Overview of Microarray Analysis of Gene Expression and its Applications to Cervical Cancer Investigation

Angel Chao[1]*, Tzu-Hao Wang[1,2], Chyong-Huey Lai[1]

[1]Department of Obstetrics and Gynecology, Lin-Kou Medical Center, Chang Gung Memorial Hospital, Chang Gung University, and [2]Genomic Medicine Research Core Laboratory, Chang Gung Memorial Hospital, Tao-Yuan, Taiwan.

SUMMARY

Cervical cancer is one of the leading female cancers in Taiwan and ranks as the fifth cause of cancer death in the female population. Human papillomavirus has been established as the causative agent for cervical neoplasia and cervical cancer. However, the tumor biology involved in the prognoses of different cell types in early cancers and tumor responses to radiation in advanced cancers remain largely unknown. The introduction of microarray technologies in the 1990s has provided genome-wide strategies for searching tens of thousands of genes simultaneously. In this review, we first summarize the two types of microarrays: oligonucleotides microarray and cDNA microarray. Then, we review the studies of functional genomics in cervical cancer. Gene expression studies that involved cervical cancer cell lines, cervical cells of cancer versus normal ectocervix, cancer tissues of different histology, radioresistant versus radiosensitive patients, and the combinatorial gene expression associated with chromosomal amplifications are discussed. In particular, *CEACAM5*, *TACSTD1*, *S100P*, and *MSLN* have shown to be upregulated in adenocarcinoma, and increased expression levels of *CEACAM5* and *TACSTD1* were significantly correlated with poorer patient outcomes. On the other hand, 35 genes, including apoptotic genes (e.g. *BIK*, *TEGT*, *SSI-3*), hypoxia-inducible genes (e.g. *HIF1A*, *CA12*), and tumor cell invasion and metastasis genes (e.g. *CTSL*, *CTSB*, *PLAU*, *CD44*), have been noted to echo the hypothesis that increased tumor hypoxia leads to radiation resistance in cervical cancer during radiation. [*Taiwan J Obstet Gynecol* 2007;46(4):363–373]

Key Words: cervix, gene expression, microarray, neoplasm

Introduction

Cervical cancer continues to be one of the leading female genital cancers worldwide [1]. The cervical cancer burden in Taiwan is still heavy with the incidence of invasive cancer being 24.5/100,000, ranking it the fifth leading cause of cancer death for women [2]. In addition to the causal link of human papillomavirus to cervical cancer [3–7], what other cofactors exist in the process of carcinogenesis and contribute to the resulting different cell types? Additionally, why do patients with adenocarcinoma of the cervix have poor outcomes? Genetic insults, such as chromosome abnormalities, altered expression of growth factors or their receptors, disturbances in tumor suppressor gene function or amplification of cellular oncogenes, may lead to cervical carcinogenesis [8].

Primary radiation therapy is the treatment modality of choice in early stage cervical cancer, depending mainly on the availability of appropriate expertise [8]. In cases of locally advanced cervical cancer, concurrent chemoradiation has emerged as the recommended treatment [9]. Radiation provides humans with lifesaving technologies; but, at the same time, it causes DNA damage that initiates a plethora of overlapping physiologic

*Correspondence to: Dr Angel Chao, Department of Obstetrics and Gynecology, Chang Gung Memorial Hospital, Chang Gung University, Tao-Yuan 333, Taiwan.
E-mail: angel945@adm.cgmh.org.tw
Accepted: September 19, 2007

responses responsible for maintaining cellular homeostasis and promoting interactions with adjacent cells [10]. These responses include induction of cell cycle checkpoints, senescence, and apoptosis. Therefore, the cellular responses of patients who undergo radiation alone or a combined radiation and chemotherapy treatment (the dual-modality treatment) warrant further investigation.

Functional genomics focuses on genome-wide patterns of gene expression and the mechanisms by which gene expression are coordinated. The use of microarrays has spread rapidly throughout the research community since the 1990s when it was first introduced [11–13]. The completion of the human genome project as well as the current availability of high-throughput technologies have enabled us to move from the study of single molecules to the simultaneous analysis of 24,000 human genes in the complex biological systems of human life.

Cancer, a complex and prevalent disease, has become a prime target for the application of novel technologies of genomics, proteomics, and functional genomics, since intensive research may help in the development of a predictive, individualized approach to cancer care and can facilitate the selection of the best treatment method for each individual patient [14]. Microarrays give researchers the opportunity to analyze tens of thousands of genes on one assay, allowing scientists to perform experiments at a previously unimaginable scale. In this review article, we first provide a summary of oligonucleotides microarray and cDNA microarray, experimental designs, gene expression analysis, and verification of microarray data. Next, we review the studies of functional genomics in cervical cancer.

Overview of Microarray Analysis of Gene Expression (MAGE)

A microarray or DNA chip is an orderly arrangement of DNA molecules that have been chemically bonded to a fine grid of surfaces [15]. The purpose of MAGE is to analyze the expression states of genes in complementary DNA (cDNA) prepared from mRNA in which the hybridization takes place on the array via the Watson–Crick duplex formation method. A "probe" is the tethered nucleic acid of known sequence on the fine grid surface, whereas a "target" is the free nucleic acid in the biological samples to be determined [16]. The power of a microarray is its large-scale ability to analyze the expression of thousands of genes simultaneously.

The two basic types of DNA microarrays are oligonucleotide microarrays and cDNA microarrays, although variations are being developed. For the majority of oligonucleotide microarrays, the oligonucleotides are directly synthesized and arrayed in a discrete regular grid on silicon or glass chip surfaces. For some oligonucleotide microarrays, the oligonucleotides that are first synthesized are spotted onto the slides. cDNA microarrays are previously prepared DNA clones (usually a polymerase chain reaction [PCR] product of longer than 100 nucleotides) which are immobilized to a glass slide.

In the late 1980s, Stephen Fodor and his colleagues adopted the process of photolithography to make microarrays with chemically synthesized oligonucleotides [11,12]. Affymetrix microarrays contain between 11 and 20 pairs of oligonucleotide probes for each target RNA, in which one of the pair is the reverse complement to a unique 25-mer in the RNA and the other contains a mutated middle base pair and serves as a measure of stray signal (http://www.affymetrix.com/index.affx). The chip is arrayed with oligonucleotides that are individually synthesized directly on the chip by automatic procedures until four nucleotides occupy the place. This cycle of reproduction and addition of a nucleotide is repeated until the array carries oligonucleotides that are 20–25 nucleotides in length. In recent years, investigators have developed additional methods for constructing microarrays by synthesizing oligonucleotides *in situ*. In one approach, Agilent Technologies (http://www.chem.agilent.com) uses ink-jet printer to prepare nucleotides for each DNA synthesis step on a glass slide. In a different approach, NimbleGen Systems (http://www.nimblegen.com) uses "maskless synthesis" photolithography that employs digital light processing to generate ultra-high density microarrays on slides, with longer oligonucleotides of 50–70 nucleotides.

