A Transcriptome- and Marker-Based Systemic Analysis of Cervical Cancer

Carlos G. Acevedo-Rocha, José A. Munguía-Moreno, Rodolfo Ocádiz-Delgado and Patricio Gariglio

1 Max-Planck-Institut für Kohlenforschung, Organische Chemie
2 Philipps-Universität-Marburg, Fakultät für Chemie
3 Departamento de Genética y Biología Molecular, CINVESTAV-IPN
1,2 Germany
3 Mexico

1. Introduction

The 20th century witnessed a great development of genetics and molecular biology, laying the foundations for a new era in medicine. The elucidation of the mechanism of heredity, for example, helped us understanding the connection between cells, chromosomes, DNA and the genetic code, an historical journey to the center of biology (Lander & Weinberg, 2000). This process strongly consolidated when “the central dogma of molecular biology” (Crick, 1970) was proposed long time ago, whereby the genetic information flows from DNA to RNA to protein. Since then, however, our understanding in the molecular and cellular organization, as well as physiology of living systems has radically changed, partially challenging the validity of the central ‘dogma’—by the way, dogma strictly means a belief that people are expected to accept without any doubts, a word to be expectedly seen outside the scientific method lexicon – of molecular biology (Shapiro, 2009). The main paradigm is that cells are able to make decisions based on actively sensing their environment; hence, information processing in living systems can be regarded at least bidirectional. In any case, the recent sequencing of the human genome is a great milestone (Human Genome Sequencing, 2004), whereby the language of the “common thread of humanity” in this new medicine era is just “the end of the beginning” (Stein, 2004).

Genomics studies the total DNA sequence of an organism. Of the approximately 3,000 million base pairs that comprise the human genome, only 1% was firstly estimated to correspond to as low as 25,000 proteins (Southan, 2004), a number that has been changing since the initial sequence drafts of the Human Genome Project (HGP). One motivation behind genome-sequencing projects is the assumption that the nucleotide sequence of an organism provides a description of the genes, its products and interaction networks that orchestrate programs like those sustaining the metabolic activity of a cell or deploying a body plan. However, new discoveries in transcriptome functions significantly expand—and even challenge—the classical concept of the gene and how post-transcriptional molecular events are becoming key to understand gene regulation in higher eukaryotes.
The success of the HGP has provided a blueprint of genes encoding the entire human protein set potentially expressed in any of the approximately 230 cell types comprising the human proteome. Considering that both the current and sometimes limited knowledge of only two-thirds of the 20,300 protein-coding human genes mapped through the HGP is at hand (Legrain et al., 2011), the recently launched Human Proteome Project (HPP) aims to provide for the remaining one-third of proteins experimental evidence related to abundance, distribution, subcellular localization, functions, and interactions (Bustamante et al., 2011).

In the current "post-genomic era" scientists aim not only to build a catalog of all genes, but also to translate the knowledge obtained into benefits for humanity (Collins et al., 2003). By examining tumors at the genomic, transcriptomic, and proteomic levels, for instance, it is possible to better understand cancer biology and improve patient care, diagnosis, prognosis, and therapy (Lin & Li, 2008). Importantly, one key development that has emerged between the interface of the HGP and the HPP is the area of functional genomics or transcriptomics, which aims to assign a function to all transcripts. But this is not a trivial task because talking about transcriptomes involves considering these as entities as diverse as the cell types, developmental stages, environmental conditions and pathological states that an organism harbors or faces. Therefore, we must include a global vision for the process of transcription, i.e. the process by which information contained in DNA is converted (or transcribed) into RNA and how this process is regulated by protein(s) (Fig. 1).

Importantly, it should bear on mind that 57% - a scalable number up to 90% (Costa, 2010) - of the genome is transcribed into RNA but does not code for proteins (Frith et al., 2005). Moreover, very recently non-coding RNAs (microRNAs, small RNAs, small interfering RNAs or siRNAs as well as medium and large RNAs) have emerged as key elements in carcinogenesis. The amazing complexity of the transcriptome and its expansion (Mendes Soares & Valcarcel, 2006), has led to scientists eager to hunt transcriptomes. Fortunately, there are tools to examine the expression of genes at many levels, allowing us to globally understand complex diseases like cancer.

The current manuscript introduces the most common techniques to study the transcription of the 1% protein-coding genes encoded in the human genome, followed by a review of microarray studies that had provided invaluable information of the carcinogenesis of cervical cancer (CC), the most and second most common cancer disease in women from the developing and developed world, respectively. The integration of all this information is very important to not only understand CC from a global perspective, but also to identify key tumor markers that could help for CC diagnosis, prognosis and/or therapy, as discussed in the last part of the manuscript. As for cancer progression involving noncoding RNAs – importantly considered the “masters of regulation” (Costa, 2010), the reader is encouraged to read an excellent recent review (Gibb et al., 2011).

Importantly, CC is largely associated to Human Papillomavirus (HPV) infection, from which there are over hundred types but of these 40 infecting the genital tract and 15 of high-risk related to the development of CC. Thus, HPV is a common sexually transmitted agent after a woman starts her first sexual relationship and responsible of ca. 30% of the global cancer burden associated to infective agents (20% of the total) (zur Hausen, 2009).
2. Probing the transcriptome

The relationship between a particular molecule and cellular phenotype has allowed us to better understand the molecular mechanisms of complex diseases such as cancer. In the course of molecular biology many useful techniques to analyze DNA, RNA and proteins were developed. For about half century, reasonably, the practice of molecular biology was comfortable with its reductionism; however, in the coming era of genomics, the tendency to probe in a single experiment hundreds or thousands of biomolecules allows us talking of two mechanisms: (i) The “reductionist mechanism” employs tools to analyze one or few different molecules in a single experiment; it is a slow but comprehensive conclusions can be reliably obtained; (ii) The “holistic mechanism” allows the assessment of thousands of different molecules in a single experiment; it is a fast mechanism but the obtained
hypotheses remained to be tested (Coulton, 2004). While single gene analyses gradually shifted towards large mutational screens and complete genome mapping, whole genome sequencing moved towards bioinformatics with exhaustive functional genomics and proteomics data. Systems biology aims to understand this complexity. Ironically, the holism in systems biology has re-emerged out of the traditional molecular biology, carrying with it the *reductionism-holism* debate since the past years (Gatherer, 2010). Interestingly, it has been boldly argued that traditional molecular biology represents a greedy reductionist approach (to some authors a naively reductionist one) that requires either extensive complementation from, or even replacement by systems biology. However, as we discuss along the text, it is more meaningful to combine both approaches.

The study of transcription is important because the levels of mRNA transcripts in a cell correlate frequently with the expression levels of the corresponding proteins. There are several techniques used in transcriptomics, which are based on gene amplification by the polymerase chain reaction (PCR), hybridization and sequencing. All these tools permit analyzing differential expression, and determine what transcripts are mainly expressed in cancerous tissue in comparison with normal tissue and *vice versa*. This is important because knowing what and how genes are differentially expressed suggests that these may play an important role in carcinogenesis. This scenario can be found in the case of proto-oncogenes and anti-oncogenes (or tumor suppressor genes) that promote and prevent cell growth, respectively. In other words, the levels of expression of many oncogenes (normally known as proto-oncogenes) may be very high and the levels of expression of tumor suppressor genes may be low. Following the reductionist and holistic classification, the most common techniques used in transcriptomics can be classified into high, medium, and low performance, with respect to its ability to analyze different molecules in a single experiment.

2.1 Low-throughput techniques

One of the first developed methods to detect a mRNA transcript was *in situ* hybridization (ISH) (Harrison *et al.*, 1973). ISH requires labelling either fluorescently or radioactively a RNA or complementary DNA (cDNA) corresponding to the transcript of interest. Through the formation of hybrid cDNA:RNA or RNA:RNA duplexes, the amount of the specific transcript can be determined as well as its cellular position can be localized. Thereafter, the popular technique Northern Blot (NB) was developed. NB uses a labeled probe that recognizes the transcript of interest in a similar manner to the ISH, but the hybridization is performed on cellulose, because the RNA of a tissue is previously separated by electrophoresis and transferred to a special paper surface. If the transcript of interest forms a hybrid with the radioactively labeled probe, it will reveal the presence of a band in a autoradiography upon exposure (Alwine *et al.*, 1977). Because of its sensitivity, this technique is commonly used in molecular biology. Another similar technique, called ribonuclease protection assay (RPA), is based on hybrid formation between the mRNA of the gene in question and a labeled probe (RNA or cDNA), being the non-hybridized single-strand RNA part degraded by a RNAse enzyme (Berk & Sharp, 1977). This way, the hybrids can be detected because the RNA chain is radiolabelled; this method is 50 times more sensitive than NB (Bartlett, 2002).

Another old technique is subtractive hybridization (SH), which employs single-strand RNA or cDNA labeled probes. Using SH one can remove commonly expressed genes between
two samples (e.g. cancerous and normal tissue) by hybrid formation between cDNA:RNA and identified those differentially expressed genes in a particular tissue (Zimmermann et al., 1980). The tumor-suppressor gene p21WAF1/CIP1, also known as CDKN1A, involved in the negative regulation of the cell cycle as well as the induction of apoptosis, was identified using SH (el-Deiry et al, 1993). Finally, the Retro-Transcription coupled to PCR (RT-PCR) allows the amplification of a cDNA synthesized from a specific mRNA using a reverse transcriptase (Rappolee et al., 1988). RT-PCR can also be applied to tissues (in situ RT-PCR) similarly to the ISH but the sensitivity differs: While ISH can detect from 20 to 200 copies of transcript per cell, in situ RT-PCR can detect one transcript per cell (Bartlett, 2002). The enormous sensitivity of RT-PCR has allowed the development of a technique to quantify quickly and accurately the amount of transcripts in a given biological sample. It is called quantitative RT-PCR or Real-Time PCR (qRT-PCR) (Bustin, 2000). All these methods mainly based on hybridization and PCR can generally characterize one transcript per experiment.

2.2 Medium-throughput techniques

When a mRNA is converted to cDNA, the fragments obtained can be cloned or inserted into a vector (plasmid), which can be introduced into bacteria to obtain many copies of the transcript. At the end, the fragment of interest must be sequenced. In this way, various types of sequences can be generated: A EST (Expressed Sequence Tags) corresponds to an arbitrary portion of a cDNA sequence, i.e. a random sequence that allows identification of a transcript (Adams et al., 1991), whereas "ORESTES" (Open Reading Frame Expressed Sequence Tags) contain an open reading frame, which generally corresponds to a central portion of the cDNA sequence (Dias Neto et al., 2000); it is also possible to alternatively clone the entire sequence of cDNA without tag (Strausberg et al., 2002). Importantly, all these partial or complete cDNA sequences had enabled the characterization of large numbers of transcripts and their differential expression depending on their frequency and tissue of origin.

