaluce R, Bearss DJ, Barrera J, Zhao Y, Han H, Beck SK, McDaniel K & Nagle RB**. 2004. Laminin-5 beta3A expression in LNCaP human prostate carcinoma cells increases cell migration and tumorigenicity. *Neoplasia*. **6**:468-479.

**Calin GA & Croce CM**. 2006. MicroRNA-cancer connection: the beginning of a new tale. *Cancer Res*. **66**:7390-7394.

**Calin GA, Dumitru CD, Shimizu M, Bichi R, Zupo S, Noch E, Aldler H, Rattan S, Keating M, Rai K, Rassenti L, Kipps T, Negrini M, Bullrich F & Croce CM**. 2002. Frequent deletions and down-regulation of micro- RNA genes miR15 and miR16 at 13q14 in chronic lymphocytic leukemia. *Proc Natl Acad Sci U S A*. **99**:15524-15529.

**Calin GA, Sevignani C, Dumitru CD, Hyslop T, Noch E, Yendamuri S, Shimizu M, Rattan S, Bullrich F, Negrini M & Croce CM**. 2004. Human microRNA genes are frequently located at fragile sites and genomic regions involved in cancers. *Proc Natl Acad Sci U S A*. **101**:2999-3004.

**Calo V, Migliavacca M, Bazan V, Macaluso M, Buscemi M, Gebbia N & Russo A**. 2003. STAT proteins: from normal control of cellular events to tumorigenesis. *J Cell Physiol*. **197**:157-168.

**Cane S, Bignotti E, Bellone S, Palmieri M, De las Casas L, Roman JJ, Pecorelli S, Cannon MJ, O'Brien T & Santin AD**. 2004. The novel serine protease tumor-associated differentially expressed gene-14 (KLK8/Neuropsin/Ovasin) is highly overexpressed in cervical cancer. *Am J Obstet Gynecol*. **190**:60-66.

**Cates W, Jr.** 1999. Estimates of the incidence and prevalence of sexually transmitted diseases in the United States. American Social Health Association Panel. *Sex Transm Dis*. **26**:S2-7.

**Chang YE & Laimins LA**. 2000. Microarray analysis identifies interferon-inducible genes and Stat-1 as major transcriptional targets of human papillomavirus type 31. *J Virol*. **74**:4174-4182.

**Chen Y, Miller C, Mosher R, Zhao X, Deeds J, Morrissey M, Bryant B, Yang D, Meyer R, Cronin F, Gostout BS, Smith-McCune K & Schlegel R**. 2003. Identification of cervical cancer markers by cDNA and tissue microarrays. *Cancer Res*. **63**:1927-1935.

**Chow LT & Broker TR**. 1994. Papillomavirus DNA replication. *Intervirology*. **37**:150-158.

**Cimmino A, Calin GA, Fabbri M, Iorio MV, Ferracin M, Shimizu M, Wojcik SE, Aqeilan RI, Zupo S, Dono M, Rassenti L, Alder H, Volinia S, Liu CG, Kipps TJ, Negrini M & Croce CM**. 2005. miR-15 and miR-16 induce apoptosis by targeting BCL2. *Proc Natl Acad Sci U S A*. **102**:13944-13949.

**Conrad M, Bubb VJ & Schlegel R**. 1993. The human papillomavirus type 6 and 16 E5 proteins are membrane-associated proteins which associate with the 16-kilodalton pore-forming protein. *J Virol*. **67**:6170-6178.

**Conway T & Schoolnik GK**. 2003. Microarray expression profiling: capturing a genome-wide portrait of the transcriptome. *Mol Microbiol*. **47**:879-889.

**Costinean S, Zanesi N, Pekarsky Y, Tili E, Volinia S, Heerema N & Croce CM**. 2006. Pre-B cell proliferation and lymphoblastic leukemia/high-grade lymphoma in E(mu)-miR155 transgenic mice. *Proc Natl Acad Sci U S A*. **103**:7024-7029.

**Culp TD, Budgeon LR, Marinkovich MP, Meneguzzi G & Christensen ND**. 2006. Keratinocyte-secreted laminin 5 can function as a transient receptor for human papillomaviruses by binding virions and transferring them to adjacent cells. *J Virol*. **80**:8940-8950.

**Cummins JM, He Y, Leary RJ, Pagliarini R, Diaz LA, Jr., Sjoblom T, Barad O, Bentwich Z, Szafranska AE, Labourier E, Raymond CK, Roberts BS, Juhl H, Kinzler KW, Vogelstein B & Velculescu VE**. 2006. The colorectal microRNAome. *Proc Natl Acad Sci U S A*. **103**:3687-3692.

**Dahlstrand H, Dahlgren L, Lindquist D, Munck-Wikland E & Dalianis T**. 2004. Presence of human papillomavirus in tonsillar cancer is a favourable prognostic factor for clinical outcome. *Anticancer Res*. **24**:1829-1835.

**Dajee M, Lazarov M, Zhang JY, Cai T, Green CL, Russell AJ, Marinkovich MP, Tao S, Lin Q, Kubo Y & Khavari PA**. 2003. NF-kappaB blockade and oncogenic Ras trigger invasive human epidermal neoplasia. *Nature*. **421**:639-643.

**Dallol A, Da Silva NF, Viacava P, Minna JD, Bieche I, Maher ER & Latif F**. 2002. SLIT2, a human homologue of the Drosophila Slit2 gene, has tumor suppressor activity and is frequently inactivated in lung and breast cancers. *Cancer Res*. **62**:5874-5880.

**Datta S & Datta S**. 2003. Comparisons and validation of statistical clustering techniques for microarray gene expression data. *Bioinformatics*. **19**:459-466.

**Davies R, Hicks R, Crook T, Morris J & Vousden K**. 1993. Human papillomavirus type 16 E7 associates with a histone H1 kinase and with p107 through sequences necessary for transformation. *J Virol*. **67**:2521-2528.

**de la Guardia C, Casiano CA, Trinidad-Pinedo J & Baez A**. 2001. CENP-F gene amplification and overexpression in head and neck squamous cell carcinomas. *Head Neck*. **23**:104-112.

**Dell G & Gaston K**. 2001. Human papillomaviruses and their role in cervical cancer. *Cell Mol Life Sci*. **58**:1923-1942.

**Dobrossy L**. 2005. Epidemiology of head and neck cancer: magnitude of the problem. *Cancer Metastasis Rev*. **24**:9-17.