In the mid-1990s, Pat Brown and his colleagues developed the "dot blot" microarray. The array is made in three steps: (1) preparing the plasmid cDNA clones to be put on the array, using PCRs to amplify the cDNA inserts; large sets of these clones are available from the expressed sequence tag and cDNA sequencing is begun early in the human genome project; (2) spotting the DNA onto the glass surface of the array with a spotting robot; and (3) postspotting processing of the glass slide [13]. During gene expression analysis with cDNA microarrays, four steps are involved: (1) sample preparation and labeling, (2) hybridization, (3) washing, and (4) image acquisition. With these two-color arrays, two different sets of mRNA samples are used, each labeled separately with different fluorescent tags, usually Cy3 (green) or Cy5 (red). One is the control mRNA sample, and the other is the test sample for which the gene expression pattern is to be determined. The two labeled samples are mixed together

and competitively hybridized to the glass slide containing cDNA clones. The scanner output is usually two monochrome images, one for each of the two lasers in the scanner. A detector measures the ratio of hybridization signals of the two labels. These are combined to create the red–green color images of microarrays. The images are usually stored as tagged image files. The array data are stored in 16 bits [15]. cDNA microarrays can also be produced as home-spotted arrays, like the Genomic Medicine Research Core Laboratory's (GMRCL) Human 15K set, version 2, which has been deposited in the Gene Expression Omnibus repository with the accession number GPL5354 [17].

Design of microarray expression experiments

Three questions are specifically associated with microarray experiments when choosing the arrangement of samples [15,18,19]: (1) Are Affymetrix arrays or a two-color array system more effective? (2) If a two-color array system is used, what is the reference sample? (3) If a two-color array system is used, what is the best arrangement of samples on the slides? Although there are no standard answers to these questions, there are some statistical reasons why some of these designs are better than others.

In an example of study of cervical cancer using MAGE, samples were obtained from diseased and corresponding control tissues from 30 patients suffering from cervical cancer (Figure) [6]. Five different experimental designs could be used for this experiment [15]. In Design 1, sixty two-color microarrays were used. Every control and tumor sample was prepared with Cy5. A reference sample of relevant cell lines was labeled with Cy3. Each array was hybridized with a cervical sample labeled in Cy5 and the reference sample in Cy3. In Design 2, sixty single-color Affymetrix arrays were used. Each array was hybridized with a different sample. In Design 3, thirty two-color, cDNA microarrays were used. On each array, the control sample was labeled with Cy3 and the tumor sample from the same patient was labeled with Cy5. In Design 4, thirty two-color cDNA microarrays were used. On fifteen arrays, the control sample was labeled with Cy3 and the tumor sample was labeled with Cy5. On the other fifteen arrays, the control sample was labeled with Cy5 and the tumor sample was labeled with Cy3. In Design 5,

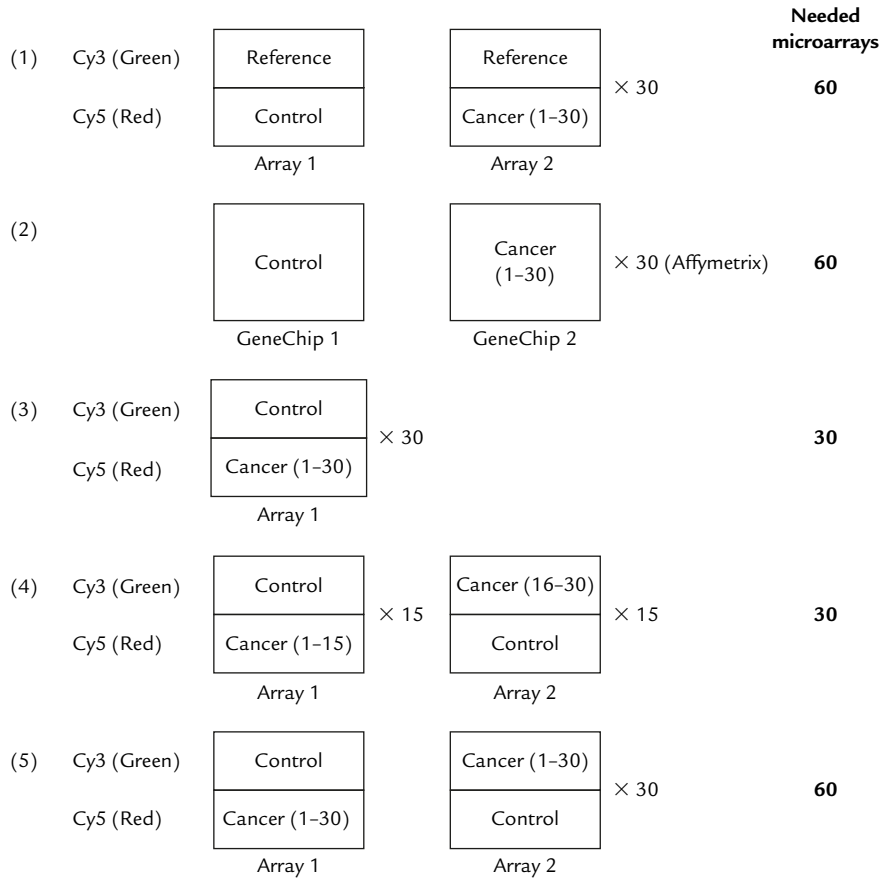

Figure. Choice of microarray platforms and various experimental designs for microarray analyses of gene expression. Figure adapted from Stekel et al [15].

sixty two-color cDNA microarrays were used in this dye-swapping design. Each pair of the control and tumor samples was analyzed twice. Control and tumor samples from each patient were hybridized to two arrays, once with Cy3 to label the control and Cy5 for the tumor, and once with Cy5 on the control and Cy3 on the tumor.