Similarly, the techniques of Differential Display (DD) and Representational Differential Analysis (RDA) permit the identification of differentially expressed transcripts e.g. coming from different sources or coming from the same source but subjected to different conditions. DD is essentially based on a series of RT-PCR amplifications where the transcripts of two samples are fluorescently or radioactively labeled, compared by electrophoresis, selected and finally sequenced (Liang & Pardee, 1992). The RDA technique is based on SH and RT-PCR, so that common transcripts between two samples are removed after the formation of hybrid cDNA:cDNA and genes only expressed in a tissue are amplified in a sensitive and accurate way (Hubank & Schatz, 1994). Since both techniques are of easy accessibility and use, their use has allowed the identification of many genes altered in cancer (Liang & Pardee, 2003; Hollestelle & Schutte, 2005). For example, while the Cyclin G was identified using the DD technique (Okamoto & Beach, 1994) the anti-oncogene PTEN was characterized through RDA (Li et al., 1997). The medium-throughput methods basically depend on sequencing and differ from those of low-performance because many transcripts can be characterized at a single experiment, but not as many as when using high-performance ones.

2.3 High-throughput techniques

In general, these methods are based on sequencing and hybridization. Sequencing includes Serial Analysis of Gene Expression (SAGE) and Massively Parallel Signature Sequencing
(MPSS) and, in the case of hybridization, the best example is DNA microarrays. SAGE is similar to the sequencing of ESTs or cDNA clones, but the performance is much higher because in a single vector a lot of small tags corresponding to different mRNAs can be inserted. After sequencing, the abundance of these tags can be measured doing a bioinformatics analysis, whereby the fold expression change of a gene in different tissues/conditions can be estimated (Velculescu et al., 1995). MPSS is similar to SAGE but the main difference is that in the former small tags are attached to microbead arrays, increasing the capacity of the system (Brenner et al., 2000). Although MPSS is similar to SAGE, the later method has been widely used, uncovering many genes with a potential role in cancer (Yamashita et al., 2008), and allowing the identification of known oncogenes such as ERBB2 and EGFR (Polyak & Riggins, 2001; Forrest et al., 2006).

The DNA microarrays are a set of gene sequences (which may correspond to transcripts) arranged on a flat surface. There are two types of DNA microarrays: cDNA microarrays, in which transcripts of interest are amplified by PCR and deposited on sites identified in a paper or small glass slide (Schena et al., 1995) and oligonucleotide microarrays, in which small sequences corresponding to a gene are synthesized and arranged on a particular area of a slide (Lockhart et al., 1996; Singh-Gasson et al., 1999; Hughes et al., 2001). While the former arrays are normally produced in-house by researchers, the latter one are usually obtained from companies, being the most known “Affymetrix”. During the experiment, the mRNA of a tissue of interest is firstly converted into cDNA and labeled either radioactively or fluorescently. Then, through the formation of hybrids between the labeled cDNA and the unlabelled cDNA or oligonucleotides attached to the surface, differentially expressed genes between two samples can be identified. Finally, the ratios of frequency can be estimated using different bioinformatics methods. Both cDNA and oligonucleotide microarrays have been widely used, the difference lies in the number of genes per square centimeter: On paper there may be hundreds of genes, whereas in a glass slide it is possible to bear sequences representing up to 10,000 and 25,000 genes in the case of cDNA and oligonucleotide microarrays, respectively. This allows the simultaneous quantification of thousands of gene transcripts in two samples when they are tagged with different fluorophores, for example, if the transcripts from tumor cells are stained with red (e.g. Cy5) and those from normal cells with green (e.g. Cy3), upon locating spots on a cDNA microarray, while the red and green ones would respectively correspond to genes differentially expressed in the tumor and normal tissue, the yellow (and alike degrees of color) would correspond to genes similarly expressed in both tissues. This is usually done on cDNA microarrays because the spots can be compared directly in one experiment, but in the case of oligonucleotide microarrays, the spots are compared indirectly in separate experiments because the detection and analysis methods differ. In either case, the different spot intensities can be transformed into transcript levels present in each sample. The numerical data are analyzed with a computer and mathematical algorithms, allowing various genes to display a characteristic pattern or “Gene Expression Profile” (GEP) related to the phenotype of the different samples. Depending on the intensity in which the various genes from the GEP are expressed, the sample acquires a particular “expression signature”.

The transcriptome should study not only the expression of transcripts, but also the DNA sites where transcription factors bind as well as chromatin modifications that regulate gene expression. Chromatin Immunoprecipitation (ChIP) is a old technique to identify genes that can be activated by a protein in vivo (Orlando, 2000), but can be of high-throughput when it
is coupled with: i) DNA microarrays (Ren et al., 2000), also known as "ChIP-on-Chip", for instance, many genes that can be activated by the transcription factors E2F have been identified (Bracken et al., 2004); or 2) Sequencing-based techniques like Paired-end di-tags (PET) that is equivalent to SAGE but in contrast to a tag, two gene extremes are joined (Ng et al., 2005). Using ChIP-PET, several TP53-regulated genes have been identified (Wei et al., 2006). TP53 and E2F are the most important transcription factors known in cancer development, activating or deactivating genes involved in cell cycle and apoptosis.

Last but not least, another successful tool combined with microarrays is Laser Capture Microdissection (LCM), which uses a laser beam targeted to specific tissue sections under microscopic control to isolate cell clusters, allowing the molecular comparison of cell populations that are histologically or pathologically distinct but topographically contiguous (Kalantari et al., 2009). The main limitation of this technique, however, is that it requires trained personnel to visually select cell populations of interest. One approach to increase dissection performance is to utilize molecular probes to facilitate the process. Expression microdissection (xMD) is such an example, where an antibody is used for cell targeting in place of an investigator (Tangrea et al., 2004; Hanson et al., 2011). In fact, large numbers of cells can be greatly analyzed by using the recently described SIVQ feature matching algorithm, making possible the development of a high-throughput cell procurement instrument. This approach permits histologically constrained morphologies (e.g. automated selection of only the malignant epithelium of solid tissue tumours) to be acquired in a semi-autonomous fashion, allowing the generation of large, preparative quantities of DNA, RNA, or protein for subsequent high-throughput analysis. In fact, SIVQ–LCM holds unique potential as a discovery tool for molecular pathology, since individual cells with particular computer-defined morphologic features can be microdissected and profiled, thus generation new integrated and composite morphological data types (e.g. morpho-genomics or -proteomics) (Hipp et al., 2011). Importantly, there is increasing evidence demonstrating the necessity of upfront malignant cell enrichment techniques for specific molecular profiles, being especially desirable for clinical trials that require accurate, disease cell-specific molecular measurements (Harrell et al., 2008; Klee et al., 2009; Silvestri et al., 2010). This technique has opened new and promising avenues to molecularly enquire histology and pathology in many fields of cancer research (Fuller et al., 2003; Domazet et al., 2008).

All the techniques mentioned above (Fig. 2) have favorable characteristics, while the high-throughput methods have a great capacity for data management; the low-throughput ones confer higher specificity, sensitivity, and reproducibility. Due to this, high- and medium-performance techniques are complementary, but they must be validated with those of low-performance. These tools have generated much information that should be integrated to extract biological meaning, allowing the complete characterization of the transcriptome of a cell. Indeed, a complete integrative analysis of the cancer transcriptome cannot only be obtained by analyzing the genome, transcriptional networks and the interactome, (Rhodes & Chinnaiyan, 2005), but also by delineating the subtypes of cancer obtained from DNA microarrays with relation to a particular phenotype.

3. A brief overview on microarrays and cancer

Microarrays are one of the most versatile tools used in transcriptomics, whereby many benefits for oncogenomics have been found. For example, thanks to the determination of
Fig. 2. Probing the transcriptome at different performances. a) Upon DNA transcription, messenger RNA (mRNA) molecules can be analyzed in a single experiment: i) For one or few transcripts, low-throughput methods include *in situ* hybridization (ISH), subtractive hybridization (SH), “Northern Blot” (NB), Ribonuclease-Protection Assay (RPA), Reverse-Transcriptase Polymerase Chain Reaction (RT-PCR), *in situ* RT-PCR and quantitative or real time RT-PCR (qRT-PCR). ii) For various transcripts there are medium-throughput tools based on cDNA clones, Expressed Sequence Tags (ESTs), Open Reading Frame ESTs (ORESTES), Differential Display (DD) and Representative Differential Analysis (RDA). iii) For thousands of transcripts, Serial Analysis of Gene Expression (SAGE), Massive-Parallel Signature Sequencing as well as DNA and oligonucleotide microarrays are high-throughput approaches. b) Transcription regulation: To identify Transcription Factors (TF) that bind specifically to DNA sites, one can use Chromatin Immunoprecipitation in a low- (ChIP) or high-throughput manner when it is coupled to microarrays (ChIP-to-Chip) or Pair-End di-Tags (ChIP-PET). These methods are classified according to the holism-reductionism approach.

Gene Expression Profiles (GEPs) using DNA microarrays, a new molecular classification and subclassification, as well as clinical prediction and diagnosis of many cancer (sub)types...
A Transcriptome- and Marker-Based Systemic Analysis of Cervical Cancer

have been developed (Macoska, 2002; Ciro et al., 2003; Wadlow & Ramaswamy, 2005). Likewise, new potential markers for therapy have been identified and there is a better understanding of the molecular mechanisms of cancer (Clarke et al., 2004). There are classic studies that have demonstrated the potentials of microarray technology, for instance, one of the first reports was the molecular classification of human acute leukemias using an oligonucleotide microarray (Affymetrix) representing 6817 genes (Golub et al., 1999). In this study, 50 genes were found aid to distinguish between acute myeloid leukemia and acute lymphoblastic leukemia. To validate the gene set, 34 samples were analyzed without knowledge of its type (unsupervised analysis) and classified in their respective type with a high accuracy. This was a very important achievement because the right diagnosis of this cancer is often difficult but essential to discern because an effective treatment relies on an accurate identification of the cancer subtype.