**Dopazo J, Zanders E, Dragoni I, Amphlett G & Falciani F**. 2001. Methods and approaches in the analysis of gene expression data. *J Immunol Methods*. **250**:93-112.

**Dresios J, Aschrafi A, Owens GC, Vanderklish PW, Edelman GM & Mauro VP**. 2005. Cold stress-induced protein Rbm3 binds 60S ribosomal subunits, alters microRNA levels, and enhances global protein synthesis. *Proc Natl Acad Sci U S A*. **102**:1865-1870.

**Duensing S, Duensing A, Flores ER, Do A, Lambert PF & Munger K**. 2001. Centrosome abnormalities and genomic instability by episomal expression of human papillomavirus type 16 in raft cultures of human keratinocytes. *J Virol*. **75**:7712-7716.

**Duensing S, Lee LY, Duensing A, Basile J, Piboonniyom S, Gonzalez S, Crum CP & Munger K**. 2000. The human papillomavirus type 16 E6 and E7 oncoproteins cooperate to induce mitotic defects and genomic instability by uncoupling centrosome duplication from the cell division cycle. *Proc Natl Acad Sci U S A*. **97**:10002-10007.

**Duensing S & Munger K**. 2002. Human papillomaviruses and centrosome duplication errors: modeling the origins of genomic instability. *Oncogene*. **21**:6241-6248.

**Duffy CL, Phillips SL & Klingelhutz AJ**. 2003. Microarray analysis identifies differentiation-associated genes regulated by human papillomavirus type 16 E6. *Virology*. **314**:196-205.

**Durst M, Kleinheinz A, Hotz M & Gissmann L**. 1985. The physical state of human papillomavirus type 16 DNA in benign and malignant genital tumours. *J Gen Virol*. **66 ( Pt 7)**:1515-1522.

**Dyson N, Howley PM, Munger K & Harlow E**. 1989. The human papilloma virus-16 E7 oncoprotein is able to bind to the retinoblastoma gene product. *Science*. **243**:934-937.

**Eisen MB, Spellman PT, Brown PO & Botstein D**. 1998. Cluster analysis and display of genome-wide expression patterns. *Proc Natl Acad Sci U S A*. **95**:14863-14868.

**Esquela-Kerscher A & Slack FJ**. 2006. Oncomirs - microRNAs with a role in cancer. *Nat Rev Cancer*. **6**:259-269.

**Evander M, Frazer IH, Payne E, Qi YM, Hengst K & McMillan NA**. 1997. Identification of the alpha6 integrin as a candidate receptor for papillomaviruses. *J Virol*. **71**:2449-2456.

**Favre M**. 1975. Structural polypeptides of rabbit, bovine, and human papillomaviruses. *J Virol*. **15**:1239-1247.

**Ferris RL, Martinez I, Sirianni N, Wang J, Lopez-Albaitero A, Gollin SM, Johnson JT & Khan S**. 2005. Human papillomavirus-16 associated squamous cell carcinoma of the head and neck (SCCHN): a natural disease model provides insights into viral carcinogenesis. *Eur J Cancer*. **41**:807-815.

**Frattini MG, Lim HB, Doorbar J & Laimins LA**. 1997. Induction of human papillomavirus type 18 late gene expression and genomic amplification in organotypic cultures from transfected DNA templates. *J Virol*. **71**:7068-7072.

**Fuchs PG & Pfister H**. 1994. Transcription of papillomavirus genomes. *Intervirology*. **37**:159-167.

**Fuks F, Burgers WA, Godin N, Kasai M & Kouzarides T**. 2001. Dnmt3a binds deacetylases and is recruited by a sequence-specific repressor to silence transcription. *Embo J*. **20**:2536-2544.

**Furniss CS, McClean MD, Smith JF, Bryan J, Nelson HH, Peters ES, Posner MR, Clark JR, Eisen EA & Kelsey KT**. 2007. Human papillomavirus 16 and head and neck squamous cell carcinoma. *Int J Cancer*.

**Gardiol D, Kuhne C, Glaunsinger B, Lee SS, Javier R & Banks L**. 1999. Oncogenic human papillomavirus E6 proteins target the discs large tumour suppressor for proteasome-mediated degradation. *Oncogene*. **18**:5487-5496.

**Gillison ML, Koch WM, Capone RB, Spafford M, Westra WH, Wu L, Zahurak ML, Daniel RW, Viglione M, Symer DE, Shah KV & Sidransky D**. 2000. Evidence for a causal association between human papillomavirus and a subset of head and neck cancers. *J Natl Cancer Inst*. **92**:709-720.

**Gillison ML & Shah KV**. 2001. Human papillomavirus-associated head and neck squamous cell carcinoma: mounting evidence for an etiologic role for human papillomavirus in a subset of head and neck cancers. *Curr Opin Oncol*. **13**:183-188.

**Giroglou T, Florin L, Schafer F, Streeck RE & Sapp M**. 2001. Human papillomavirus infection requires cell surface heparan sulfate. *J Virol*. **75**:1565-1570.

**Golub TR, Slonim DK, Tamayo P, Huard C, Gaasenbeek M, Mesirov JP, Coller H, Loh ML, Downing JR, Caligiuri MA, Bloomfield CD & Lander ES**. 1999. Molecular classification of cancer: class discovery and class prediction by gene expression monitoring. *Science*. **286**:531-537.

**Greenlee RT, Hill-Harmon MB, Murray T & Thun M**. 2001. Cancer statistics, 2001. *CA Cancer J Clin*. **51**:15-36.

**Griffiths-Jones S**. 2004. The microRNA Registry. *Nucleic Acids Res*. **32**:D109-111.

**Griffiths-Jones S, Grocock RJ, van Dongen S, Bateman A & Enright AJ**. 2006. miRBase: microRNA sequences, targets and gene nomenclature. *Nucleic Acids Res*. **34**:D140-144.

**Guo X, Lui WO, Qian CN, Chen JD, Gray SG, Rhodes D, Haab B, Stanbridge E, Wang H, Hong MH, Min HQ, Larsson C & Teh BT**. 2002. Identifying cancer-related genes in nasopharyngeal carcinoma cell lines using DNA and mRNA expression profiling analyses. *Int J Oncol*. **21**:1197-1204.

**Hebner CM & Laimins LA**. 2006. Human papillomaviruses: basic mechanisms of pathogenesis and oncogenicity. *Rev Med Virol*. **16**:83-97.