In the above example, since the two samples from each patient were obviously related to each other, it was more suitable to use a two-color array in which the two samples from each patient were hybridized to the same array. There are several levels of variability in the measured gene expression of the microarray experiments [15]. The four levels of variability are: (1) between replicate features on the same array spots; (2) between two separately labeled samples hybridized to the same array; (3) between samples hybridized to different arrays (inter-sample); and (4) between different individuals in a population hybridized to different arrays (inter-individual). Since the variability between arrays is higher than the variability between signals on the same array, this experiment was better performed on two-color arrays so that the two samples from each patient could be hybridized to the same array rather than on different arrays, which would be the case with Affymetrix arrays. Experimental design 5 was a full-factorial design, because each patient had each sample hybridized twice, once with each Cy dye. There are two main advantages to this method. Firstly, there are two measurements of log ratio for each patient, both using the same number of arrays in experimental design 1. Secondly, the data can be analyzed with a *t* test or bootstrap *t* test, whereas Design 4 would require ANOVA analysis. The only disadvantage of Design 5 is that it requires twice as many arrays as Design 4 and twice as many labeling reactions as any of the other designs.

Another example is the study of cervical cancer in patients undergoing concurrent chemoradiation viewed in time series [20]. Peripheral blood samples were taken at four time points: twice before treatment and twice after treatment. We were interested in identifying genes that were overexpressed but showed similar profiles throughout the time course. The patient samples were collected at four time points and a common reference was prepared. About 10 healthy control samples were pooled and used as the common reference. The samples and the common reference were then analyzed in a dye-swapping fashion. With this design, each sample is compared to the common reference. The hypothesis is that, if an array is particularly bright, it will be bright for both the sample and the reference, so the log ratio will be free from this artifact. The data from a failed array can be omitted, and data of other time points are still usable for analyzing the overall change of gene expression in patients during concurrent chemoradiation.

Image analysis

The image of the microarray generated by the scanner is the raw data of a microarray [21,22]. Feature extraction software converts the image into numerical data that quantify gene expression by using the positions of the features on the microarray (such as the pixels on the image that are part of the features and the nearby pixels that will be used for background calculation) to calculate the numerical information of the intensity of the feature, the intensity of the background, and quality control information. The spot and background intensities can be acquired with GenePix Pro 4.1 (Axon Instruments Inc., Union City, CA, USA) software [17].

Normalization

Normalization is used to resolve the systematic errors and bias in the microarray experimental platform [18,23–25]. It can be classified into within-array normalization and between-array normalization. Within-array normalization is achieved by using the tools in the MATLAB 6.0 (The MathWorks Inc., Natick, MA, USA) software to eliminate the three sources of systematic bias [23]: (1) the difference in DNA abundance incorporating into the Cy3 and Cy5 labels; (2) the different emission responses of Cy3 and Cy5 dyes; and (3) the difference in Cy3 and Cy5 intensities measured on different areas on the array. These biases can be corrected by using a linear regression of Cy5 against Cy3, a linear regression of log ratio against average intensity, and a non-linear (LOWESS) regression of the log ratio against average intensity, respectively [18]. The software and instructions for performing the LOWESS algorithm can be obtained from http://www.stat.Berkeley.EDU/users/terry/zarray/Html/index.html.

Between-array normalization is a comparison between samples hybridized to different arrays, which could be either two-color arrays or Affymetrix arrays. There are three standard methods: scaling, centering and distribution normalization. Between-assay normalization is used to minimize experimental variability without interfering with biological variability. Data are scaled to ensure that the means and standard deviations of all the distributions are equal.

Clustering methods to analyze gene expression

Analytical analysis of gene expression includes hierarchical clustering, self-organized mapping, principal components analysis, *K*-means clustering, and nearest-neighbor clustering. Hierarchical clustering is the most commonly used method of identifying groups of closely

related genes or samples [26,27]. Hierarchical clustering is a method that successively links genes or samples with similar profiles to form a tree structure that is very much like a phylogenetic tree. This technique has become popular for four reasons: (1) it simplifies large volumes of data; (2) the analysis reveals groups of similar genes; (3) it visualizes the data in a hierarchical way using interactive computer programs; and (4) the results are visualized in a style which is familiar to many geneticists. The algorithm calculates distances between genes or between samples using single linkage, complete linkage, or average linkage. Complete linkage defines the distance between two clusters as the average of the distances between all pairs of points in the two clusters. Average linkage is an intermediate value between single and complete linkage and tends to perform well in many microarray applications.

Principal components analysis is used to reduce multidimensional data to a single x-y graph [28–30]. The data from expression arrays are often of high dimensionality. This analysis allows the data to be reduced into a lower dimensionality that is more suitable for visual analysis. When reducing the gene dimensions, the first step is to construct a variance–covariance matrix for the genes. Then, a new variable called the eigenvector is transformed from the matrices. The first principal component is the axis that captures the most variation between patients and has the highest eigenvalue. The second principal component is orthogonal to the first, maximizing the remaining variance.

Self-organized mapping is a related method that allocates genes or samples to a predefined number of clusters that relate to each other on a spatial grid [31]. *K*-means clustering is a non-hierarchical clustering that allows genes or samples to cluster appropriately. *K*-nearest-neighbors clustering is the simplest method for grouping samples [26]. The general form of *k*-nearest-neighbors clustering uses *k* nearest neighbors and proceeds as follows: (1) for each patient, the *k* nearest neighbors is found according to the distance metric used, and (2) the class that is most common among the *k* neighbors is predicted.

Minimal Information About a Microarray Experiment (MIAME) compliance

When interpreting microarray data, authors are often requested by editors of many journals to submit sufficient information in the MIAME format that was proposed by the Microarray Gene Expression Data group [32]. Authors can deposit the data in either one of the two main public repositories, Gene Expression Omnibus (http://www.ncbi.nlm.nih.gov/geo/) and ArrayExpress (http://www.ebi.ac.uk/arrayexpress/). There are six components to MIAME, including experimental design, array design, samples, hybridizations, measurements, and normalization of control subjects, all of which should be included in every microarray publication [1]. The reason for this requirement by many journals is that microarrays are frequently manufactured independently of the experiment and are then used for a wide range of experiments.

Target preparation

RNA preparation from tissues that were snap-frozen in liquid nitrogen and those that were stored in the reagent RNAlater (Ambion, Austin, TX, USA) were equally intact [17]. However, rapid degradation of RNA was noted when the tissues were kept at room temperature for as short a time period as 10 minutes. A lab-on-a-chip device RNA LabChip read on the Bioanalyzer 2100 (Agilent Technologies, Palo Alto, CA, USA) is usually used to evaluate RNA quality and quantity. With as little as 25–500 ng of RNA input, the Bioanalyzer 2100 can calculate the ratio of 28S to 18S ribosomal RNA and indicate the concentrations of total RNA. On average, the 28S/18S ratios of RNA isolated from tissues are between 1.6 and 2.0.