Another classical study was applied on the Diffuse Large-B-Cell Lymphoma since it is known that patients exhibit different prognoses and variable responses to therapy. Using a microarray containing over 18,000 cDNA clones, a GEP with little more than 100 genes and 96 different samples was established (Alizadeh et al., 2000). This pattern allowed the classification of this cancer into two subtypes regarding the status of differentiation of B cells: one similar to germ B cells and other similar to activated B cells in vitro. Interestingly, the two subtypes showed a strong correlation with clinical prognosis, which was the best for the subtype bearing germinal B cells. These patients are usually treated with a combination of chemotherapy based on anthracyclines, but if they don’t have a good prognosis then a bone marrow transplantation is rather recommended. Therefore, the GEP of about hundred genes can help to determine what kind of treatment and prognosis a patient should have. Thereafter, this work was validated by using 240 samples that allowed the identification of only 17 genes capable to correlate disease with prognosis (Rosenwald et al., 2002). Similarly, another laboratory studied the prognosis of the same cancer type but whose patients received different treatments, allowing the identification of two groups of patients with different life expectancy for 5 years (72% good versus 12% bad prognosis) using only a predictor of 13 out of 6,817 genes included in a "Genechip" from Affymetrix (Shipp et al., 2002). It is noteworthy that 3 tumor markers were detected in both the 17 and 13 gene predictors developed independently by those laboratories.

The best example of GEPs, nonetheless, has been demonstrated in the prognosis of breast cancer. Using an oligonucleotide microarray of 25,000 and 78 samples of primary breast tumors obtained from patients with negative lymph node status for metastasis, a 70-gene "poor-prognosis" molecular signature was identified (van 't Veer et al., 2002). This signature corresponds to a high probability of developing metastasis in the short term and most likely die. What is interesting about this study is that tumors are not “good” nor “bad” when the disease progresses as was proposed not so long time ago with the clonal model of development (Couzin, 2003); rather, the malignant cell is destined to metastasize very early. Through this genetic signature, experts can decide what patients should receive adjuvant therapy consisting of Tamoxifen (an antagonist of the estrogen receptor in breast tissue via its active metabolite, hydroxytamoxifen). Shortly after, this study was clinically validated using 217 new samples, which confirmed that the signature of 70 genes is the best criterion for deciding whether a patient requires adjuvant therapy or not (van de Vijver et al., 2002).
Since the first two studies were developed using samples from young patients with relatively early tumours from the same institution, it was not clear whether the 70-gene could also be applied to other patients. Interestingly, the TRANSBIG consortium, a network of 28 institutions promoting international collaboration in translational research across 11 countries, independently validated the 70-gene signature using 302 samples from patients from different age groups (up to 61 years) and from 5 different European hospitals (Buyse et al., 2006). Despite its achievements, the same group questioned whether this 70-gene signature could be used as a standard high-throughput diagnostic test, so, using the samples from the first two mentioned reports, they validated a customized mini-array containing a reduced set of 1,900 probes known as the “MammaPrint” (Glas et al., 2006). The “MammaPrint” prognostic assay is currently being validated under the clinical MINDACT (Microarray in Node-Negative Disease May Avoid Chemotherapy) randomized trial that includes 6,000 patient samples from various centers, even though the 70-gene signature has been validated several times in patients with negative (Bueno-de-Mesquita et al., 2009) or positive (Mook et al., 2009) lymph-node status as well as from other populations, including Japanese (Ishitobi et al., 2010). Remarkably, the MammaPrint 70-gene signature, whose genes reflect the hallmarks of cancer (Tian et al., 2010), can be considered as a milestone in the personalized care for breast cancer patients (Slodkowska & Ross, 2009).

4. Microarrays and cervical cancer

The origin of cervical cancer (CC) is linked to the infection of High-Risk Human Papilloma Virus (HR-HPV) mainly type 16 and 18. The genome of these viruses contain 8 viral oncogenes, 2 of which code for the early-expressed oncoproteins E6 and E7 that inhibit the activity of the anti-oncoproteins p53 and pRb, respectively. This way, the oncoproteins deregulate the necessary balance between proliferation and apoptosis, promoting the development of cancer. These imbalances have been studied at the transcriptional level and in a comprehensive manner using microarrays in both clinical samples and cell lines derived from CC with and without therapy. Although there are much fewer reports of microarrays compared to other tissues e.g. for every CC microarray paper, there are 7 for breast cancer (Acevedo Rocha et al., 2007); these few studies have provided invaluable information on the molecular mechanisms of CC.

4.1 Studying carcinogenesis using in vitro HPV models

A key event in the development of CC is the infection by HR-HPVs. Using microarray technology, gene expression profiles in cell lines as well as keratinocytes containing HR-HPVs have been assessed (Chang & Laimins, 2000; Nees et al., 2000; Nees et al., 2001; Duffy et al., 2003; Garner-Hamrick et al., 2004; Lee et al., 2004; Toussaint-Smith et al., 2004). Similarly, the overall effect upon infection of cultured human keratinocytes with low-risk HPVs (LR-HPVs) has been described (Thomas et al., 2001). Interestingly, in contrast to HR-HPVs, LR-HPVs induce the overexpression of a larger number of genes from the family TGF-β (Tumor Growth Factor) and apparently, LR-HPVs do not suppress interferon-inducible genes (Thomas et al., 2001). This is very interesting as members of the TGF-β family play a role as tumor suppressor genes (at least at the early development of CC) and interferons are key molecules that counteract viral infections mediated by the immune system. These findings help to explain why the LR-HPVs episomes, conversely to those of HR-HPVs, are easily eliminated in many cases.
Another important event in the carcinogenesis of CC is the integration of viral genomes into the cellular genome. It is known that upon viral DNA integration into the host genome, the E2 protein expression is usually lost. Since E2 normally represses both E6 and E7, its absence deregulates the latter oncoproteins. Using microarrays, the overall effect upon viral genome integration of HR-HPV type 16, 18, and 33 into cell lines and keratinocytes has been determined (Alazawi et al., 2002; Ruutu et al., 2002; Pett et al., 2006). Notably, these studies found that the integration of the viral genome into the host genome is a critical step because, besides the high chromosomal instability of the infected cells, interferon-inducible genes are accordingly activated, thus eliminating the cells containing mainly viral episomes but promoting the selection of the more unstable cells.

In addition, the overall effect of expressing E2 in some cervical carcinoma cell lines has been also determined (Thierry et al., 2004), inducing in some cases cellular senescence or exit to the G0 cell cycle phase (Wells et al., 2003). Last but not least, the general effect of eliminating the gene E6AP, an important gene involved in the E6-mediated TP53 protein degradation, has been also assessed in multiple CC-derived (HPV+) cell lines (Kelley et al., 2005). All the studies mentioned in this section have identified significant changes in the expression patterns of hundreds of genes including cyclins, kinases, oncopgenes, and anti-oncopgenes; some known to be involved in CC but other previously unknown, so all these gathered information is essential to systematically study the HPV-mediated CC carcinogenesis.

4.2 Studying carcinogenesis using patient samples

To identify key genes in the development of CC several strategies have been followed. Some of them had focused on the progress of the lesions while others had compared their origin, i.e. squamous and/or glandular lesions vs normal tissue. In any case, these studies had allowed the identification of gene expression profiles useful for the molecular classification and subclassification of CC.

In the first attempts to classify CC, an expression profile of only 18 differentially expressed genes involved in apoptosis, cell adhesion, and transcription regulation was found between cervical squamous cell carcinoma (SCC) and normal cervical tissue using a microarray of 588 genes (Shim et al., 1998). In another interesting study, employing a 10,000-gene microarray, 40 genes allowed the classification of 34 samples of patients into a normal and a tumoral group (Wong et al., 2003). Moreover, from the 34 samples, 16 could be sub-classified as patients with grade IB and IIB tumors, from which four genes displayed key expression levels in both the previous classification and subclassification, suggesting their role as possible tumor markers (Wong et al., 2003). In a similar analysis but using only 1,276 genes together with 10 samples of SCC and 20 of cervical intraepithelial neoplasia grade 3 (CIN3), a gene expression profile showed that, from all the samples corresponding to CIN3, some correlated with the progression to cancer while others did not, implying the existence of a new subdivision of precancerous lesions histologically indistinguishable (Sopov et al., 2004).

The selection and characterization of tumor samples is critical as this has permitted the establishment of significant gene expression differences between samples from squamous and glandular origin in both normal and pathological conditions (Contag et al., 2004). Obviously, these differences arise by the transcriptional activity of genes particularly expressed in the histological subtypes of CC, but other strategies had also compared the
expression profiles between normal and squamous (Cheng et al., 2002b; Chen et al., 2003; Wong et al., 2006) or glandular (Chen et al., 2003; Fujimoto et al., 2004; Chao et al., 2006) tumor samples. Importantly, with a correct histological characterization of the samples, other factors can also be correlated, for example, using more than 40 samples derived from invasive CC (HPV+), it was found that a high burden of viral DNA correlates with high levels of E6 and E7 transcripts, poor prognosis, genomic instability and overexpression of more than 100 genes related to the cell cycle, from which many were identified as oncogenes and at least 50 target genes for the relevant E2F transcription factor family (Rosty et al., 2005). Although the sample description in other studies has remained considerably poor (Ahn et al., 2004a; Guelaguetza Vázquez-Ortíz, 2005; Santin et al., 2005; Vazquez-Ortiz et al., 2005a; Vazquez-Ortiz et al., 2005b), these also had generated long lists of genes possibly important for the molecular study of CC.

Lastly, there are two more examples displaying the great power of microarray technology as these have enriched samples from cytological screening (Papanicolaou). For instance, by obtaining normal and cancerous cells from a cytobrush and from simple exfoliated cells, it was possible to identify known and potential tumor markers in epithelial cells (Hudelist et al., 2005) and CIN3 lesions (Steinau et al., 2005), respectively.

4.3 Treatment

In the CC treatment, besides surgery there is radiotherapy and chemotherapy. However, it's not possible to predict the individual response of patients. The ability of tumor cells to evade treatments suggests that there are different resistance-induced mechanisms. It is believed that by monitoring the genes involved in the resistance against therapy, will help not only to understand the molecular mechanisms of CC, but also to improve its treatment. Accordingly, depending on the gene expression profiles of tumor samples that indicate sensitivity to radiation or chemotherapy, it could be possible to classify patients, allowing a customized CC treatment (Chin et al., 2005).