**Herrero R, Castellsague X, Pawlita M, Lissowska J, Kee F, Balaram P, Rajkumar T, Sridhar H, Rose B, Pintos J, Fernandez L, Idris A, Sanchez MJ,**

**Nieto A, Talamini R, Tavani A, Bosch FX, Reidel U, Snijders PJ, Meijer CJ, Viscidi R, Munoz N & Franceschi S**. 2003. Human papillomavirus and oral cancer: the International Agency for Research on Cancer multicenter study. *J Natl Cancer Inst*. **95**:1772-1783.

**Hirt B**. 1967. Selective extraction of polyoma DNA from infected mouse cell cultures. *J Mol Biol*. **26**:365-369.

**Honda K, Yamada T, Hayashida Y, Idogawa M, Sato S, Hasegawa F, Ino Y, Ono M & Hirohashi S**. 2005. Actinin-4 increases cell motility and promotes lymph node metastasis of colorectal cancer. *Gastroenterology*. **128**:51-62.

**Horvath LG, Henshall SM, Kench JG, Saunders DN, Lee CS, Golovsky D, Brenner PC, O'Neill GF, Kooner R, Stricker PD, Grygiel JJ & Sutherland RL**. 2004. Membranous expression of secreted frizzled-related protein 4 predicts for good prognosis in localized prostate cancer and inhibits PC3 cellular proliferation in vitro. *Clin Cancer Res*. **10**:615-625.

**Huang M, Stolina M, Sharma S, Mao JT, Zhu L, Miller PW, Wollman J, Herschman H & Dubinett SM**. 1998. Non-small cell lung cancer cyclooxygenase-2-dependent regulation of cytokine balance in lymphocytes and macrophages: up-regulation of interleukin 10 and down-regulation of interleukin 12 production. *Cancer Res*. **58**:1208-1216.

**Huibregtse JM, Scheffner M & Howley PM**. 1991. A cellular protein mediates association of p53 with the E6 oncoprotein of human papillomavirus types 16 or 18. *Embo J*. **10**:4129-4135.

**Hutvagner G, McLachlan J, Pasquinelli AE, Balint E, Tuschl T & Zamore PD**. 2001. A cellular function for the RNA-interference enzyme Dicer in the maturation of the let-7 small temporal RNA. *Science*. **293**:834-838.

**Ibrahim SO, Aarsaether N, Holsve MK, Kross KW, Heimdal JH, Aarstad JH, Liavaag PG, Elgindi OA, Johannessen AC, Lillehaug JR & Vasstrand EN**. 2003. Gene expression profile in oral squamous cell carcinomas and matching normal oral mucosal tissues from black Africans and white Caucasians: the case of the Sudan vs. Norway. *Oral Oncol*. **39**:37-48.

**Imai MA, Moriya T, Imai FL, Shiiba M, Bukawa H, Yokoe H, Uzawa K & Tanzawa H**. 2005. Down-regulation of DMBT1 gene expression in human oral squamous cell carcinoma. *Int J Mol Med*. **15**:585-589.

**Iorio MV, Ferracin M, Liu CG, Veronese A, Spizzo R, Sabbioni S, Magri E, Pedriali M, Fabbri M, Campiglio M, Menard S, Palazzo JP, Rosenberg A, Musiani P, Volinia S, Nenci I, Calin GA, Querzoli P, Negrini M & Croce CM**. 2005. MicroRNA gene expression deregulation in human breast cancer. *Cancer Res*. **65**:7065-7070.

**Iwanaga R, Komori H, Ishida S, Okamura N, Nakayama K, Nakayama KI & Ohtani K**. 2006. Identification of novel E2F1 target genes regulated in cell cycle-dependent and independent manners. *Oncogene*. **25**:1786-1798.

**Jeon GA, Lee JS, Patel V, Gutkind JS, Thorgeirsson SS, Kim EC, Chu IS, Amornphimoltham P & Park MH**. 2004. Global gene expression profiles of human head and neck squamous carcinoma cell lines. *Int J Cancer*. **112**:249-258.

**Jeon S, Allen-Hoffmann BL & Lambert PF**. 1995. Integration of human papillomavirus type 16 into the human genome correlates with a selective growth advantage of cells. *J Virol*. **69**:2989-2997.

**Jiang M & Milner J**. 2002. Selective silencing of viral gene expression in HPV-positive human cervical carcinoma cells treated with siRNA, a primer of RNA interference. *Oncogene*. **21**:6041-6048.

**Jin XW, Cash J & Kennedy AW**. 1999. Human papillomavirus typing and the reduction of cervical cancer risk. *Cleve Clin J Med*. **66**:533-539.

**Johnson SM, Grosshans H, Shingara J, Byrom M, Jarvis R, Cheng A, Labourier E, Reinert KL, Brown D & Slack FJ**. 2005. RAS is regulated by the let-7 microRNA family. *Cell*. **120**:635-647.

**Kannan K, Amariglio N, Rechavi G, Jakob-Hirsch J, Kela I, Kaminski N, Getz G, Domany E & Givol D**. 2001. DNA microarrays identification of primary and secondary target genes regulated by p53. *Oncogene*. **20**:2225-2234.

**Kayashima T, Yamasaki K, Yamada T, Sakai H, Miwa N, Ohta T, Yoshiura K, Matsumoto N, Nakane Y, Kanetake H, Ishino F, Niikawa N & Kishino T**. 2003. The novel imprinted carboxypeptidase A4 gene ( CPA4) in the 7q32 imprinting domain. *Hum Genet*. **112**:220-226.

**Kell B, Jewers RJ, Cason J, Pakarian F, Kaye JN & Best JM**. 1994. Detection of E5 oncoprotein in human papillomavirus type 16-positive cervical scrapes using antibodies raised to synthetic peptides. *J Gen Virol*. **75 ( Pt 9)**:2451-2456.

**Kessis TD, Connolly DC, Hedrick L & Cho KR**. 1996. Expression of HPV16 E6 or E7 increases integration of foreign DNA. *Oncogene*. **13**:427-431.

**Khvorova A, Reynolds A & Jayasena SD**. 2003. Functional siRNAs and miRNAs exhibit strand bias. *Cell*. **115**:209-216.