The PAXgene Blood RNA Tubes and PAXgene Blood RNA Kit (QIAGEN, Mississauga, Ontario, Canada) system contains a lysis buffer and a stabilization solution, which is a mixture of cationic detergent and salts. The system allows standardized blood collection in optimal conditions [33–35]. High-throughput microarray technology in peripheral blood transcriptome has enabled us to differentiate between healthy and diseased individuals [36,37] and recognize diseases such as schizophrenia and bipolar disorder [38], diabetic retinopathy [39], and cancer [36,40,41].

Verification of microarray results

Microarrays are powerful tools for examining the expression of thousands of mRNAs, but two potential sources of error exist. First, as microarrays rely on hybridization, accurate measurement is possible only when each feature on the microarray is in great excess over the amount of mRNA in the test sample. Second, cross-hybridization, especially in spotted cDNA microarrays, can be another source of error. Therefore, confirmation of the results of gene expression levels is frequently required. One common method is using real-time quantitative PCR (RT-QPCR). Conventional PCR is not quantitative, because the end product is observed only after the bulk of the product has been synthesized. When the rate of synthesis of new molecules has reached a plateau, small differences in the amount of target at the start of the reaction are masked. Real-time measurement of

fluorescent product accumulation provides the number of cycles required to attain the linear phase of increase in product biosynthesis [42]. With each cycle, the fluorescence intensity reflects the number of molecules of double-stranded DNA that have been synthesized.

The end point of RT-QPCR analysis is the threshold cycle or C_T. The C_T is determined from a log-linear plot of the PCR signal versus the cycle number. The two most commonly used methods of analyzing data from RT-QPCR experiments are absolute quantification and relative quantification. Absolute quantification determines the input copy number, usually by relating the PCR signal to a standard curve [42]. Relative quantification measures the PCR signal of the target transcript in a treatment group and compares it to that of another sample (usually an untreated control). The $2^{-\Delta\Delta C_T}$ method is convenient in analyzing the relative changes in gene expression from RT-QPCR experiments [42]. In the demonstration by Livak and Schmittgen [42], the efficiency of the amplification of the target gene (*c-myc*) and the internal control (glyceraldehyde-3-phosphate dehydrogenase [GAPDH]) was examined using RT-QPCR and TaqMan detection. Using reverse transcriptase, cDNA was synthesized from 1 μg of total RNA isolated from human Raji cells. Serial dilutions of cDNA were amplified in a real-time PCR using gene-specific primers. The sample contained cDNA derived from 1 ng of total RNA. The ΔC_T (C_T c-myc – C_T GAPDH) was calculated for each cDNA dilution. If the absolute value of the slope is close to zero, the efficiencies of the target and reference genes are similar, and the $\Delta\Delta C_T$ calculation or the relative quantification of target may be used.

Important steps in the design and evaluation of a RT-QPCR experiment are: (1) to select an internal control gene, (2) to validate the internal control to determine that it is not affected by experimental treatment, and (3) to perform a PCR on serial dilutions of RNA or cDNA for both the target and internal control genes to ensure that the efficiencies are similar. Finally, statistical data should be converted to the linear form and should not be presented as raw C_T values. The purpose of the internal control gene is to normalize the PCRs for the amount of RNA added to the reverse transcription reactions. Standard housekeeping genes usually suffice as internal control genes. Suitable internal controls for RT-QPCR include GAPDH, β-actin, β_2-microglobulin, and rRNA.

MAGE in Cervical Cancer Research

Implications from cervical cancer cell lines

Gene expression profiling in two morphologically different uterine cervical carcinoma cell lines derived from a single donor was studied using a human cancer cDNA array [43]. Among the 10 selected genes, insulin-like growth factor binding protein-3, inhibitor of apoptosis protein-1, and cadherin-13 were more frequently expressed in squamous cell carcinoma (SC) cell lines. Interferon-inducible genes, *Sno* oncogenes, and transforming growth factor (TGF)-β2 receptor were expressed in both SC and adenocarcinoma (AC) cell lines. It was suggested that these genes are involved in the morphologic characteristics and carcinogenesis of cervical cancer.

With the introduction of microarray technology, large-scale transcriptional changes have been found after irradiation. Patterns of gene sets characterizing persistent responses to radiation and those characterizing immediate responses, as well as genes linked to radiosensitivity and resistance towards irradiation in clinical settings, have been reported [44–48].

Two cell lines with different radiosensitivity originated from the same tumor of a cervical cancer patient who, prior to treatment, exhibited different gene expression patterns from the gene expression exhibited when receiving treatment [49]. The overexpressed genes of the radioresistant cell line included metal-regulatory transcription factor-1, cytochrome P450 *CYP1B1*, adenomatosis polyposis coli, translation elongation factor-1, and cytochrome-c oxidase; whereas, the radiosensitive cell line overexpressed the transcription factor NF-κB, metalloproteinase inhibitor-1 precursor, superoxide dismutase 2, insulin-like growth factor binding protein-3, guanine nucleotide-binding protein, and TGF-β-induced protein. This early report has shed some light on our understanding of the biological pathways contributing to radiation resistance and may lead to the development of alternative treatment modalities.

After exposure to 2 Gy of X-rays, Chaudhry et al examined the expression patterns in HeLa cells in the G1 and G2 phases of the cell cycle and identified some previously known and unknown genes associated with radiation [50]. For example, none of the genes related to p53 was upregulated, since HeLa cells are deficient in p53. In addition, both phases of the cell cycle encode ribosomal proteins. Chen et al demonstrated that NF-κB was responsible for a major part of the radioresistance observed in a cell population (HK18-IR) derived from HK18 cells by fractionated ionizing radiation (2 Gy/fraction; total dose, 60 Gy) [51].