4.3.1 Radiation

The survival of patients diagnosed with cervical cancer has been improved by combining radiotherapy and chemotherapy. However, it has been estimated that about 65% of patients can be cured with radiation alone (Usmani et al., 2005) but such patients have not been identified so far and therefore they suffer the unnecessary and lethal chemotherapy effects. The long-term goal of the first report using microarrays in combination with radiotherapy against CC, was to find a GEP that would help deciding whether a patient would benefit or not with this treatment, avoiding in this way chemotherapy (Achary et al., 2000). Using a microarray of 5,776 genes, 70 identified genes allowed the differentiation of two cell lines derived from a single carcinoma, which had been previously characterized as radiosensitive and radiotolerant. Interestingly, some genes were previously associated with a cellular response against radiation, suggesting a key role in therapy resistance (Achary et al., 2000). Likewise, it was possible to classify 19 samples with 100% accuracy in two groups: sensitive and tolerant to ionizing radiation (IR) by using 62 out of 23,000 (Kitahara et al., 2002). Moreover, from the genes identified in the previous study, it was found that low levels of the gene XRCC5, and its corresponding protein Ku80, correlated with a good prognosis in CC patients (Harima et al., 2003). Thereafter, but using instead 35 genes, the same group
classified samples from patients treated with radiation and hyperthermia in two groups: sensitive and tolerant (Harima et al., 2004). Importantly, not only the combined treatment offered a better prognosis than radiotherapy alone, but a long list of genes with a possible role in the molecular mechanisms associated with therapy was obtained.

There are other studies where samples of patients with CC were classified in radiotolerant or radiosensitive (Wong et al., 2003), as well as in different radiosensitivity degrees (Tewari et al., 2005). In addition, in vitro studies using human keratinocytes (Chen et al., 2002), cervical carcinoma cell lines lacking HPV (Liu et al., 2003) and harboring HPV type either 16 (Liu et al., 2003; Chung et al., 2005) or 18 (Crawford & Piwnica-Worms, 2001; Chaudhry et al., 2003) have been also useful to improve the understanding of the molecular mechanisms that occur when tumor cells are treated with IR. Moreover, high levels of cyclin D1 mRNA (a molecule that promotes the progression of cell cycle) and low mRNA levels of the “Insulin-like Growth Factor-Binding Protein 2” or IGFBP2 (protein that can inhibit or promote tumor growth in many cancers) (Hoeflich et al., 2001) correlate with a radioresistant phenotype in immortalized human keratinocytes and CC cell lines (Chen et al., 2002; Liu et al., 2003; Chung et al., 2005). Other up-regulated genes, primarily involved in the cell cycle, that were detected in patients and radio-resistant cell lines include GAPDH (Kitahara et al., 2002; Harima et al., 2004), E2F3 (Chaudhry et al., 2003; Liu et al., 2003), DDB1 (Chaudhry et al., 2003; Wong et al., 2003) and ICAM5 (Achary et al., 2000; Chung et al., 2005). However, cyclin B1 and D1 have been determined to be overexpressed in immortalized human keratinocytes and several CC-derived radio-resistant cell lines (Chen et al., 2002; Liu et al., 2003), but suppressed in radiosensitive cell lines (Crawford & Piwnica-Worms, 2001; Chaudhry et al., 2003).

Unfortunately, is difficult to find a clear correlation of differentially expressed genes between different microarray studies related to radiation therapy because the response is not only different in every patient, but it also depends on the dose, type, time, etc. In spite of this, other radiation-related tumor markers (Haffty & Glazer, 2003) have also been detected including cyclin D1 (CCND1), the factor vascular endothelial growth factor (VEGF) and the proliferating cell nuclear antigen (PCNA), though in isolated studies (Chen et al., 2002; Chaudhry et al., 2003; Liu et al., 2003).

4.3.2 Chemotherapy

Similar to radiation, there are several studies but using instead chemical agents. For example, using cell lines derived from CC with and without HPV infection, the effect of anticancer substances that stop cell cycle like lovastatin has been study in a comprehensive manner (Dimitroulakos et al., 2002). Other chemicals have been used like the apoptosis-inducing di-indol-methane (Carter et al., 2002), catechin EGCG (found in green tea) (Ahn et al., 2003), arsenic-derived (As2O3 and As4O6) (Ahn et al., 2004b), and platinum-derived compounds (Gatti et al., 2004) as well as the antibiotic zeocin (Hwang et al., 2005). In addition, several effects exerted by chemicals that inhibit the epidermal growth factor receptor (EGFR) oncogene (Woodworth et al., 2005) and phosphatidylinositol kinase (PIK3CA) (Lee et al., 2006) signaling pathways had been also assessed. However, since these compounds are highly toxic, with broad action spectra, similar to those of radiotherapy, only very slight correlations of activated or deactivated genes across all these studies can be observed. For example, the expression of pro-metastatic factor JAG2 is suppressed when CC
cell lines were treated with platinum-containing compounds (Gatti et al., 2004) or di-
indolylmethane (Carter et al., 2002). Di-indolylmethane (Carter et al., 2002) or arsenic
compounds (Ahn et al., 2004b), on the other hand, suppressed the transcripts of the
proliferation marker PCNA.

It has been likewise reported that the transcription factor E2F4 can be suppressed by the
competitive inhibition (in the ATP binding-site) of the EGFR (Woodworth et al., 2005) or
simply using zeocin (Hwang et al., 2005). Another gene involved in cell proliferation is
CHEK1, which can be suppressed by zeocin (Hwang et al., 2005) and derivatives of arsenic
(Ahn et al., 2004b). Lastly, the membrane marker CD83 (antigen involved in immunologic
response) has also been down-regulated using arsenic compounds (Ahn et al., 2004b) and
EGCG (Ahn et al., 2003). Despite efforts to improve the prognosis of patients through the use
of diverse chemotherapy regimens, radiation and their combinations, the quality of life,
generally speaking, has not been yet improved significantly (Duenas-Gonzalez et al., 2003).
Owed to this, the search for new tumor markers and the development of drugs specifically
targeted against these molecules is an important step to control CC.

5. A systematic view on cervical cancer

Systems biology (SB) seeks to explain biological phenomena through the study of networks
that emerge because of the interactions of the cellular and biochemical components of a cell
or organism (Kitano, 2002). This can be achieved with the aid of bioinformatics, as it allows
the integration of large amounts of information that are generated every day as well as the
construction of biology-oriented mathematical models. In fact, not only transcriptional
network models for the understanding of cancer have been simulated, but also the
integration of microarray-derived data has been a useful tool for identifying gene modules
involved in different cancer-altered pathways (Segal et al., 2005). Furthermore, it has been
shown that cancer alterations can be better correlated when these are compared to different
organisms, suggesting that combining data obtained from both cell lines and various
techniques can provide more compelling ideas to understand biological phenomena.

5.1 Systems biology and cervical cancer

All available information from the cancer transcriptome could be easily correlated if the
respective studies would share a universal language e.g. MIAME (Minimal Information
About a Microarray Experiment) (Quackenbush, 2004).

Most microarray reports and in particular those in CC, however, contain no standardized
data. Using a database and different computational tools (Kent, 2002; Wain et al., 2004;
Wheeler et al., 2008) to assign all genes the same nomenclature, it is nonetheless possible to
assess their expression levels and correlate them in different scenarios. For example, from all
the aforementioned CC microarray studies, when assessing only “on”/“off” expression, we
observed genes commonly found between some studies (Table 1).

Many of the genes in Table 1 have been implicated before in CC. Nevertheless, these genes
can be related to other high performance techniques, such as the identification of tumor
suppressor genes among a big set of genes that increase their expression during loss of
tumorigenicity in HeLa cells (Mikheev et al., 2004) or the quantification of transcripts present
in samples of CC (Frigessi et al., 2005) or normal cervix (Perez-Plasencia et al., 2005).
### Up-regulated genes in cervical cancer

<table>
<thead>
<tr>
<th>Gene</th>
<th>References</th>
</tr>
</thead>
<tbody>
<tr>
<td>TOP2A</td>
<td>(Nees et al., 2000), (Nees et al., 2001), (Garner-Hamrick et al., 2004), (Thierry et al., 2004), (Sopov et al., 2004), (Chen et al., 2003), (Rosty et al., 2005), (Santin et al., 2005)</td>
</tr>
<tr>
<td>CCNA2</td>
<td>(Nees et al., 2000), (Nees et al., 2001), (Garner-Hamrick et al., 2004), (Thierry et al., 2004), (Sopov et al., 2004), (Rosty et al., 2005), (Santin et al., 2005)</td>
</tr>
<tr>
<td>CCNB1</td>
<td>(Nees et al., 2000), (Nees et al., 2001), (Garner-Hamrick et al., 2004), (Santin et al., 2005), (Vazquez-Ortiz et al., 2005a), (Liu et al., 2003)</td>
</tr>
<tr>
<td>CDKN2A</td>
<td>(Nees et al., 2001), (Garner-Hamrick et al., 2004), (Wong et al., 2006), (Rosty et al., 2005), (Santin et al., 2005), (Kelley et al., 2005)</td>
</tr>
<tr>
<td>PLK1</td>
<td>(Nees et al., 2000), (Nees et al., 2001), (Pett et al., 2006), (Wells et al., 2003), (Rosty et al., 2005), (Santin et al., 2005)</td>
</tr>
<tr>
<td>BIRC5</td>
<td>(Nees et al., 2000), (Nees et al., 2001), (Garner-Hamrick et al., 2004), (Rosty et al., 2005), (Santin et al., 2005)</td>
</tr>
<tr>
<td>MCM2</td>
<td>(Garner-Hamrick et al., 2004), (Wells et al., 2003), (Wong et al., 2006), (Rosty et al., 2005), (Santin et al., 2005)</td>
</tr>
<tr>
<td>NEK2</td>
<td>(Nees et al., 2001), (Garner-Hamrick et al., 2004), (Thierry et al., 2004), (Rosty et al., 2005), (Santin et al., 2005)</td>
</tr>
<tr>
<td>BUB1</td>
<td>(Nees et al., 2001), (Garner-Hamrick et al., 2004), (Wells et al., 2003), (Rosty et al., 2005)</td>
</tr>
<tr>
<td>CCNB2</td>
<td>(Garner-Hamrick et al., 2004), (Thierry et al., 2004), (Rosty et al., 2005), (Santin et al., 2005)</td>
</tr>
<tr>
<td>CDC2</td>
<td>(Garner-Hamrick et al., 2004), (Wells et al., 2003), (Rosty et al., 2005)</td>
</tr>
<tr>
<td>CDC20</td>
<td>(Nees et al., 2001), (Wells et al., 2003), (Rosty et al., 2005), (Santin et al., 2005)</td>
</tr>
<tr>
<td>CKS1B</td>
<td>(Nees et al., 2001), (Thierry et al., 2004), (Rosty et al., 2005), (Santin et al., 2005)</td>
</tr>
<tr>
<td>E2F1</td>
<td>(Wells et al., 2003), (Rosty et al., 2005), (Santin et al., 2005), (Hudelist et al., 2005)</td>
</tr>
<tr>
<td>FOXM1</td>
<td>(Garner-Hamrick et al., 2004), (Thierry et al., 2004), (Rosty et al., 2005), (Santin et al., 2005)</td>
</tr>
<tr>
<td>KRT18</td>
<td>(Garner-Hamrick et al., 2004), (Thierry et al., 2004), (Sopov et al., 2004), (Rosty et al., 2005)</td>
</tr>
<tr>
<td>MEST</td>
<td>(Duffy et al., 2003), (Chen et al., 2003), (Rosty et al., 2005), (Santin et al., 2005)</td>
</tr>
<tr>
<td>MKI67</td>
<td>(Garner-Hamrick et al., 2004), (Thierry et al., 2004), (Rosty et al., 2005), (Vazquez-Ortiz et al., 2005b)</td>
</tr>
<tr>
<td>MSH6</td>
<td>(Garner-Hamrick et al., 2004), (Sopov et al., 2004), (Rosty et al., 2005), (Santin et al., 2005)</td>
</tr>
</tbody>
</table>