**Kim J & MacNeill SA**. 2003. Genome stability: a new member of the RFC family. *Curr Biol*. **13**:R873-875.

**Kim VN**. 2005. MicroRNA biogenesis: coordinated cropping and dicing. *Nat Rev Mol Cell Biol*. **6**:376-385.

**Klingelhutz AJ, Foster SA & McDougall JK**. 1996. Telomerase activation by the E6 gene product of human papillomavirus type 16. *Nature*. **380**:79-82.

**Klussmann JP, Gultekin E, Weissenborn SJ, Wieland U, Dries V, Dienes HP, Eckel HE, Pfister HJ & Fuchs PG**. 2003. Expression of p16 protein identifies a distinct entity of tonsillar carcinomas associated with human papillomavirus. *Am J Pathol*. **162**:747-753.

**Kohlberger P, Beneder C, Horvat R, Leodolter S & Breitenecker G**. 2003. Immunohistochemical expression of laminin-5 in cervical intraepithelial neoplasia. *Gynecol Oncol*. **89**:391-394.

**Kornberg LJ, Villaret D, Popp M, Lui L, McLaren R, Brown H, Cohen D, Yun J & McFadden M**. 2005. Gene expression profiling in squamous cell carcinoma of the oral cavity shows abnormalities in several signaling pathways. *Laryngoscope*. **115**:690-698.

**Kumar RV, Kadkol SS, Daniel R, Shenoy AM & Shah KV**. 2003. Human papillomavirus, p53 and cyclin D1 expression in oropharyngeal carcinoma. *Int J Oral Maxillofac Surg*. **32**:539-543.

**Kuo WP, Jenssen TK, Park PJ, Lingen MW, Hasina R & Ohno-Machado L**. 2002. Gene expression levels in different stages of progression in oral squamous cell carcinoma. *Proc AMIA Symp*.415-419.

**Lechner MS, Mack DH, Finicle AB, Crook T, Vousden KH & Laimins LA**. 1992. Human papillomavirus E6 proteins bind p53 in vivo and abrogate p53-mediated repression of transcription. *Embo J*. **11**:3045-3052.

**Lee C & Laimins LA**. 2004. Role of the PDZ domain-binding motif of the oncoprotein E6 in the pathogenesis of human papillomavirus type 31. *J Virol*. **78**:12366-12377.

**Lee KA, Shim JH, Kho CW, Park SG, Park BC, Kim JW, Lim JS, Choe YK, Paik SG & Yoon DY**. 2004. Protein profiling and identification of modulators regulated by the E7 oncogene in the C33A cell line by proteomics and genomics. *Proteomics*. **4**:839-848.

**Lee RC, Feinbaum RL & Ambros V**. 1993. The C. elegans heterochronic gene lin-4 encodes small RNAs with antisense complementarity to lin-14. *Cell*. **75**:843-854.

**Lee Y, Ahn C, Han J, Choi H, Kim J, Yim J, Lee J, Provost P, Radmark O, Kim S & Kim VN**. 2003. The nuclear RNase III Drosha initiates microRNA processing. *Nature*. **425**:415-419.

**Lee Y, Jeon K, Lee JT, Kim S & Kim VN**. 2002. MicroRNA maturation: stepwise processing and subcellular localization. *Embo J*. **21**:4663-4670.

**Lee Y, Kim M, Han J, Yeom KH, Lee S, Baek SH & Kim VN**. 2004. MicroRNA genes are transcribed by RNA polymerase II. *Embo J*. **23**:4051-4060.

**Leethanakul C, Knezevic V, Patel V, Amornphimoltham P, Gillespie J, Shillitoe EJ, Emko P, Park MH, Emmert-Buck MR, Strausberg RL, Krizman DB & Gutkind JS**. 2003. Gene discovery in oral squamous cell carcinoma

through the Head and Neck Cancer Genome Anatomy Project: confirmation by microarray analysis. *Oral Oncol*. **39**:248-258.

**Lewis BP, Burge CB & Bartel DP**. 2005. Conserved seed pairing, often flanked by adenosines, indicates that thousands of human genes are microRNA targets. *Cell*. **120**:15-20.

**Li C & Wong WH**. 2001. Model-based analysis of oligonucleotide arrays: expression index computation and outlier detection. *Proc Natl Acad Sci U S A*. **98**:31-36.

**Li J, Zhang Z, Bidder M, Funk MC, Nguyen L, Goodfellow PJ & Rader JS**. 2005. IGSF4 promoter methylation and expression silencing in human cervical cancer. *Gynecol Oncol*. **96**:150-158.

**Li W, Thompson CH, Cossart YE, O'Brien CJ, McNeil EB, Scolyer RA & Rose BR**. 2004. The expression of key cell cycle markers and presence of human papillomavirus in squamous cell carcinoma of the tonsil. *Head Neck*. **26**:1-9.

**Li X, Gu W, Mohan S & Baylink DJ**. 2002. DNA microarrays: their use and misuse. *Microcirculation*. **9**:13-22.

**Lim KC, Lakshmanan G, Crawford SE, Gu Y, Grosveld F & Engel JD**. 2000. Gata3 loss leads to embryonic lethality due to noradrenaline deficiency of the sympathetic nervous system. *Nat Genet*. **25**:209-212.

**Lin SC, Wang CP, Chen YM, Lu SY, Fann MJ, Liu CJ, Kao SY & Chang KW**. 2002. Regulation of IGFBP-5 expression during tumourigenesis and differentiation of oral keratinocytes. *J Pathol*. **198**:317-325.

**Liu CG, Calin GA, Meloon B, Gamliel N, Sevignani C, Ferracin M, Dumitru CD, Shimizu M, Zupo S, Dono M, Alder H, Bullrich F, Negrini M & Croce CM**. 2004. An oligonucleotide microchip for genome-wide microRNA profiling in human and mouse tissues. *Proc Natl Acad Sci U S A*. **101**:9740-9744.

**Livak KJ & Schmittgen TD**. 2001. Analysis of relative gene expression data using real-time quantitative PCR and the 2(-Delta Delta C(T)) Method. *Methods*. **25**:402-408.

**Loning T, Ikenberg H, Becker J, Gissmann L, Hoepfer I & zur Hausen H**. 1985. Analysis of oral papillomas, leukoplakias, and invasive carcinomas for human papillomavirus type related DNA. *J Invest Dermatol*. **84**:417-420.