In cervical carcinoma, extracellular levels of TGF-β1 are increased in the late stage of malignancy [52,53]. Pathways such as the tumor necrosis factor-α (TNF-α), mitogen-activated protein kinase (MAPK), wingless type (Wnt) and Smad pathways that are involved in the cellular processes of proliferation, differentiation, angiogenesis, wound healing and inhibition of epithelial

cell growth of cervical cancer have been investigated at the transcriptional level. Kloth et al observed substantial changes in the gene expression of Wnt, MAPK and TNF-α pathways, which appear to be induced by TGF-β1 in cervical cancer cell lines [54]. Changes in gene expression in TGF-β1 that stimulated cervical cancer cell lines with high (CC10B), intermediate (SiHa) and low (HeLa) sensitivity to the antiproliferative effect of TGF-β1 were investigated at time points 0, 6, 12 and 24 hours. Increased resistance to TGF-β1-induced cell growth inhibition was correlated with an elevated production of TGF-β1 [54].

Differentiation between malignant and normal cervical epithelia

Much effort has been given to the investigation of differentially expressed genes in cervical cancer (Table) and the evaluation of the feasibility of gene expression profiling on squamous cervical cells obtained from cervical swabs [55]. Hudelist et al showed the feasibility of gene expression profiling of cervical squamous cells obtained with cytobrushes by identifying a characteristic gene expression pattern that clearly distinguishes between malignant and normal cervical epithelia of squamous type [55]. Cervical cells from three women with high-risk human papillomavirus-positive invasive squamous cervical carcinoma and from three human papillomavirus-negative women with normal ectocervical smears were analyzed with cDNA array. Immunoblot analysis was performed to detect the proteins that corresponded to the highest upregulated genes with cDNA array. mRNA expressions of *ERBB2*, *KIT*, *FLT1*, *MYCN*, *RAS*, *CDKN2A*, *CCND1*, *NME1*, *NME2*, *MET*, *FGF7*, *FGFR2*, and *STAT1* were increased in malignant

Table. Studies on gene expression profiling in cervical cancer patients

Study	Upregulated gene expression changes	Reference
Cell swab		
Smear of HR-HPV SC ($n=3$) and normal ectocervix ($n=3$)	≥ 10 fold in cancer cells of *MET* proto-oncogene, *NME1*, *NME2*, *EGFR*, *FGFR2*, *ERBB2* proto-oncogene, *CDKN2A/p16^{INK2A}*	Hudelist et al [55]
Tissues		
Stage IB–IIA cancer and normal tissue pairs ($n=5$)	Guanine nucleotide-binding protein Gs, leukocyte adhesion protein (LFA1-β), NF45, Hox-A1, β-catenin in cancer	Shim et al [57]
Various stages of cancer ($n=38$) and normal tissue ($n=5$)	*G32C4B* (NADH dehydrogenase 4 gene) and *G30CC* (ribosomal protein S12 gene) as early pre-transformation diagnostic markers	Cheng et al [56]
SC ($n=9$), AC ($n=3$), HSIL ($n=5$), LSIL ($n=5$), normal ($n=12$)	Cellular proliferation, extracellular matrix-associated genes in cancer	Chen et al [58]
SC, AC, NGE, NSE ($n=10$ in each group)	*c-myc, RhGC, KRT-4* in SC	Contag et al [61]
Early (I–II; $n=19$) and late stage (III–IV; $n=10$) in SC	*CTGF* and *RGS1* in late stage cancer	Wong et al [62]
Combined CGH and expression profiles ($n=31$)	Amplification-associated genes at 8q24.3, 11q22.2, 20q13 involved in cervical cancer pathogenesis	Narayan et al [68]
Response to radiation		
Radiation-sensitive ($n=17$) and radiation-resistant SC tumors	*ICAM3* in radiation-resistant tumors	Chung et al [47]
Prognostic outcome		
Radiation patients ($n=19$)	A predictive score with 35 genes in apoptosis (*BIK*, *TEGT*, *SSI-3*), hypoxia-inducible genes (*HIF1A*, *CA12*), tumor invasion (*CTSL*, *CTSB*, *PLAU*, *CD44*)	Harima et al [46]
SC ($n=16$), AC ($n=9$), and ASC ($n=5$)	*TACSTD1* and *CEACAM5* as poor prognostic factor in AC and ASC patients	Chao et al [6]

HR-HPV = high-risk human papillomavirus; SC = squamous cell carcinoma; AC = adenocarcinoma; HSIL = high-grade squamous intraepithelial lesion; LSIL = low-grade squamous intraepithelial lesion; NGE = normal glandular epithelium; NSE = normal squamous epithelium; CGH = comparative genomic hybridization; ASC = adenosquamous cell carcinoma.

samples. Several genes that are associated with anti-apoptosis (such as *BCL2*), cell structuring or cell attachment were also upregulated in carcinoma cells [55].

Pattern distinction between different cell types and biomarker discovery of aggressive cancer

Utilizing RT-QPCR and cDNA microarrays, Cheng et al identified cellular genes that are involved in the multistep carcinogenesis of SC [56]. Thirty-eight cervical cancer patients in various stages of the disease and five non-cervical cancer patients were studied. Northern blot analysis and RNA–RNA *in situ* hybridization studies were also used for verification. When compared with sequences available in the GenBank database, two clones (*G32C4B* and *G30CC*) were identified to be the NADH dehydrogenase 4 gene and the gene that encodes ribosomal protein S12, respectively. Increased expression of these two genes was detected in the matched normal tissues collected together with the late International Federation of Gynecology and Obstetrics (FIGO) stages of cervical cancer biopsies. In comparison, upregulation of these two genes was not detected in benign cervical squamous epithelia, suggesting that expression of these two genes in the adjacent histopathologically "normal" cervical squamous epithelial tissue may be different from that in cervical cancer [56].

Systematic analysis of the expression pattern during the various steps of cervical tumorigenesis revealed that: (1) there are specific patterns of gene expression in cellular processes during carcinogenesis, and these profiles can discriminate between normal cervical tissue and low-grade cervical intraepithelial neoplasia (CIN) from high-grade CIN and cervical cancers; (2) many of these genes are also expressed in the stroma adjacent to the cancer tissue; and (3) the extent of gene overexpression is increased during the progression from low-grade to high-grade CIN and finally to cancer [57–60].