### Down-regulated genes in cervical cancer

<table>
<thead>
<tr>
<th>Gene</th>
<th>References</th>
</tr>
</thead>
<tbody>
<tr>
<td>CDKN1A</td>
<td>(Chang &amp; Laimins, 2000), (Nees et al., 2000), (Nees et al., 2001), (Duffy et al., 2003), (Thierry et al., 2004), (Wells et al., 2003), (Kelley et al., 2005)</td>
</tr>
<tr>
<td>FN1</td>
<td>(Nees et al., 2000), (Nees et al., 2001), (Toussaint-Smith et al., 2004), (Kelley et al., 2005), (Santin et al., 2005)</td>
</tr>
<tr>
<td>TRIM22</td>
<td>(Chang &amp; Laimins, 2000), (Nees et al., 2001), (Duffy et al., 2003), (Pett et al., 2006), (Kelley et al., 2005), (Santin et al., 2005)</td>
</tr>
<tr>
<td>ILIRN</td>
<td>(Chang &amp; Laimins, 2000), (Duffy et al., 2003), (Ruutu et al., 2002), (Wong et al., 2006), (Santin et al., 2005)</td>
</tr>
<tr>
<td>SPRR1A</td>
<td>(Chang &amp; Laimins, 2000), (Duffy et al., 2003), (Alazawi et al., 2002), (Wong et al., 2006), (Santin et al., 2005)</td>
</tr>
<tr>
<td>TNC</td>
<td>(Duffy et al., 2003), (Garner-Hamrick et al., 2004), (Pett et al., 2006), (Kelley et al., 2005), (Santin et al., 2005)</td>
</tr>
<tr>
<td>IGFBP6</td>
<td>(Garner-Hamrick et al., 2004), (Wong et al., 2006), (Hudelist et al., 2005), (Liu et al., 2003)</td>
</tr>
<tr>
<td>LCN2</td>
<td>(Chang &amp; Laimins, 2000), (Nees et al., 2001), (Duffy et al., 2003), (Santin et al., 2005)</td>
</tr>
<tr>
<td>ABCA1</td>
<td>(Garner-Hamrick et al., 2004), (Kelley et al., 2005), (Santin et al., 2005)</td>
</tr>
<tr>
<td>BNIP2</td>
<td>(Nees et al., 2001), (Thierry et al., 2004), (Wong et al., 2006)</td>
</tr>
<tr>
<td>CSPG2</td>
<td>(Duffy et al., 2003), (Ruutu et al., 2002), (Santin et al., 2005)</td>
</tr>
<tr>
<td>DDB2</td>
<td>(Duffy et al., 2003), (Thierry et al., 2004), (Kelley et al., 2005)</td>
</tr>
<tr>
<td>GSN</td>
<td>(Garner-Hamrick et al., 2004), (Thierry et al., 2004), (Kelley et al., 2005)</td>
</tr>
<tr>
<td>INPP5D</td>
<td>(Nees et al., 2001), (Duffy et al., 2003), (Wells et al., 2003)</td>
</tr>
<tr>
<td>IVL</td>
<td>(Duffy et al., 2003), (Garner-Hamrick et al., 2004), (Wong et al., 2006)</td>
</tr>
<tr>
<td>KLK7</td>
<td>(Chang &amp; Laimins, 2000), (Duffy et al., 2003), (Wong et al., 2006)</td>
</tr>
<tr>
<td>KRT4</td>
<td>(Duffy et al., 2003), (Ruutu et al., 2002), (Wong et al., 2006)</td>
</tr>
<tr>
<td>KRT16</td>
<td>(Alazawi et al., 2002), (Ruutu et al., 2002), (Wong et al., 2006)</td>
</tr>
<tr>
<td>LAMA3</td>
<td>(Chang &amp; Laimins, 2000), (Kelley et al., 2005), (Santin et al., 2005)</td>
</tr>
</tbody>
</table>
Up-regulated genes in cervical cancer | Down-regulated genes in cervical cancer
--- | ---
**Gene** | **References** | **Gene** | **References**
MYBL2 | (Thierry et al., 2004), (Chen et al., 2003), (Rosty et al., 2005), (Santin et al., 2005) | SMPG | (Garner-Hamrick et al., 2004), (Wong et al., 2006), (Santin et al., 2005)
PRIM1 | (Nees et al., 2001), (Garner-Hamrick et al., 2004), (Rosty et al., 2005), (Santin et al., 2005) | PI3 | (Duffy et al., 2003), (Alazawi et al., 2002), (Santin et al., 2005)
RRM2 | (Nees et al., 2001), (Thierry et al., 2004), (Wong et al., 2006), (Rosty et al., 2005) | PPP2R5B | (Garner-Hamrick et al., 2004), (Ruutu et al., 2002), (Santin et al., 2005)
SPARC | (Nees et al., 2001), (Duffy et al., 2003), (Chen et al., 2003), (Ahn et al., 2004a) | SERPINB2 | (Chang & Laimins, 2000), (Ruutu et al., 2002), (Santin et al., 2005)
TTK | (Garner-Hamrick et al., 2004), (Wells et al., 2003), (Rosty et al., 2005), (Santin et al., 2005) | SPRR2B | (Duffy et al., 2003), (Wong et al., 2006), (Santin et al., 2005)
VEGF | (Garner-Hamrick et al., 2004), (Toussaint-Smith et al., 2004), (Wong et al., 2006), (Vazquez-Ortiz et al., 2005a) | SULT2B1 | (Chang & Laimins, 2000), (Duffy et al., 2003), (Wong et al., 2006)

Table 1. Genes primarily found to be up- or down-regulated in cervical cancer across different DNA microarray platforms comparing non-pathogenic vs tumor samples and cell lines. The internationally accepted nomenclature for each gene can be found in: [http://www.genenames.org/](http://www.genenames.org/) or [http://cgap.nci.nih.gov/Genes/GeneFinder](http://cgap.nci.nih.gov/Genes/GeneFinder).

Moreover, it is even possible to combine all this information with that derived of techniques of medium- (Nees et al., 1998; Cheng et al., 2002a; Brentani et al., 2003; Ahn et al., 2005; Ranamukhaarachchi et al., 2005; Seo et al., 2005; Sgarlato et al., 2005) and low- (Helliwell, 2001; Keating et al., 2001; Follen et al., 2003; Gray & Herrington, 2004) performance in CC.

In addition, the to-be-integrated information can be further correlated with genes that have been (a) implied as potential markers in several metastatic solid tumors, including some of uterine origin (Ramawamy et al., 2003); (b) associated with cervical cancer and other kind of cancers whose somatic or germline mutations frequently favor the development of neoplasia (Forbes et al., 2006); or (c) proposed as common tumor proliferation markers overexpressed across microarray reports in very diverse tumor tissues (Whitfield et al., 2006). Last but not least, a more comprehensive systematic analysis of CC can be done by correlating gene up-regulation mediated via the transcription factors E2F (Bracken et al., 2004) and TP53 (Wei et al., 2006), being this integration crucial for a general understanding of the transcriptional regulation during CC development because the functions E2F and TP53 are respectively altered by the oncoproteins E7 and E6. The idea of integrating all these additional supporting studies from many sources poses great potential in the diagnosis, prevention, and treatment of cancer as has been shown in liver carcinoma (Thorgerisson et al., 2006).

### 5.2 Systematic model of HPV-mediated cervical carcinogenesis

The invaluable information provided by all the aforementioned microarray-based CC reports can be related to those additional supporting studies through an integrative disease model as the HPV-mediated cervix carcinogenesis develops in a complex multiple-step process (Sherman & Kurman, 1998; Klaes et al., 1999; zur Hausen, 2002; Sherman, 2003; Ahn et al., 2004a; Frazer, 2004; Pett et al., 2006; Snijders et al., 2006). It starts with the HR-HPV infection and episomes formation thereof, followed by the production of virions and/or the integration of the viral genome into the host one that can lead to precancerous and...
cancerous lesions of squamous and/or glandular origin and ultimately to death. In other words, with this model (Fig. 3) it is not only possible to correlate the up/down regulation of

![Fig. 3. Cervix carcinogenesis systematic model. The various nomenclatures employed in the histopathology of cervical cancer are aligned by dashed lines and extended to key cellular and molecular events that occur during the transformation of the epithelium. The solid lines indicate a direct relationship between key events. A key event is the infection of cells in the basal membrane by HR-HPVs. These can turn into episomal bodies, which will be in charge of, on one hand, producing infective virions and, on the other, integrating into the genome of epithelial cells. Upon infection, an average of 2-3 years are necessary to develop cervical intraepithelial neoplasia of low- (CIN1/2) and/or high-grade (CIN3), often characterized by the integration of the viral genome, another key event for the disease progression as this often triggers the deregulation of the oncoproteins E6 and E7, mayor chromosomal alterations and cellular immortalization. The immune system plays a key role during carcinogenesis since the majority of HR-HPV infections (80%) as well as most low-grade lesions (80%) regress. Due to this and the long periods of time between viral infection and the progression to invasive disease, the infection by HR-HPVs is necessary but not sufficient for the development of cervical cancer; in addition, the inactivation of anti-oncoproteins (besides pRb and p53) and activation of oncoproteins are necessary to consequently provoke changes at the (epi-)genomic, transcriptomic and proteomic level. D = Displasia; L- or HSIL = Low- or High-grade squamous intraepithelial lesion.
specific genes upon presence/absence of HR-HPVs episomes or genome integration as well as that of the oncoproteins E6 and/or E7, but also to identify specific carcinogenesis targets.