**Lu J, Getz G, Miska EA, Alvarez-Saavedra E, Lamb J, Peck D, Sweet-Cordero A, Ebert BL, Mak RH, Ferrando AA, Downing JR, Jacks T, Horvitz HR & Golub TR**. 2005. MicroRNA expression profiles classify human cancers. *Nature*. **435**:834-838.

**Lund E, Guttinger S, Calado A, Dahlberg JE & Kutay U**. 2004. Nuclear export of microRNA precursors. *Science*. **303**:95-98.

**Mansur CP & Androphy EJ**. 1993. Cellular transformation by papillomavirus oncoproteins. *Biochim Biophys Acta*. **1155**:323-345.

**Mantovani F & Banks L**. 1999. The interaction between p53 and papillomaviruses. *Semin Cancer Biol*. **9**:387-395.

**Marsit CJ, Okpukpara C, Danaee H & Kelsey KT**. 2005. Epigenetic silencing of the PRSS3 putative tumor suppressor gene in non-small cell lung cancer. *Mol Carcinog*. **44**:146-150.

**Martinez I, Wang J, Hobson KF, Ferris RL & Khan SA**. 2007. Identification of differentially expressed genes in HPV-positive and HPV-negative oropharyngeal squamous cell carcinomas. *Eur J Cancer*. **43**:415-432.

**Massimi P, Pim D, Storey A & Banks L**. 1996. HPV-16 E7 and adenovirus E1a complex formation with TATA box binding protein is enhanced by casein kinase II phosphorylation. *Oncogene*. **12**:2325-2330.

**May M, Dong XP, Beyer-Finkler E, Stubenrauch F, Fuchs PG & Pfister H**. 1994. The E6/E7 promoter of extrachromosomal HPV16 DNA in cervical cancers escapes from cellular repression by mutation of target sequences for YY1. *Embo J*. **13**:1460-1466.

**Mazzarelli JM, Atkins GB, Geisberg JV & Ricciardi RP**. 1995. The viral oncoproteins Ad5 E1A, HPV16 E7 and SV40 TAg bind a common region of the TBP-associated factor-110. *Oncogene*. **11**:1859-1864.

**McBride AA, Romanczuk H & Howley PM**. 1991. The papillomavirus E2 regulatory proteins. *J Biol Chem*. **266**:18411-18414.

**McCance DJ**. 2005. Transcriptional regulation by human papillomaviruses. *Curr Opin Genet Dev*. **15**:515-519.

**McIntyre MC, Ruesch MN & Laimins LA**. 1996. Human papillomavirus E7 oncoproteins bind a single form of cyclin E in a complex with cdk2 and p107. *Virology*. **215**:73-82.

**McMurray HR, Nguyen D, Westbrook TF & McAnce DJ**. 2001. Biology of human papillomaviruses. *Int J Exp Pathol*. **82**:15-33.

**Meissner JD**. 1999. Nucleotide sequences and further characterization of human papillomavirus DNA present in the CaSki, SiHa and HeLa cervical carcinoma cell lines. *J Gen Virol*. **80 ( Pt 7)**:1725-1733.

**Michael MZ, SM OC, van Holst Pellekaan NG, Young GP & James RJ**. 2003. Reduced accumulation of specific microRNAs in colorectal neoplasia. *Mol Cancer Res*. **1**:882-891.

**Mollenhauer J, Helmke B, Muller H, Kollender G, Krebs I, Wiemann S, Holmskov U, Madsen J, Otto HF & Poustka A**. 2002. An integrative model on the role of DMBT1 in epithelial cancer. *Cancer Detect Prev*. **26**:266-274.

**Muller H, Bracken AP, Vernell R, Moroni MC, Christians F, Grassilli E, Prosperini E, Vigo E, Oliner JD & Helin K**. 2001. E2Fs regulate the expression of genes involved in differentiation, development, proliferation, and apoptosis. *Genes Dev*. **15**:267-285.

**Munger K & Howley PM**. 2002. Human papillomavirus immortalization and transformation functions. *Virus Res*. **89**:213-228.

**Munger K & Phelps WC**. 1993. The human papillomavirus E7 protein as a transforming and transactivating factor. *Biochim Biophys Acta*. **1155**:111-123.

**Munger K, Werness BA, Dyson N, Phelps WC, Harlow E & Howley PM**. 1989. Complex formation of human papillomavirus E7 proteins with the retinoblastoma tumor suppressor gene product. *Embo J*. **8**:4099-4105.

**Naiki T, Shimomura T, Kondo T, Matsumoto K & Sugimoto K**. 2000. Rfc5, in cooperation with rad24, controls DNA damage checkpoints throughout the cell cycle in Saccharomyces cerevisiae. *Mol Cell Biol*. **20**:5888-5896.

**Nakanishi T, Oka T & Akagi T**. 2001. Recent advances in DNA microarrays. *Acta Med Okayama*. **55**:319-328.

**O'Donnell KA, Wentzel EA, Zeller KI, Dang CV & Mendell JT**. 2005. c-Myc-regulated microRNAs modulate E2F1 expression. *Nature*. **435**:839-843.

**Okuducu AF, Janzen V, Ko Y, Hahne JC, Lu H, Ma ZL, Albers P, Sahin A, Wellmann A, Scheinert P & Wernert N**. 2005. Cellular retinoic acid-binding protein 2 is down-regulated in prostate cancer. *Int J Oncol*. **27**:1273-1282.

**Ota A, Tagawa H, Karnan S, Tsuzuki S, Karpas A, Kira S, Yoshida Y & Seto M**. 2004. Identification and characterization of a novel gene, C13orf25, as a target for 13q31-q32 amplification in malignant lymphoma. *Cancer Res*. **64**:3087-3095.

**Pallante P, Visone R, Ferracin M, Ferraro A, Berlingieri MT, Troncone G, Chiappetta G, Liu CG, Santoro M, Negrini M, Croce CM & Fusco A**. 2006. MicroRNA deregulation in human thyroid papillary carcinomas. *Endocr Relat Cancer*. **13**:497-508.

**Pasquinelli AE, Reinhart BJ, Slack F, Martindale MQ, Kuroda MI, Maller B, Hayward DC, Ball EE, Degnan B, Muller P, Spring J, Srinivasan A, Fishman M, Finnerty J, Corbo J, Levine M, Leahy P, Davidson E & Ruvkun G**. 2000. Conservation of the sequence and temporal expression of let-7 heterochronic regulatory RNA. *Nature*. **408**:86-89.