Comparisons of gene expression in SC and AC of the uterine cervix using microarray experiments have been reported [6,61]. Selected differentially expressed genes between AC and SC were further verified using RTQ-PCR and immunohistochemistry. Genes including *CEACAM5*, *TACSTD1*, *S100P* and *MSLN* were upregulated in AC. On the contrary, genes involved in the epidermal differentiation complex, such as *S100A9* and *ANXA8*, were upregulated in SC. In a group of patients with known long-term outcomes ($n=63$; median follow-up, 70.3 months; range, 4–208 months), the correlation between the selected six differentially expressed genes and histology was highly significant. Increased expression levels of *CEACAM5* and *TACSTD1* were significantly correlated with poorer patient outcomes using a multivariate Cox proportional hazards regression analysis. The combination of the cDNA microarray, RTQ-PCR and immunohistochemical results of this study showed that it is possible to define different gene profiles for AC and SC. Moreover, *TACSTD1* expression may be a novel poor prognostic factor [6,61]. In another study in Hong Kong, two genes, *CTGF* and *RGS1*, were found to be more upregulated in late-stage cancer than in early-stage cancer, suggesting that they may be involved in cancer progression. The pathway analysis of expression data showed that the *SPP1*, *VEGF*, *CDC2* and *CKS2* genes were coordinately and differentially regulated between cancer and normal tissues [62].

Association of expression profiling with radiation

Gene expression profiling by DNA microarray has been applied to classify disease stages and predict treatment response to radiotherapy in cervical cancer [44–48,63]. Wong et al classified tumors into two major groups based on their responses to radiotherapy, and they were able to predict the response of these patients to radiotherapy from their expression profiles using a prediction model [45]. The top 100 differentially expressed genes represented a wide spectrum of cellular functions, including transcription, cell adhesion, membrane and cytoskeletal, and signal transduction functions [45].

Kitahara et al studied the gene profiles of 10 radioresistant cervical SCs versus nine radiosensitive samples that had a complete response to radiation [44]. Sixty-two genes were used to classify the tumors, and the resistant cells had higher levels of the DNA repair gene, X-ray repair cross-complementing 5 (*XRCC5*), and its gene product Ku80. Subsequent follow-up analysis of Ku80 levels in 89 cervical cancer patients revealed that low expression of Ku80 protein was correlated with sensitivity to radiation [63].

Harima et al identified 35 genes that could clearly separate the thermoradiosensitive group from the thermoradioresistant group and further divided these genes into three groups: apoptosis genes (e.g. *BIK*, *TEGT*, *SSI-3*), hypoxia-inducible genes (e.g. *HIF1A*, *CA12*), and tumor cell invasion and metastasis genes (e.g. *CTSL*, *CTSB*, *PLAU*, *CD44*) [46]. These findings echoed the hypothesis that increased tumor hypoxia leads to radiation resistance and increased tumor angiogenesis [64,65].

Narrowing down the genome-wide expression profiling, Chung et al first compared gene expression profiles between the parental SiHa cervical cancer cells and the derived radiation-resistant SiHa/R cells. Expression of *ICAM-3* was upregulated in SiHa/R cells. Further verification showed that the overexpression of *ICAM-3* was significantly more frequent in radiation-resistant cervical cancer specimens when compared with radiation-sensitive specimens (83.3% *vs.* 35.3%; $p=0.015$). These

observations suggest that the expression of *ICAM-3* may be used as a biomarker to predict the radiation resistance in cervical cancer that occurs during radiotherapy [47].

Using an *in vitro* chemoradiation response assay after exposure to 3 Gy γ-irradiation, Tewari et al divided eight viable primary untreated cervical cancer specimens into extreme ($n=4$), intermediate ($n=2$) and low radiation resistance ($n=2$) categories. They identified a radiation response gene set of 54 genes transcripts with 100% accuracy for classifying tumors according to their radiation response [66]. The functional classes of these genes include signal transduction, apoptotic pathway, cell cycle regulation, DNA repair, RNA transcription, and protein transport.

To identify the genes that may be involved in the development of late tissue injury, genes that were differentially expressed in mouse tissues with specific vascular damage were studied [67]. Microarray experiments were performed using amplified RNA isolated from irradiated mouse kidney and rectum and from sham-irradiated controls at 10 and 20 weeks after treatment. The mouse kidney experiments showed that jagged 1 and Kruppel-like factor 5 (*KLF5*), which are reported to play a role in vascular development and remodeling, were upregulated in both tissues.

Chromosomal amplifications and associated expression changes in cervical cancer

The combining of genomic alteration and gene expression to examine the molecular changes in cervical cancer has been reported [68]. The use of cDNA array comparative genomic hybridization and fluorescence *in situ* hybridization to analyze 29 cervical cancers demonstrated the presence of high-level amplifications at the 8q24.3, 11q22.2 and 20q13 regions. These regions contain several known tumor-associated genes, such as those involved in transcription, apoptosis, cytoskeletal remodeling, ion-transport, drug metabolism, and immune response. The changes in expression of the amplification-associated genes were then analyzed with Affymetrix U133A expression arrays and semi-quantitative reverse-transcription PCR in 31 cervical cancers [68]. The identification of characteristic focal amplicons and expression profiles has facilitated the understanding of cervical cancer pathogenesis.

Conclusion

Microarray technology has been used in the search for gene profiles that are associated with different histologic types, responses to radiation, and prognostic outcomes. Results of these studies have led us to better understand the molecular events in cervical cancer. Nevertheless, many pertinent questions remain unanswered, such as the transcriptional changes in radiation and the functions of differentially expressed genes of cervical cancer during treatment. At this moment, microarrays are providing an opportunity for expansive research that can help unravel the mysteries of tumor biology, but they are still too costly for widespread clinical use in gynecologic cancer.

Acknowledgments

The authors would like to thank Dr Yun-Shien Lee (Department of Biotechnology, Ming Chuan University) for his long-term contribution to the Genomic Medicine Research Core Laboratory and Shih-Yee Mimi Wang (University of Illinois, Chicago College of Medicine) for editing the manuscript. This study was supported by grants NSC 93-2314-B-182A-102 (to TH Wang) and NMRPG460011 (to A Chao) from National Science Council, and CMRPG 340462 (to TH Wang) and CMRPG360951 (to A Chao) from Chang Gung Memorial Hospital, Taiwan.