Depending on the study, however, the gene correlation has to be carefully done, for example, the processes of cell differentiation and senescence (Nees et al., 2000; Wells et al., 2003; Ranamukhaarachchi et al., 2005) have been considered as anti-cell proliferation molecular events (Gandarillas, 2000). Similarly, an indirect correlation could be observed for gene activation mediated by LR-HPVs (Thomas et al., 2001) but not HR-HPVs or E6 and E7 oncogene suppression (Wells et al., 2003; Thierry et al., 2004; Kelley et al., 2005) by E2 (Dowhanick et al., 1995) or RNA interference (Novina & Sharp, 2004). More importantly, nonetheless, as will be discussed in the coming subsections, this model allows the comparison of candidate tumor markers to data obtained from other CC studies at the genomic (Lazo, 1999; Wilting et al., 2006), transcriptomic (Martin et al., 2006), proteomic (Bae et al., 2005; Choi et al., 2005; Yim & Park, 2006) and epigenomic (Duenas-Gonzalez et al., 2005; Sova et al., 2006) level.

5.3 Up-regulated candidate tumor markers

Although many genes frequently activated in CC have been reported using microarrays, other techniques and analyses strongly suggest that these are tumor markers. This can be illustrated with the inhibitor of cyclin-dependent kinases (CDKs) p16\textsuperscript{INK4a} or CDKN2A, which is involved in cell cycle and has been categorized as a tumor marker in the development of CC (Keating et al., 2001). Overexpression of p16\textsuperscript{INK4a} at both the transcript and protein level can be detected in samples of cervical dysplasia, squamous and glandular HR-HPV positive and negative lesions when compared with normal cervix by low-throughput techniques (Martin et al., 2006). As summarized in Table 2, p16\textsuperscript{INK4a} up-regulation has been also found using medium- (Brentani et al., 2003) and high-performance methods when the oncoprotein E7 is expressed in cell lines in vitro (Nees et al., 2001; Garner-Hamrick et al., 2004), in patient samples in vivo (Rosty et al., 2005) and when comparing tumors vs normal tissue (Hudelist et al., 2005; Rosty et al., 2005; Santin et al., 2005; Wong et al., 2006).

Interestingly, p16\textsuperscript{INK4a} is one of the genes that can display somatic mutations in CC; an abnormal status that has been linked to the development of cervical squamous cell cancer (SCC) (Forbes et al., 2006). Dozens of references in the literature demonstrate that the overexpression of p16\textsuperscript{INK4a} is useful as a CC tumor marker; however, using patient samples, others have determined transcript inactivation due to strong hypermethylation on its promoter region (Duenas-Gonzalez et al., 2005). Although these findings are contradictory at first glance, some subpopulations of dysplastic cervical cells can also display epigenetic silencing of p16\textsuperscript{INK4a} and associated low protein levels (Nuovo et al., 1999). This suggests that the expression of p16\textsuperscript{INK4a} is inhibited in some cells within the tumor, whereas its overexpression can be abundant in other cells, most probably expressing the oncoprotein E7. In spite of this, the detection of p16\textsuperscript{INK4a} is very useful in the cytological diagnosis of CC and, furthermore, recent evidence suggests that the determination of the p16\textsuperscript{INK4a} protein may be even more useful than the already-established HR-HPVs detection in the cytological diagnosis (Nieh et al., 2005).

Another important up-regulated gene is “survivin” or BIRC5 (Table 2). Although surviving expression is undetectable in normal adult tissues, its expression can be detected normally
<table>
<thead>
<tr>
<th>Biological Process</th>
<th>Gene (Locus)</th>
<th>Throughput&lt;sup&gt;B&lt;/sup&gt;</th>
<th>Marker&lt;sup&gt;C&lt;/sup&gt;</th>
<th>TF&lt;sup&gt;D&lt;/sup&gt;</th>
<th>Analysis&lt;sup&gt;E&lt;/sup&gt;</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>High-</td>
<td>Medium-</td>
<td>Low-</td>
<td>Metastasis</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>Cancer</td>
<td>Progesterone</td>
<td>E2F</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>TP53</td>
<td>Genome</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>Transcriptome</td>
<td>Proteome</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>Epigenome</td>
<td></td>
</tr>
<tr>
<td>Apoptosis</td>
<td>BIRC5 (17q25)</td>
<td>(Nees et al., 2000), (Nees et al., 2001), (Garner-Hamrick et al., 2004), (Rosty et al., 2005), (Santin et al., 2005)</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td></td>
<td>SMC4L1 (3q26.1)</td>
<td>(Rosty et al., 2005), (Santin et al., 2005)</td>
<td>(Brentani et al., 2003)</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td></td>
<td>ZWINT (10q21-22)</td>
<td>(Thierry et al., 2004), (Rosty et al., 2005), (Santin et al., 2005)</td>
<td>(Brentani et al., 2003)</td>
<td>(Sgarlato et al., 2005)</td>
<td>(Willing et al., 2006)</td>
</tr>
<tr>
<td></td>
<td>CKS1B (1q21.2)</td>
<td>(Nees et al., 2001), (Thierry et al., 2004), (Wells et al., 2003), (Rosty et al., 2005), (Santin et al., 2005)</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td></td>
<td>MAD2L1 (4q27)</td>
<td>(Nees et al., 2000), (Thierry et al., 2004), (Wells et al., 2003), (Rosty et al., 2005), (Santin et al., 2005)</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td></td>
<td>MSH6 (2p16)</td>
<td>(Garner-Hamrick et al., 2004), (Sopov et al., 2004), (Rosty et al., 2005), (Santin et al., 2005)</td>
<td>(Ranamukhaar achchi et al., 2005)</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td></td>
<td>CCNA2 (4q25-31)</td>
<td>(Nees et al., 2000), (Nees et al., 2001), (Garner-Hamrick et al., 2004), (Thierry et al., 2004), (Wells et al., 2003), (Rosty et al., 2005), (Santin et al., 2005)</td>
<td>(Brentani et al., 2003)</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>Cell Cycle</td>
<td>PLK1 (16p12.1)</td>
<td>(Nees et al., 2000), (Nees et al., 2001), (Pett et al., 2006), (Wells et al., 2003), (Rosty et al., 2005), (Santin et al., 2005)</td>
<td>(Brentani et al., 2003)</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td></td>
<td>CCNB1 (5q12)</td>
<td>(Nees et al., 2000), (Nees et al., 2001), (Garner-Hamrick et al., 2004), (Rosty et al., 2005), (Santin et al., 2005)</td>
<td>(Brentani et al., 2003)</td>
<td>(Verkhusha et al., 2006)</td>
<td>-</td>
</tr>
<tr>
<td></td>
<td>CDKN2A (p16INK4a) (9p21)</td>
<td>(Garner-Hamrick et al., 2004), (Wong et al., 2006), (Rosty et al., 2005), (Santin et al., 2005), (Hudelist et al., 2005)</td>
<td>(Brentani et al., 2003)</td>
<td>(Ranamukhaar achchi et al., 2005)</td>
<td>(Willing et al., 2006)</td>
</tr>
<tr>
<td></td>
<td>MKI67 (Ag Ki-67) (10q25-ter)</td>
<td>(Garner-Hamrick et al., 2004), (Thierry et al., 2004), (Rosty et al., 2005), (Vazquez-Ortiz et al., 2005b)</td>
<td>(Brentani et al., 2003)</td>
<td>(Koelzer et al., 2003)</td>
<td>-</td>
</tr>
</tbody>
</table>
Table 2. Genes frequently reported as up-regulated in cervical cancer (CC). A) For each biological process, genes are listed in descending order by the mayor number of reports related to CC e.g. the gene MKI67 has been reported at least 190 times in CC. Genes in bold have been used as therapeutic targets in cancer, whereas genes in italics are not so known in CC. The chromosomal localization of the gene is shown in brackets. B) Techniques of high-throughput are DNA microarrays; medium- DD, RDA and ESTs; and low- are tumor markers previously defined in CC. C) Genes proposed as metastasis markers (in solid tumors), tumoral cancer markers (due to frequent mutations) and proliferation markers (large number of cancers). D) Transcription factor (TF) that might regulate the corresponding gene. E) The analysis of the genome refers to the most common chromosomal gains in CC (1q, 3q, 5p, 8q, 20q and Xq); transcriptome to the importance of genes in CC; proteome to overexpressed proteins in CC and; epigenome to genes whose promoter has been found methylated in samples derived from CC. For gene nomenclature see Table 1.