**Patel D, Huang SM, Baglia LA & McCance DJ**. 1999. The E6 protein of human papillomavirus type 16 binds to and inhibits co-activation by CBP and p300. *Embo J*. **18**:5061-5072.

**Patel S & Lyons-Weiler J**. 2004. caGEDA: a web application for the integrated analysis of global gene expression patterns in cancer. *Appl Bioinformatics*. **3**:49-62.

**Penning TM, Steckelbroeck S, Bauman DR, Miller MW, Jin Y, Peehl DM, Fung KM & Lin HK**. 2006. Aldo-keto reductase (AKR) 1C3: role in prostate disease and the development of specific inhibitors. *Mol Cell Endocrinol*. **248**:182-191.

**Pett MR, Alazawi WO, Roberts I, Dowen S, Smith DI, Stanley MA & Coleman N**. 2004. Acquisition of high-level chromosomal instability is associated with integration of human papillomavirus type 16 in cervical keratinocytes. *Cancer Res*. **64**:1359-1368.

**Pim D, Storey A, Thomas M, Massimi P & Banks L**. 1994. Mutational analysis of HPV-18 E6 identifies domains required for p53 degradation in vitro, abolition of p53 transactivation in vivo and immortalisation of primary BMK cells. *Oncogene*. **9**:1869-1876.

**Pisani P, Bray F & Parkin DM**. 2002. Estimates of the world-wide prevalence of cancer for 25 sites in the adult population. *Int J Cancer*. **97**:72-81.

**Portugal LG, Goldenberg JD, Wenig BL, Ferrer KT, Nodzenski E, Sabnani JB, Javier C, Weichselbaum RR & Vokes EE**. 1997. Human papillomavirus expression and p53 gene mutations in squamous cell carcinoma. *Arch Otolaryngol Head Neck Surg*. **123**:1230-1234.

**Ren B, Cam H, Takahashi Y, Volkert T, Terragni J, Young RA & Dynlacht BD**. 2002. E2F integrates cell cycle progression with DNA repair, replication, and G(2)/M checkpoints. *Genes Dev*. **16**:245-256.

**Risinger JI, Maxwell GL, Chandramouli GV, Aprelikova O, Litzi T, Umar A, Berchuck A & Barrett JC**. 2005. Gene expression profiling of microsatellite unstable and microsatellite stable endometrial cancers indicates distinct pathways of aberrant signaling. *Cancer Res*. **65**:5031-5037.

**Ritchie JM, Smith EM, Summersgill KF, Hoffman HT, Wang D, Klussmann JP, Turek LP & Haugen TH**. 2003. Human papillomavirus infection as a

prognostic factor in carcinomas of the oral cavity and oropharynx. *Int J Cancer*. **104**:336-344.

**Rodriguez A, Griffiths-Jones S, Ashurst JL & Bradley A**. 2004. Identification of mammalian microRNA host genes and transcription units. *Genome Res*. **14**:1902-1910.

**Saeed AI, Sharov V, White J, Li J, Liang W, Bhagabati N, Braisted J, Klapa M, Currier T, Thiagarajan M, Sturn A, Snuffin M, Rezantsev A, Popov D, Ryltsov A, Kostukovich E, Borisovsky I, Liu Z, Vinsavich A, Trush V & Quackenbush J**. 2003. TM4: a free, open-source system for microarray data management and analysis. *Biotechniques*. **34**:374-378.

**Sanclemente G & Gill DK**. 2002. Human papillomavirus molecular biology and pathogenesis. *J Eur Acad Dermatol Venereol*. **16**:231-240.

**Santin AD, Zhan F, Bignotti E, Siegel ER, Cane S, Bellone S, Palmieri M, Anfossi S, Thomas M, Burnett A, Kay HH, Roman JJ, O'Brien TJ, Tian E, Cannon MJ, Shaughnessy J, Jr. & Pecorelli S**. 2005. Gene expression profiles of primary HPV16- and HPV18-infected early stage cervical cancers and normal cervical epithelium: identification of novel candidate molecular markers for cervical cancer diagnosis and therapy. *Virology*. **331**:269-291.

**Saunders NA & Frazer IH**. 1998. Simplifying the molecular mechanisms of human papillomavirus. *Dermatol Clin*. **16**:823-827, xv.

**Saviozzi S, Iazzetti G, Caserta E, Guffanti A & Calogero RA**. 2004. Microarray data analysis and mining. *Methods Mol Med*. **94**:67-90.

**Scaltriti M, Brausi M, Amorosi A, Caporali A, D'Arca D, Astancolle S, Corti A & Bettuzzi S**. 2004. Clusterin (SGP-2, ApoJ) expression is downregulated in low- and high-grade human prostate cancer. *Int J Cancer*. **108**:23-30.

**Schaeffer AJ, Nguyen M, Liem A, Lee D, Montagna C, Lambert PF, Ried T & Difilippantonio MJ**. 2004. E6 and E7 oncoproteins induce distinct patterns of chromosomal aneuploidy in skin tumors from transgenic mice. *Cancer Res*. **64**:538-546.

**Scheffner M, Munger K, Byrne JC & Howley PM**. 1991. The state of the p53 and retinoblastoma genes in human cervical carcinoma cell lines. *Proc Natl Acad Sci U S A*. **88**:5523-5527.

**Scheffner M, Werness BA, Huibregtse JM, Levine AJ & Howley PM**. 1990. The E6 oncoprotein encoded by human papillomavirus types 16 and 18 promotes the degradation of p53. *Cell*. **63**:1129-1136.

**Schena M, Shalon D, Heller R, Chai A, Brown PO & Davis RW**. 1996. Parallel human genome analysis: microarray-based expression monitoring of 1000 genes. *Proc Natl Acad Sci U S A*. **93**:10614-10619.

**Schwartz SM, Daling JR, Doody DR, Wipf GC, Carter JJ, Madeleine MM, Mao EJ, Fitzgibbons ED, Huang S, Beckmann AM, McDougall JK & Galloway DA**. 1998. Oral cancer risk in relation to sexual history and evidence of human papillomavirus infection. *J Natl Cancer Inst*. **90**:1626-1636.

**Schwarz DS, Hutvagner G, Du T, Xu Z, Aronin N & Zamore PD**. 2003. Asymmetry in the assembly of the RNAi enzyme complex. *Cell*. **115**:199-208.

**Sedman SA, Barbosa MS, Vass WC, Hubbert NL, Haas JA, Lowy DR & Schiller JT**. 1991. The full-length E6 protein of human papillomavirus type 16 has transforming and trans-activating activities and cooperates with E7 to immortalize keratinocytes in culture. *J Virol*. **65**:4860-4866.

**Seedorf K, Krammer G, Durst M, Suhai S & Rowekamp WG**. 1985. Human papillomavirus type 16 DNA sequence. *Virology*. **145**:181-185.

**Shai A, Brake T, Somoza C & Lambert PF**. 2007. The human papillomavirus E6 oncogene dysregulates the cell cycle and contributes to cervical carcinogenesis through two independent activities. *Cancer Res*. **67**:1626-1635.

**Shakib K, Norman JT, Fine LG, Brown LR & Godovac-Zimmermann J**. 2005. Proteomics profiling of nuclear proteins for kidney fibroblasts suggests hypoxia, meiosis, and cancer may meet in the nucleus. *Proteomics*. **5**:2819-2838.

**Skyldberg B, Salo S, Eriksson E, Aspenblad U, Moberger B, Tryggvason K & Auer G**. 1999. Laminin-5 as a marker of invasiveness in cervical lesions. *J Natl Cancer Inst*. **91**:1882-1887.

**Slebos RJ, Yi Y, Ely K, Carter J, Evjen A, Zhang X, Shyr Y, Murphy BM, Cmelak AJ, Burkey BB, Netterville JL, Levy S, Yarbrough WG & Chung CH**. 2006. Gene expression differences associated with human papillomavirus status in head and neck squamous cell carcinoma. *Clin Cancer Res*. **12**:701-709.

**Slonim DK**. 2002. From patterns to pathways: gene expression data analysis comes of age. *Nat Genet*. **32 Suppl**:502-508.

**Southern SA, Noya F, Meyers C, Broker TR, Chow LT & Herrington CS**. 2001. Tetrasomy is induced by human papillomavirus type 18 E7 gene expression in keratinocyte raft cultures. *Cancer Res*. **61**:4858-4863.

**Sparkowski J, Anders J & Schlegel R**. 1995. E5 oncoprotein retained in the endoplasmic reticulum/cis Golgi still induces PDGF receptor autophosphorylation but does not transform cells. *Embo J*. **14**:3055-3063.

**Stanhope-Baker P, Kessler PM, Li W, Agarwal ML & Williams BR**. 2004. The Wilms tumor suppressor-1 target gene podocalyxin is transcriptionally repressed by p53. *J Biol Chem*. **279**:33575-33585.

**Stanley MA, Browne HM, Appleby M & Minson AC**. 1989. Properties of a non-tumorigenic human cervical keratinocyte cell line. *Int J Cancer*. **43**:672-676.

**Steenbergen RD, Kramer D, Braakhuis BJ, Stern PL, Verheijen RH, Meijer CJ & Snijders PJ**. 2004. TSLC1 gene silencing in cervical cancer cell lines and cervical neoplasia. *J Natl Cancer Inst*. **96**:294-305.

**Steenbergen RD, OudeEngberink VE, Kramer D, Schrijnemakers HF, Verheijen RH, Meijer CJ & Snijders PJ**. 2002. Down-regulation of GATA-3 expression during human papillomavirus-mediated immortalization and cervical carcinogenesis. *Am J Pathol*. **160**:1945-1951.

**Steinberg BM & DiLorenzo TP**. 1996. A possible role for human papillomaviruses in head and neck cancer. *Cancer Metastasis Rev*. **15**:91-112.

**Stoppler H, Stoppler MC & Schlegel R**. 1994. Transforming proteins of the papillomaviruses. *Intervirology*. **37**:168-179.

**Sutherland LC, Rintala-Maki ND, White RD & Morin CD**. 2005. RNA binding motif (RBM) proteins: a novel family of apoptosis modulators? *J Cell Biochem*. **94**:5-24.


**Syrjanen SM & Syrjanen KJ**. 1999. New concepts on the role of human papillomavirus in cell cycle regulation. *Ann Med*. **31**:175-187.


**Takamizawa J, Konishi H, Yanagisawa K, Tomida S, Osada H, Endoh H, Harano T, Yatabe Y, Nagino M, Nimura Y, Mitsudomi T & Takahashi T**. 2004. Reduced expression of the let-7 microRNAs in human lung cancers in association with shortened postoperative survival. *Cancer Res*. **64**:3753-3756.


**Tan PK, Downey TJ, Spitznagel EL, Jr., Xu P, Fu D, Dimitrov DS, Lempicki RA, Raaka BM & Cam MC**. 2003. Evaluation of gene expression measurements from commercial microarray platforms. *Nucleic Acids Res*. **31**:5676-5684.


**Tang S, Tao M, McCoy JP, Jr. & Zheng ZM**. 2006. The E7 oncoprotein is translated from spliced E6*I transcripts in high-risk human papillomavirus type 16- or type 18-positive cervical cancer cell lines via translation reinitiation. *J Virol*. **80**:4249-4263.


**Thomas M, Pim D & Banks L**. 1999. The role of the E6-p53 interaction in the molecular pathogenesis of HPV. *Oncogene*. **18**:7690-7700.


**Tran N, Rose BR & O'Brien CJ**. 2007. Role of human papillomavirus in the etiology of head and neck cancer. *Head Neck*. **29**:64-70.


**van 't Veer LJ, Dai H, van de Vijver MJ, He YD, Hart AA, Mao M, Peterse HL, van der Kooy K, Marton MJ, Witteveen AT, Schreiber GJ, Kerkhoven RM, Roberts C, Linsley PS, Bernards R & Friend SH**. 2002. Gene expression profiling predicts clinical outcome of breast cancer. *Nature*. **415**:530-536.


**van de Vijver MJ, He YD, van't Veer LJ, Dai H, Hart AA, Voskuil DW, Schreiber GJ, Peterse JL, Roberts C, Marton MJ, Parrish M, Atsma D, Witteveen A, Glas A, Delahaye L, van der Velde T, Bartelink H, Rodenhuis S, Rutgers ET, Friend SH & Bernards R**. 2002. A gene-expression signature as a predictor of survival in breast cancer. *N Engl J Med*. **347**:1999-2009.

**Venuti A, Badaracco G, Rizzo C, Mafera B, Rahimi S & Vigili M**. 2004. Presence of HPV in head and neck tumours: high prevalence in tonsillar localization. *J Exp Clin Cancer Res*. **23**:561-566.


**Vikram B, Strong EW, Shah JP & Spiro R**. 1984. Second malignant neoplasms in patients successfully treated with multimodality treatment for advanced head and neck cancer. *Head Neck Surg*. **6**:734-737.


**Volinia S, Calin GA, Liu CG, Ambs S, Cimmino A, Petrocca F, Visone R, Iorio M, Roldo C, Ferracin M, Prueitt RL, Yanaihara N, Lanza G, Scarpa A, Vecchione A, Negrini M, Harris CC & Croce CM**. 2006. A microRNA expression signature of human solid tumors defines cancer gene targets. *Proc Natl Acad Sci U S A*. **103**:2257-2261.


**Walboomers JM, Jacobs MV, Manos MM, Bosch FX, Kummer JA, Shah KV, Snijders PJ, Peto J, Meijer CJ & Munoz N**. 1999. Human papillomavirus is a necessary cause of invasive cervical cancer worldwide. *J Pathol*. **189**:12-19.


**Walter M, Davies JP & Ioannou YA**. 2003. Telomerase immortalization upregulates Rab9 expression and restores LDL cholesterol egress from Niemann-Pick C1 late endosomes. *J Lipid Res*. **44**:243-253.


**Watabe K, Ito A, Koma YI & Kitamura Y**. 2003. IGSF4: a new intercellular adhesion molecule that is called by three names, TSLC1, SgIGSF and SynCAM, by virtue of its diverse function. *Histol Histopathol*. **18**:1321-1329.


**Wei CL, Wu Q, Vega VB, Chiu KP, Ng P, Zhang T, Shahab A, Yong HC, Fu Y, Weng Z, Liu J, Zhao XD, Chew JL, Lee YL, Kuznetsov VA, Sung WK, Miller LD, Lim B, Liu ET, Yu Q, Ng HH & Ruan Y**. 2006. A global map of p53 transcription-factor binding sites in the human genome. *Cell*. **124**:207-219.


**Werness BA, Levine AJ & Howley PM**. 1990. Association of human papillomavirus types 16 and 18 E6 proteins with p53. *Science*. **248**:76-79.


**Wong DT & Munger K**. 2000. Association of human papillomaviruses with a subgroup of head and neck squamous cell carcinomas. *J Natl Cancer Inst*. **92**:675-677.

**Wu EW, Clemens KE, Heck DV & Munger K**. 1993. The human papillomavirus E7 oncoprotein and the cellular transcription factor E2F bind to separate sites on the retinoblastoma tumor suppressor protein. *J Virol*. **67**:2402-2407.

**Xi S, Dyer KF, Kimak M, Zhang Q, Gooding WE, Chaillet JR, Chai RL, Ferrell RE, Zamboni B, Hunt J & Grandis JR**. 2006. Decreased STAT1 expression by promoter methylation in squamous cell carcinogenesis. *J Natl Cancer Inst*. **98**:181-189.

**Xi Y, Shalgi R, Fodstad O, Pilpel Y & Ju J**. 2006. Differentially regulated micro-RNAs and actively translated messenger RNA transcripts by tumor suppressor p53 in colon cancer. *Clin Cancer Res*. **12**:2014-2024.

**Xiang Z, Yang Y, Ma X & Ding W**. 2003. Microarray expression profiling: analysis and applications. *Curr Opin Drug Discov Devel*. **6**:384-395.

**Xie D, Jauch A, Miller CW, Bartram CR & Koeffler HP**. 2002. Discovery of over-expressed genes and genetic alterations in breast cancer cells using a combination of suppression subtractive hybridization, multiplex FISH and comparative genomic hybridization. *Int J Oncol*. **21**:499-507.

**Yamamichi N, Yamamichi-Nishina M, Mizutani T, Watanabe H, Minoguchi S, Kobayashi N, Kimura S, Ito T, Yahagi N, Ichinose M, Omata M & Iba H**. 2005. The Brm gene suppressed at the post-transcriptional level in various human cell lines is inducible by transient HDAC inhibitor treatment, which exhibits antioncogenic potential. *Oncogene*. **24**:5471-5481.

**Yee C, Krishnan-Hewlett I, Baker CC, Schlegel R & Howley PM**. 1985. Presence and expression of human papillomavirus sequences in human cervical carcinoma cell lines. *Am J Pathol*. **119**:361-366.

**Yousef GM, Yacoub GM, Polymeris ME, Popalis C, Soosaipillai A & Diamandis EP**. 2004. Kallikrein gene downregulation in breast cancer. *Br J Cancer*. **90**:167-172.

**Zaug AJ & Cech TR**. 1986. The intervening sequence RNA of Tetrahymena is an enzyme. *Science*. **231**:470-475.

**Zhang B, Spandau DF & Roman A**. 2002. E5 protein of human papillomavirus type 16 protects human foreskin keratinocytes from UV B-irradiation-induced apoptosis. *J Virol*. **76**:220-231.

**Zhang L, Huang J, Yang N, Greshock J, Megraw MS, Giannakakis A, Liang S, Naylor TL, Barchetti A, Ward MR, Yao G, Medina A, O'Brien-Jenkins A, Katsaros D, Hatzigeorgiou A, Gimotty PA, Weber BL & Coukos G**. 2006. microRNAs exhibit high frequency genomic alterations in human cancer. *Proc Natl Acad Sci U S A*. **103**:9136-9141.

**Zhao R, Gish K, Murphy M, Yin Y, Notterman D, Hoffman WH, Tom E, Mack DH & Levine AJ**. 2000. Analysis of p53-regulated gene expression patterns using oligonucleotide arrays. *Genes Dev*. **14**:981-993.

**Zhu H, Cong JP, Mamtora G, Gingeras T & Shenk T**. 1998. Cellular gene expression altered by human cytomegalovirus: global monitoring with oligonucleotide arrays. *Proc Natl Acad Sci U S A*. **95**:14470-14475.

**zur Hausen H**. 1991. Viruses in human cancers. *Science*. **254**:1167-1173.

**zur Hausen H**. 2000. Papillomaviruses causing cancer: evasion from host-cell control in early events in carcinogenesis. *J Natl Cancer Inst*. **92**:690-698.

**zur Hausen H**. 2002. Papillomaviruses and cancer: from basic studies to clinical application. *Nat Rev Cancer*. **2**:342-350.