References

1. Brazma A, Hingamp P, Quackenbush J, et al. Minimum information about a microarray experiment (MIAME)-toward standards for microarray data. *Nat Genet* 2001;29:365–71.
2. *Cancer Registration Annual Report in Taiwan*. Department of Health, Executive Yuan, Taiwan, 2002.
3. Munoz N, Bosch FX, de Sanjose S, et al. Epidemiologic classification of human papillomavirus types associated with cervical cancer. *N Engl J Med* 2003;348:518–27.
4. Bosch FX, Manos MM, Munoz N, et al. Prevalence of human papillomavirus in cervical cancer: a worldwide perspective. International Biological Study on Cervical Cancer (IBSCC) Study Group. *J Natl Cancer Inst* 1995;87:796–802.
5. zur Hausen H. Papillomaviruses causing cancer: evasion from host-cell control in early events in carcinogenesis. *J Natl Cancer Inst* 2000;92:690–8.
6. Chao A, Wang TH, Lee YS, et al. Molecular characterization of adenocarcinoma and squamous carcinoma of the uterine cervix using microarray analysis of gene expression. *Int J Cancer* 2006;119:91–8.
7. Lai CH, Huang HJ, Hsueh S, et al. Human papillomavirus genotype in cervical cancer: a population-based study. *Int J Cancer* 2007;120:1999–2006.
8. Hacker NF. Cervical Cancer. In: Berek JS, Hacker NF, eds. *Practical Gynecologic Oncology*, 3rd edition. Philadelphia: Lippincott Williams & Wilkins, 2000:345–405.
9. Green JA, Kirwan JM, Tierney JF, Symonds P, Fresco L, Collingwood M, Williams CJ. Survival and recurrence after

concomitant chemotherapy and radiotherapy for cancer of the uterine cervix: a systematic review and meta-analysis. *Lancet* 2001;358:781–6.
10. Snyder AR, Morgan WF. Gene expression profiling after irradiation: clues to understanding acute and persistent responses? *Cancer Metastasis Rev* 2004;23:259–68.
11. Fodor SP, Read JL, Pirrung MC, Stryer L, Lu AT, Solas D. Light-directed, spatially addressable parallel chemical synthesis. *Science* 1991;251:767–73.
12. Fodor SP, Rava RP, Huang XC, Pease AC, Holmes CP, Adams CL. Multiplexed biochemical assays with biological chips. *Nature* 1993;364:555–6.
13. Schena M, Shalon D, Davis RW, Brown PO. Quantitative monitoring of gene expression patterns with a complementary DNA microarray. *Science* 1995;270:467–70.
14. Wang TH, Chao A. Microarray analysis of gene expression of cancer to guide the use of chemotherapeutics. *Taiwan J Obstet Gynecol* 2007;46:222–9.
15. Stekel D. *Microarray Bioinformatics*. Cambridge: Cambridge University Press, 2003.
16. Duggan DJ, Bittner M, Chen Y, Meltzer P, Trent JM. Expression profiling using cDNA microarrays. *Nat Genet* 1999;21:10–4.
17. Wang TH, Lee YS, Chen ES, et al. Establishment of cDNA microarray analysis at the Genomic Medicine Research Core Laboratory (GMRCL) of Chang Gung Memorial Hospital. *Chang Gung Med J* 2004;27:243–60.
18. Yang YH, Speed T. Design issues for cDNA microarray experiments. *Nat Rev Genet* 2002;3:579–88.
19. Witmer JA, Samuels ML. *Statistics for the Life Sciences*, 3rd edition. New Jersey: Prentice Hall, 2002.
20. Chao A, Wang TH, Lee YS, et al. Analysis of functional groups of differentially expressed genes in the peripheral blood of patients with cervical cancer undergoing concurrent chemoradiation. *Radiat Res* (In press).
21. Adams R, Bischof L. Seeded region growing. *IEEE Trans Pattern Anal Mach Intell* 1994;16:641–47.
22. Kamberova G, Shah S. *DNA Array Image Analysis: Nuts and Bolts*. Eagleville: DNA Press, 2002.
23. Quackenbush J. Microarray data normalization and transformation. *Nat Genet* 2002;32:496–501.
24. Kooperberg C, Fazzio TG, Delrow JJ, Tsukiyama T. Improved background correction for spotted DNA microarrays. *J Comput Biol* 2002;9:55–66.
25. Huber W, von Heydebreck A, Sultmann H, Poustka A, Vingron M. Variance stabilization applied to microarray data calibration and to the quantification of differential expression. *Bioinformatics* 2002;18(Suppl 1):S96–104.
26. Everitt BS. *Cluster Analysis*, 3rd edition. London: Edward Arnold, 1993.
27. Chatfield C, Collins AJ. *Introduction to Multivariate Analysis*. London: Chapman & Hall, 1980.
28. Dysvik B, Jonassen I. J-Express: exploring gene expression data using Java. *Bioinformatics* 2001;17:369–70.
29. Raychaudhuri S, Stuart JM, Altman RB. Principal components analysis to summarize microarray experiments: application to sporulation time series. *Pac Symp Biocomput* 2000:455–66.
30. Xia X, Xie Z. AMADA: analysis of microarray data. *Bioinformatics* 2001;17:569–70.
31. Kohonen T. *Self Organized Maps*. Berlin: Springer, 2000.
32. Microarray standards at last. *Nature* 2002;419:323.
33. Rainen L, Oelmueller U, Jurgensen S, et al. Stabilization of mRNA expression in whole blood samples. *Clin Chem* 2002;48:1883–90.
34. Thach DC, Lin B, Walter E, Kruzelock R, Rowley RK, Tibbetts C, Stenger DA. Assessment of two methods for handling blood in collection tubes with RNA stabilizing agent for surveillance of gene expression profiles with high density microarrays. *J Immunol Methods* 2003;283:269–79.
35. Chai V, Vassilakos A, Lee Y, Wright JA, Young AH. Optimization of the PAXgene blood RNA extraction system for gene expression analysis of clinical samples. *J Clin Lab Anal* 2005;19:182–8.
36. Whitney AR, Diehn M, Popper SJ, Alizadeh AA, Boldrick JC, Relman DA, Brown PO. Individuality and variation in gene expression patterns in human blood. *Proc Natl Acad Sci USA* 2003;100:1896–901.
37. Lee YS, Chen CH, Chao A, et al. Molecular signature of clinical severity in recovering patients with severe acute respiratory syndrome coronavirus (SARS-CoV). *BMC Genomics* 2005;6:132.
38. Tsuang MT, Nossova N, Yager T, et al. Assessing the validity of blood-based gene expression profiles for the classification of schizophrenia and bipolar disorder: a preliminary report. *Am J Med Genet B Neuropsychiatr Genet* 2005;133:1–5.
39. Hamaoui K, Butt A, Powrie J, Swaminathan R. Concentration of circulating rhodopsin mRNA in diabetic retinopathy. *Clin Chem* 2004;50:2152–5.
40. Li D, Butt A, Clarke S, Swaminathana R. Real-time quantitative PCR measurement of thyroglobulin mRNA in peripheral blood of thyroid cancer patients and healthy subjects. *Ann N Y Acad Sci* 2004;1022:147–51.
41. Muller MC, Merx K, Weisser A, Kreil S, Lahaye T, Hehlmann R, Hochhaus A. Improvement of molecular monitoring of residual disease in leukemias by bedside RNA stabilization. *Leukemia* 2002;16:2395–9.
42. Livak KJ, Schmittgen TD. Analysis of relative gene expression data using real-time quantitative PCR and the $2^{-\Delta\Delta T}$ method. *Methods* 2001;25:402–8.
43. Fujimoto T, Nishikawa A, Iwasaki M, Akutagawa N, Teramoto M, Kudo R. Gene expression profiling in two morphologically different uterine cervical carcinoma cell lines derived from a single donor using a human cancer cDNA array. *Gynecol Oncol* 2004;93:446–53.
44. Kitahara O, Katagiri T, Tsunoda T, Harima Y, Nakamura Y. Classification of sensitivity or resistance of cervical cancers to ionizing radiation according to expression profiles of 62 genes selected by cDNA microarray analysis. *Neoplasia* 2002;4:295–303.
45. Wong YF, Selvanayagam ZE, Wei N, et al. Expression genomics of cervical cancer: molecular classification and prediction of radiotherapy response by DNA microarray. *Clin Cancer Res* 2003;9:5486–92.
46. Harima Y, Togashi A, Horikoshi K, et al. Prediction of outcome of advanced cervical cancer to thermoradiotherapy according to expression profiles of 35 genes selected by cDNA microarray analysis. *Int J Radiat Oncol Biol Phys* 2004; 60:237–48.
47. Chung YM, Kim BG, Park CS, et al. Increased expression of ICAM-3 is associated with radiation resistance in cervical cancer. *Int J Cancer* 2005;117:194–201.

48. Wong YF, Sahota DS, Cheung TH, et al. Gene expression pattern associated with radiotherapy sensitivity in cervical cancer. *Cancer J* 2006;12:189–93.
49. Achary MP, Jaggernauth W, Gross E, Alfieri A, Klinger HP, Vikram B. Cell lines from the same cervical carcinoma but with different radiosensitivities exhibit different cDNA microarray patterns of gene expression. *Cytogenet Cell Genet* 2000;91:39–43.
50. Chaudhry MA, Chodosh LA, McKenna WG, Muschel RJ. Gene expression profile of human cells irradiated in G1 and G2 phases of cell cycle. *Cancer Lett* 2003;195:221–33.
51. Chen X, Shen B, Xia L, Khaletzkiy A, Chu D, Wong JY, Li JJ. Activation of nuclear factor kappaB in radioresistance of TP53-inactive human keratinocytes. *Cancer Res* 2002;62: 1213–21.
52. Chopra V, Dinh TV, Hannigan EV. Circulating serum levels of cytokines and angiogenic factors in patients with cervical cancer. *Cancer Invest* 1998;16:152–9.
53. Comerci JT Jr, Runowicz CD, Flanders KC, De Victoria C, Fields AL, Kadish AS, Goldberg GL. Altered expression of transforming growth factor-beta 1 in cervical neoplasia as an early biomarker in carcinogenesis of the uterine cervix. *Cancer* 1996;77:1107–14.
54. Kloth JN, Fleuren GJ, Oosting J, et al. Substantial changes in gene expression of Wnt, MAPK and TNFalpha pathways induced by TGF-beta1 in cervical cancer cell lines. *Carcinogenesis* 2005;26:1493–502.
55. Hudelist G, Czerwenka K, Singer C, Pischinger K, Kubista E, Manavi M. cDNA array analysis of cytobrush-collected normal and malignant cervical epithelial cells: a feasibility study. *Cancer Genet Cytogenet* 2005;158:35–42.
56. Cheng Q, Lau WM, Tay SK, Chew SH, Ho TH, Hui KM. Identification and characterization of genes involved in the carcinogenesis of human squamous cell cervical carcinoma. *Int J Cancer* 2002;98:419–26.
57. Shim C, Zhang W, Rhee CH, Lee JH. Profiling of differentially expressed genes in human primary cervical cancer by complementary DNA expression array. *Clin Cancer Res* 1998; 4:3045–50.
58. Chen Y, Miller C, Mosher R, et al. Identification of cervical cancer markers by cDNA and tissue microarrays. *Cancer Res* 2003;63:1927–35.
59. Pappa KI, Anagnou NP. Emerging issues of the expression profiling technologies for the study of gynecologic cancer. *Am J Obstet Gynecol* 2005;193:908–18.
60. Kendrick JE, Conner MG, Huh WK. Gene expression profiling of women with varying degrees of cervical intraepithelial neoplasia. *J Low Genit Tract Dis* 2007;11:25–8.
61. Contag SA, Gostout BS, Clayton AC, Dixon MH, McGovern RM, Calhoun ES. Comparison of gene expression in squamous cell carcinoma and adenocarcinoma of the uterine cervix. *Gynecol Oncol* 2004;95:610–7.
62. Wong YF, Cheung TH, Tsao GS, et al. Genome-wide gene expression profiling of cervical cancer in Hong Kong women by oligonucleotide microarray. *Int J Cancer* 2006; 118:2461–9.
63. Harima Y, Sawada S, Miyazaki Y, et al. Expression of Ku80 in cervical cancer correlates with response to radiotherapy and survival. *Am J Clin Oncol* 2003;26:e80–5.
64. Thomas G. The effect of hemoglobin level on radiotherapy outcomes: the Canadian experience. *Semin Oncol* 2001; 28:60–5.
65. Grogan M, Thomas GM, Melamed I, et al. The importance of hemoglobin levels during radiotherapy for carcinoma of the cervix. *Cancer* 1999;86:1528–36.
66. Tewari D, Monk BJ, Al-Ghazi MS, Parker R, Heck JD, Burger RA, Fruehauf JP. Gene expression profiling of *in vitro* radiation resistance in cervical carcinoma: a feasibility study. *Gynecol Oncol* 2005;99:84–91.
67. Kruse JJ, te Poele JA, Russell NS, Boersma LJ, Stewart FA. Microarray analysis to identify molecular mechanisms of radiation-induced microvascular damage in normal tissues. *Int J Radiat Oncol Biol Phys* 2004;58:420–6.
68. Narayan G, Bourdon V, Chaganti S, et al. Gene dosage alterations revealed by cDNA microarray analysis in cervical cancer: identification of candidate amplified and overexpressed genes. *Genes Chromosomes Cancer* 2007;46: 373–84.