| MYBL2 (20q13.1) | (Thierry et al., 2004), (Chen et al., 2003), (Rosty et al., 2005), (Santin et al., 2005) | (Brentani et al., 2003) | (Sgarlato et al., 2005) | - | - | (Bracken et al., 2004) | - | (Witting et al., 2006) | - | (Martin et al., 2006) | - |
| LMNB1 (5q23.3-31) | (Garner-Hamrick et al., 2004), (Rosty et al., 2005), (Santin et al., 2005) | - | - | (Santin et al., 2005) | - | - | (Bracken et al., 2004) | - | - | (Martin et al., 2006) | - |
| TOP2A (17q21-22) | (Nees et al., 2000), (Nees et al., 2001), (Garner-Hamrick et al., 2004), (Thierry et al., 2004), (Sopov et al., 2004), (Chen et al., 2003), (Rosty et al., 2005), (Santin et al., 2005) | (Brentani et al., 2003) | - | - | (Whitting et al., 2006) | - | (Martin et al., 2006) | - |
| MCM2 (3q21) | (Garner-Hamrick et al., 2004), (Wells et al., 2003), (Wong et al., 2006), (Rosty et al., 2005), (Santin et al., 2005) | (Brentani et al., 2003) | (Sgarlato et al., 2005) | - | - | - | (Bracken et al., 2004) | - | - | (Martin et al., 2006) | - |
| MCM4 (8q11.2) | (Ruuutu et al., 2002), (Chen et al., 2003), (Santin et al., 2005) | - | - | (Whitting et al., 2006) | - | - | - | (Martin et al., 2006) | - |
| KRT19 (17q21-23) | (Garner-Hamrick et al., 2004), (Alazawi et al., 2002), (Wong et al., 2006), 113 | (Brentani et al., 2003) | - | - | - | - | - | (Whitting et al., 2006) | - | - | - |
| KRT18 (12q13) | (Garner-Hamrick et al., 2004), (Thierry et al., 2004), (Sopov et al., 2004), (Rosty et al., 2005) | (Brentani et al., 2003) | - | - | - | - | - | - | (Martin et al., 2006) | - | - |
| VEGF (6p21-12) | (Garner-Hamrick et al., 2004), (Toussaint-Smith et al., 2004), (Wong et al., 2006), (Vazquez-Ortiz et al., 2005a) | - | - | (Helliwell, 2001) | - | - | - | (Martin et al., 2006) | - | - | - |
| VEGFC (4q33-34) | (Nees et al., 2001), (Duffy et al., 2003), (Pett et al., 2006) | - | - | - | - | - | - | - | - | - | - |

in embryogenesis and abnormally in cancer (mainly inhibiting apoptosis). Due to this, survivin has been generally proposed as a proliferation tumor marker in cancer (Whitfield et al., 2006).
and particularly in CC (Branca et al., 2005), opening promising therapeutic strategies (Altieri, 2006). Other emergent useful target genes are the members 2 and 4 from the "minichromosome maintenance deficient" complex or MCM found, genes primary involved in DNA replication that have been considered useful for cancer diagnosis and therapy (Rosty et al., 2005; Santin et al., 2005). Similarly, other up-regulated genes that could be specifically targeted are the 2α topoisomerase or TOP2A (Whittfield et al., 2006), cyclin B1 or CCNB1 (Yuan et al., 2004), the kinase 1 polo type or PLK1 (Strebhardt & Ullrich, 2006) and keratin 19 or KRT19 (Chang et al., 2005). The transcripts of the latter gene have been abundantly estimated not only in CC-derived samples (Frigessi et al., 2005), but also determined as overexpressed at both the messenger (Alazawi et al., 2002; Brentani et al., 2003; Garner-Hamrick et al., 2004; Wong et al., 2006), and protein (Bae et al., 2005) level in cervical neoplasia compared to normal tissue. As KRT19, a protein part of the intermediate filaments of epithelial cells, KRT18 (Table 2) could likewise play an important role in the molecular diagnosis of cancer.

It should be noted that several genes reported in Table 2 only have been linked to CC using high and average performance techniques, such as the gene involved in the structural maintenance of chromosomes "SMC4L1". As far as we known, a single report correlated the expression levels of this gene to esophageal squamous cancer (Yen et al., 2005), but a genomic analysis showed that chromosomal gains in the region 3q12.1-28 (where SMC4L1 lies) are most common in SCC (Wilting et al., 2006). This gene might be activated by E2F (Bracken et al., 2004), but it is desirable to check the expression levels of SMC4L1 with low-yield techniques to determine its relevance in CC as well as for potential metastatic markers like LMNB1 or proliferation ones like MAD2L1 gene (Table 2).

5.4 Down-regulated candidate tumor markers

Using microarrays and other techniques it has been possible to find genes frequently down-regulated in CC, suggesting that these may play a role as tumor markers e.g. the tumor suppressor gene \( p21^{\text{WAF1/CIP1}} \) or CDKN1A, which regulates the cell cycle via CDKs inhibition, senescence as well as TP53-dependent and -independent apoptosis (Table 3). Upon degradation or inactivation of the nuclear phosphoprotein TP53 by E6 or PLK1, respectively, the transcription of \( p21^{\text{WAF1/CIP1}} \) is reduced as observed in several types of cancer (Gartel & Radhakrishnan, 2005) and particularly in CC samples using DNA microarrays (Chang & Laimins, 2000; Nees et al., 2000; Nees et al., 2001; Duffy et al., 2003; Wells et al., 2003; Thierry et al., 2004; Kelley et al., 2005). In addition, it has been suggested that low \( p21^{\text{WAF1/CIP1}} \) expression correlates with poor prognosis in cervical adenocarcinoma (AC) (Lu et al., 1998). Moreover, in samples derived from CC it has been observed a decrease in cell growth and induction of \( p21^{\text{WAF1/CIP1}} \) by platinum-based chemotherapy (Gatti et al., 2004) as well as radioimmunotherapy directed against KRT19 (Chang et al., 2005).

Other down-regulated genes in CC include the gene desmoglein 1 or DSG1, which encodes a protein involved in the homeostasis of cell-cell epithelial junctions and belongs to the family of "cadherins", proteins whose expression decreases as it progresses in many kinds of cancers, such as cervical cancer (de Boer et al., 1999). It has been determined that the expression of DSG1 increases in presence of LR-HPVs episomal bodies in human keratinocytes, but its expression levels highly decrease when HR-HPV episomes are present in cell lines and SCC samples. Moreover, DSG1 importantly lies in an area that often presents chromosomal losses during CC (Table 3) and has been assigned as a pro-apoptotic factor mediated by the caspase 3 in keratinocytes (Dusek et al., 2006).
<table>
<thead>
<tr>
<th>Biological Process</th>
<th>Gene (Locus)</th>
<th>References</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Cell Cycle</strong></td>
<td>CDKN1A (p21WAF1/CIP1, 6p21.1)</td>
<td>(Chang &amp; Laimins, 2000), (Nees et al., 2000), (Nees et al., 2001), (Duffy et al., 2003), (Thierry et al., 2004), (Wells et al., 2003), (Kelley et al., 2005)</td>
</tr>
<tr>
<td></td>
<td>FN1 (2q34-36)</td>
<td>(Nees et al., 2000), (Nees et al., 2001), (Toussaint-Smith et al., 2004), (Kelley et al., 2005), (Santin et al., 2005), (Hudelist et al., 2005)</td>
</tr>
<tr>
<td></td>
<td>DSG1 (18q12.1)</td>
<td>(Chang &amp; Laimins, 2000), (Thomas et al., 2001), (Wong et al., 2006)</td>
</tr>
<tr>
<td></td>
<td>CSPG2 (5q12-14)</td>
<td>(Duffy et al., 2003), (Ruutu et al., 2002), (Santin et al., 2005)</td>
</tr>
<tr>
<td></td>
<td>SERPINE2 (18q21.3)</td>
<td>(Chang &amp; Laimins, 2000), (Ruutu et al., 2002), (Santin et al., 2005)</td>
</tr>
<tr>
<td></td>
<td>BNIP2 (10q26.3)</td>
<td>(Nees et al., 2001), (Thierry et al., 2004), (Wong et al., 2006)</td>
</tr>
<tr>
<td></td>
<td>IL1RN (2q14.2)</td>
<td>(Chang &amp; Laimins, 2000), (Duffy et al., 2003), (Ruutu et al., 2002), (Wong et al., 2006), (Santin et al., 2005)</td>
</tr>
<tr>
<td></td>
<td>TRIM22 (11p15)</td>
<td>(Chang &amp; Laimins, 2000), (Nees et al., 2001), (Duffy et al., 2003), (Pett et al., 2006), (Kelley et al., 2005), (Santin et al., 2005)</td>
</tr>
<tr>
<td></td>
<td>KLK7 (19q13.41)</td>
<td>(Chang &amp; Laimins, 2000), (Duffy et al., 2003), (Wong et al., 2006)</td>
</tr>
<tr>
<td></td>
<td>KRT4 (12p12-11)</td>
<td>(Duffy et al., 2003), (Ruutu et al., 2002), (Wong et al., 2006)</td>
</tr>
<tr>
<td></td>
<td>KRT16 (17q12-21)</td>
<td>(Alazawi et al., 2002), (Ruutu et al., 2002), (Wong et al., 2006)</td>
</tr>
<tr>
<td></td>
<td>LAMA3 (18q11.2)</td>
<td>(Chang &amp; Laimins, 2000), (Kelley et al., 2005), (Santin et al., 2005)</td>
</tr>
<tr>
<td></td>
<td>SPRR3 (1q21-22)</td>
<td>(Wong et al., 2006), (Santin et al., 2005), (Perez-Plasencia et al., 2005)</td>
</tr>
</tbody>
</table>

**Biological Process**

- **Cell Cycle**
- **Cell Adhesion**
- **Apoptosis**
- **Immune Response**
- **Epidermal Development**

**Gene (Locus)**

- CDKN1A (p21WAF1/CIP1, 6p21.1)
- FN1 (2q34-36)
- DSG1 (18q12.1)
- CSPG2 (5q12-14)
- SERPINE2 (18q21.3)
- BNIP2 (10q26.3)
- IL1RN (2q14.2)
- TRIM22 (11p15)
- KLK7 (19q13.41)
- KRT4 (12p12-11)
- KRT16 (17q12-21)
- LAMA3 (18q11.2)
- SPRR3 (1q21-22)

**References**

- (Chang & Laimins, 2000)
- (Nees et al., 2000)
- (Nees et al., 2001)
- (Toussaint-Smith et al., 2004)
- (Kelley et al., 2005)
- (Santin et al., 2005)
- (Hudelist et al., 2005)
- (Duffy et al., 2003)
- (Ruutu et al., 2002)
- (Santin et al., 2005)
- (Chang & Laimins, 2000)
- (Thomas et al., 2001)
- (Wong et al., 2006)
- (Brentani et al., 2003)
- (Wilting et al., 2006)
- (Wilting et al., 2006)
- (Wilting et al., 2006)
- (Lazo, 1999)
- (Nees et al., 2000)
- (Nees et al., 2000)
- (Wong et al., 2006)
- (Brentani et al., 2003)
- (Wilting et al., 2006)
- (Nees et al., 2000)
- (Lazo, 1999)
- (Contag et al., 2004)
- (Wilting et al., 2006)
- (Wilting et al., 2006)
- (Perez-Plasencia et al., 2005)
Table 3. Genes frequently reported as down-regulated in cervical cancer (CC). A) For each biological process, genes are listed in descending order by the mayor number of reports related to CC e.g. the gene CDKN1A has been reported at least 70 times in CC. Genes in bold represent increased expression levels upon different schemes of radio and/or chemotherapy, whereas genes in italics are not so known in CC. The chromosomal localization of the gene is shown in brackets. B) Techniques of high-throughput are mainly DNA microarrays; medium- DD, and ESTs; and low- are tumor markers previously defined in CC. C) Genes proposed as metastasis markers (in solid tumors) and tumoral cancer markers (due to frequent mutations). D) Transcription factor (TF) that might regulate the corresponding gene. E) The analysis of the genome refers to the most common chromosomal alterations in CC (2q, 3p, 4p, 5p, 5q, 6p, 6q, 11q, 13q, 18q and 19q). For gene nomenclature see Table 1.

Another gene that could be of interest in CC is SERPINB2. The gene product is an inhibitor of the serine-type proteases like the plasminogen activator (also known as PLAU). On one hand, SERPINB2 suppression has been determined using both microarrays as well as genomic studies in CC (Table 3); but on the other, its expression in HeLa cells can stabilize the expression levels of the Rb protein and suppress the oncoproteins E6 and E7 of HPV18 (Darnell et al., 2005). This suggests that low levels of SERPINB2 promote CC development, being this a potentially good molecular marker.

Of genes not known in CC there are several examples, being the gene TRIM22 or “tripartite motif-containing 22”, which has been found down-regulated in at least 6
microarray studies (Table 3). TRIM22 belongs to a conserved family of antiviral proteins, where the member 22 has been implicated in inhibiting the replication of the human immunodeficiency virus 1 (HIV1) (Nisole et al., 2005). This suggests that TRIM22 may be relevant in the immune response HR-HPVs and that these viruses may be responsible for its inhibition.

Table 3 also lists genes from the epidermal differentiation complex (EDC, located in the band 21 of the long arm of chromosome 1), for instance, using SAGE, abundant transcripts of SPRR3 have been found in normal cervical tissue, but a low SPRR3 expression has been determined in tumor tissue using microarrays (Table 3). This suggests that SPRR3 and perhaps SPRR1A, which also belongs to the EDC, may be useful tumor markers in CC. Last, other suppressed genes in CC are IGFBP6 and RUNX1 (Table 3). While the first one is responsible for inactivating a potent growth factor similar to insulin (IGF2), a gene in turn required by IGFBP6 to reduce metastatic characteristics in tumors from different origin (Bach, 2005), the second gene belongs to a family of transcription factors that can inhibit angiogenesis (Sakakura et al., 2005).

5.5 Candidate tumor markers in cervical cancer subtypes

Although HPV-16-infections are more frequently detected than HPV-18 ones in squamous cell carcinoma (SCC), the latter ones are more often associated to adenocarcinoma of the cervix (AC), whose incidence is growing at the same time as SCC incidence. Interestingly, several genes with a clinically usefulness for the molecular differentiation between the two major histological subtypes of CC have been found using DNA microarrays (Table 4). The genes TACSTD1 and CEACAM5, which encode transmembrane proteins that transmit signals for development, motility and cell growth, for example, were found to be upregulated in AC compared to SCC (Chao et al., 2006).

<table>
<thead>
<tr>
<th>Up-regulated genes in squamous cell carcinoma</th>
<th>Up-regulated genes in adenocarcinoma</th>
</tr>
</thead>
<tbody>
<tr>
<td>Gene</td>
<td>Reference</td>
</tr>
<tr>
<td>CRABP2</td>
<td>(Chao et al., 2006)</td>
</tr>
<tr>
<td>NDRG1</td>
<td>(Chao et al., 2006)</td>
</tr>
<tr>
<td>CDH13</td>
<td>(Fujimoto et al., 2004)</td>
</tr>
<tr>
<td>KRT13</td>
<td>(Chao et al., 2006)</td>
</tr>
<tr>
<td>KRT15</td>
<td>(Chao et al., 2006)</td>
</tr>
<tr>
<td>PTHHLH</td>
<td>(Fujimoto et al., 2004)</td>
</tr>
<tr>
<td>S100A9</td>
<td>(Chao et al., 2006)</td>
</tr>
<tr>
<td>SPRR1B</td>
<td>(Chao et al., 2006)</td>
</tr>
</tbody>
</table>

Table 4. Genes with a possible clinical utility for the molecular differentiation between squamous cell carcinoma (SCC) and adenocarcinoma (AC) in cervical cancer. The internationally accepted nomenclature for each gene can be found in: http://www.genenames.org/ or http://cgap.nci.nih.gov/Genes/GeneFinder.
Furthermore, high levels of the corresponding proteins served by themselves as poor prognostic factors in patients with AC compared with SCC (Chao et al., 2006). Other genes for potential use as markers in CC that have been found with microarrays are:

1. **CRABP2** (belongs to the EDC and encodes the retinoic acid binding protein 2) has been identified as up-regulated in SCC compared to normal tissue (Shim et al., 1998; Seo et al., 2005) and AC (Chao et al., 2006).

2. **NDRG1** (N-myc Downstream Regulated Gene 1) is involved in cell growth and differentiation and was found overexpressed in SCC compared to AC (Chao et al., 2006) and normal cervical tissue (Sgarlato et al., 2005).

3. Other members of the “Carcinoembryonic antigen-related cell adhesion molecule” family such as the **CEACAM-1**, -5, -6, and -7, are shown as up-regulated in AC compared to SCC (Fujimoto et al., 2004; Chao et al., 2006).

4. **MSLN** or mesothelin encodes a membrane glycoprotein involved in cell adhesion whose transcripts are detectable in normal tissue but abundant in tumors of glandular origin or HeLa cells. In CC, MSLN is overexpressed in AC compared to SCC (Chao et al., 2006) and in HPV-18-derived samples of SCC/AC compared to normal tissue (Rosty et al., 2005). It is worth noting that MSLN is a therapeutic target in various malignancies (Hassan et al., 2004).

5. Finally, high expression levels of **FOLR1** (folate receptor) have been associated with an AC phenotype (Fujimoto et al., 2004) and tumorigenicity in cell lines derived from AC (Mikheev et al., 2004). However, further studies are required to demonstrate the relevance of this receptor in both AC and SCC because it is known that via FOLR1 and folic acid, its ligand, some drugs can be bound and directed into over-expressing high levels of FOLR1, as suggested in several types of cancer (Kelemen, 2006).

The aforementioned “tumoral markers” could be potentially important for the diagnosis, prevention, and treatment of CC because these were identified using cell lines from various sources as well as samples of SCC and/or AC for comparative studies with normal tissues. Last but not least, a recent and interesting CC review not only proposed a similar systematic model of HPV infection highlighting the current debate on the viral status as hallmark of disease progression (episomal vs integrated forms where HPV-18 genome integration seems to prevail in women with advance disease in contrast to HPV-16), but also provided overlapping and additional tumor markers at some of those analyzed herein (Woodman et al., 2007). Along these lines, it would be worth saying that cancer, including its hundred subtypes, is such a complex phenomenon (Vogelstein & Kinzler, 2004), which should be rather seen as an average of key molecular events displaying often specific hallmarks (Hanahan & Weinberg, 2000) of disease progression.

### 6. Conclusions

Thanks to the comparison of the cervix in normal and abnormal conditions via transcriptomics in general and particularly using DNA microarrays, it is possible to identify known and unknown clinically relevant genes for the disease progression. The next goal is to identify and validate specific tumor markers for profiling histo- and pathological subtypes. This will allow not only a molecular subclassification and more understanding of CC, but also choosing the right treatment for each patient according to its gene expression signature if there is prior knowledge about the most likely response.
she would have. This is the only way to fully understand more about this complex disease.

The intention of this manuscript is to provide the reader a broad view of the transcriptome, an area that is developing rapidly, especially in cancer. It is worthwhile reemphasize that the transcriptome also consists of non-coding RNAs regulating the transcription of many genes and likewise acting as oncogenes or tumor suppressor genes. With such complexity, the best tackling to cancer will rely on predictive hypothesis, so it is important to take into account systems biology, which will allow us to better understand transcriptional networks and identify specific therapy targets for a tailored therapy.

Although there are improved programs for the early diagnosis of CC as well as very effective prophylactic vaccines against HR-HPVs, the high mortality rates triggered by CC will not diminish soon, not even in the medium-term after optimizing CC monitoring programs and broadly executing vaccination schemes. An alternative for CC patients is therefore to look at those tumor markers that could aid in the stratification of the disease and therapy. Unfortunately, genomics and all its derivatives are exacerbating global inequalities in terms of scientific research and health between developed and developing countries since the first cause of death of women in the former countries is breast cancer whereas in the latter ones CC kills every 2 hours, on average, a Mexican woman in productive age.

7. Acknowledgments

We thank the library at the MPI of Coal Research for financial support.

8. References


line, CaSki cells, through apoptosis, G(1) arrest, and regulation of gene expression. DNA and cell biology 22, 217-224.


Bach LA (2005) IGFBP-6 five years on; not so ‘forgotten’? Growth hormone & IGF research 15, 185-192.


Gatherer D (2010) So what do we really mean when we say that systems biology is holistic? *BMC systems biology* 4, 22.


www.intechopen.com


Harrison PR, Conkie D, Paul J & Jones K (1973) Localisation of cellular globin messenger RNA by in situ hybridisation to complementary DNA. *FEBS letters* 52, 109-112.


Nieh S, Chen SF, Chu TY, Lai HC, Lin YS, Fu E & Gau CH (2005) Is p16(INK4A) expression more useful than human papillomavirus test to determine the outcome of atypical


Sgarlato GD, Eastman CL & Sussman HH (2005) Panel of genes transcriptionally up-regulated in squamous cell carcinoma of the cervix identified by representational difference analysis, confirmed by macroarray, and validated by real-time quantitative reverse transcription-PCR. *Clinical chemistry* 51, 27-34.


A Transcriptome- and Marker-Based Systemic Analysis of Cervical Cancer


Cervical Cancer is one of the leading cancers among women, especially in developing countries. Prevention and control are the most important public health strategies. Empowerment of women, education, "earlier" screening by affordable technologies like visual inspection, and treatment of precancers by cryotherapy/LEEP are the most promising interventions to reduce the burden of cervical cancer. Dr. Rajamanickam Rajkumar had the privilege of establishing a rural population based cancer registry in South India in 1996, as well as planning and implementing a large scale screening program for cervical cancer in 2000. The program was able to show a reduction in the incidence rate of cervical cancer by 25%, and reduction in mortality rate by 35%. This was the greatest inspiration for him to work on cervical cancer prevention, and he edited this book to inspire others to initiate such programs in developing countries. InTech - Open Access Publisher plays a major role in this crusade against cancer, and the authors have contributed to it very well.

How to reference
In order to correctly reference this scholarly work, feel free to copy and paste the following: