Expression Genomics of Cervical Cancer: Molecular Classification and Prediction of Radiotherapy Response by DNA Microarray

Yick Fu Wong,1,9 Zachariah E. Selvanayagam,2,9 Nien Wei,5 Joseph Porter,6 Ragini Vittal,7 Rong Hu,5 Yong Lin,4 Jason Liao,4 Joe Weichung Shih,4 Tak Hong Cheung,1 Keith Wing Kit Lo,1 So Fan Yim,1 Shing Kai Yip,1 Danny Tse Ngong,1 Nelson Siu,1 Loucia Kit Ying Chan,1 Chun Sing Chan,1 Tony Kong,7 Elena Kutlina,8 Randall D. McKinnon,3 David T. Denhardt,6 Khew-Voon Chin,7 and Tony Kwok Hung Chung1

1Department of Obstetrics and Gynecology, Chinese University of Hong Kong, Prince of Wales Hospital, Hong Kong, People's Republic of China; Departments of 2Pediatrics and 3Surgery and 4The Cancer Institute of New Jersey, Robert Wood Johnson Medical School, University of Medicine and Dentistry of New Jersey, New Brunswick, New Jersey; 5Northridge Co., Warren, New Jersey; and 6Department of Cell Biology and Neuroscience, 7Susan Lehman Cullman Laboratory for Cancer Research, Department of Chemical Biology, and 8Department of Pharmaceutics, Ernest Mario School of Pharmacy, Rutgers University, Piscataway, New Jersey

ABSTRACT

Purpose: The incidence and mortality rates of cervical cancer are declining in the United States; however, worldwide, cervical cancer is still one of the leading causes of death in women, second only to breast cancer. This disparity is at least partially explained by the absence of or comparatively ineffective screening programs in the developing world. Recent advances in expression genomics have enabled the use of DNA microarray to profile gene expression of various cancers. These expression profiles may be suitable for molecular classification and prediction of disease outcome and treatment response. We envision that expression genomics applied in cervical cancer may provide a more rational approach to the classification and treatment of the disease.

Experimental Design: In this report, we examined the expression profiles of cervical cancer compared with normal cervical tissues in DNA microarrays that contained approximately 11,000 features that correspond to either human transcripts with known function or anonymous expressed sequence tags.

Results: Our results showed that normal cervical tissues were completely segregated from the cancer samples using about 40 genes whose expressions were significantly different between these specimens. In addition, clinical stage IB and stage IIB tumors could also be classified based on their signature expression patterns. Most importantly, some of the tumor samples were further stratified into two major groups based on their response to radiotherapy, and we were able to predict the response of these patients to radiotherapy from their expression profiles.

Conclusions: Gene expression profiling by DNA microarray may be used for further molecular classification of disease stages and prediction of treatment response in cervical cancer.

INTRODUCTION

Despite a decline in the incidence of cervical cancer in the United States, in less developed nations, cervical cancer remains one of leading causes of cancer mortality (1, 2). The absence of screening programs or comparatively ineffective screening programs lead to relatively late diagnosis of the disease (3). For advanced cervical cancer, radiotherapy remains the major treatment modality despite the comparatively low response rate (4).

Recent advances in expression genomics by DNA microarray demonstrated the potential use of expression profiles obtained from patients' specimens for molecular classification of cancer as well as disease outcome (5–9). Therefore, monitoring gene expression profiles for genome-wide changes in gene expression patterns may provide insights into the molecular fingerprint of different diseases including cancer, diseases of the central nervous system, and the cardiovascular system, treatments, environmental agents, and ultimately even distinguishing responders and nonresponders to a given drug, as well as predicting toxicity and adverse effects of drugs (10, 11).

Emergence of resistance in cancer treatments, whether by radio- or chemotherapy, remains a critical unresolved problem. Failure to respond to treatment is often the result of the empirical nature of cancer therapy or the inability to foresee or predict the response of an individual's cancer. Therefore, application of expression genomics in cervical cancer to identify genetic pathways that contribute to treatment resistance will undoubtedly provide insights into mechanisms of resistance and ultimately help guide and improve cervical cancer treatment. DNA microarray has already been used to monitor genome-wide gene expression changes in response to chemotherapeutic agents in in vitro cell culture studies (12–16), to gain further understanding about mechanisms of resistance as well as predict resistance in tumor specimen (14).

In this report, we conducted pilot studies to compare the
expression profiles of normal cervical tissues to cervical cancers by DNA microarray. We also showed that the expression profiles could classify cervical cancers into two broad clinical groups, based on the FIGO\textsuperscript{10} staging system. Most importantly, our results further showed that response to radiotherapy can be predicted by monitoring the expression profiles of some of these cervical cancers, which were stratified into responders and non-responders, based on the retrospective outcome of the treatment.

**MATERIALS AND METHODS**

**DNA Microarray.** DNA arrays were custom printed on 3 in x 10 in nylon membrane and contained approximately 11,000 DNA elements, comprising expressed sequence tags that either correspond to human transcripts with known function in the GenBank database (approximately 7,000) or are anonymous (>3,000).

**Tissue Specimens and RNA Isolation.** Patient specimens and clinical data with end points for correlation analysis were obtained from the Department of Obstetrics and Gynecology, Prince of Wales Hospital, at The Chinese University of Hong Kong. This study was approved by the Clinical Research Ethics Committee at the Chinese University of Hong Kong and the Internal Review Boards at Rutgers University. The study involves no potential risk to the patients because these are archival samples and were without identifier label.

RNAs were isolated using the RNeasy Kit (Qiagen) according to the manufacturer’s instructions. In brief, approximately 5 mg of pulverized tissue were homogenized in the RNA extraction reagent with a Dounce homogenizer. Integrity of isolated RNA was examined by gel electrophoresis.

**Target Synthesis, Array Hybridization, and Image Processing.** cDNA targets were synthesized from the isolated total RNA with \(^{33}\)PdCTP by oligo(dT)-primed polymerization using Superscript II reverse transcriptase (Life Technologies, Inc.) as described previously (16). \(^{33}\)PdCTP-labeled targets were hybridized to the membrane arrays, washed, exposed on a phosphorimage screen for approximately 15 h, and then scanned on a Molecular Dynamics Typhoon PhosphorImager. Scanned images of microarrays were analyzed using Imagene (Biodiscovery), and the output intensity data were further analyzed using custom statistical software and supervised clustering method. The analyzed results were displayed using the Stanford array analysis software suite Cluster andTreeView (17).

**Data Analysis.** To compare normal cervical tissues with cervical cancers, the expression level for each gene was subjected to Wilcoxon’s rank-sum test, and the \(P\) was adjusted according to Benjamini and Hochberg’s false discovery rate (18). A logistic regression model was applied to classify the samples as either normal or cancer using the genes with smallest Benjamini and Hochberg’s adjusted \(P\)s. Model gene predictors all have Benjamini and Hochberg’s adjusted \(P\)s of <0.01. Supervised clustering methods, involving training and testing of cervical cancer gene expression profiles and clinical treatment response, were used to predict and correlate gene expression patterns with treatment response.\textsuperscript{11}

**RESULTS**

There have been limited studies on the expression genomes of cervical cancer, and most of these published reports focused on tissue culture cell lines (19–22). To test whether monitoring the expression profiles of cervical cancer patients can be applied for molecular classification of the disease and predicting treatment outcome, we performed pilot studies on the expression profiles of cervical cancer samples from patients.

Twenty-six grossly dissected primary tumors from cervical cancer patients taken at the time of diagnosis and before radiotherapy were collected at the Prince of Wales Hospital, Department of Obstetrics and Gynecology, Chinese University of Hong Kong. The patients’ characteristics are shown in Table 1. Microscopic examination of tumor preparations verified the presence of cancer cells in all samples.

To classify the tumor samples, we used gene expression data from cDNA microarrays containing 10,692 elements (Fig. 1A). In the initial analysis, we compared the expression profiles of 8 normal cervixes derived from hysterectomy with 26 cervical cancer samples, randomly shuffled to test whether the expression profiles could distinguish normal cervix from cervical cancer. As shown in Fig. 1B, our results demonstrated that, using a permutation test based on Wilcoxon’s rank-sum test, the expression profiles could classify cervical cancers into two broad clinical groups, based on the FIGO\textsuperscript{10} staging system. Most importantly, our results further showed that response to radiotherapy can be predicted by monitoring the expression profiles of some of these cervical cancers, which were stratified into responders and non-responders, based on the retrospective outcome of the treatment.

---

\textsuperscript{10} The abbreviation used is: FIGO, International Federation of Gynecologists and Obstetrics.

\textsuperscript{11} The entire microarray dataset is available for searches at http://www.ncbi.nlm.nih.gov/geo under the data series accession number GSE527.
expression of each gene was compared between normal cervix and cervical cancer, and all of the samples were correctly classified by their expression profiles as either normal or cancer.

Cervical cancer is clinically staged according to the FIGO system into four stages. In our study, due to a limited sample size and insufficient distribution of the samples into the four stage groups, we decided to classify the disease molecularly based on the expression profiles of 16 patients with either stage IB (11 samples) or stage IIB (5 samples) cervical cancer. Our results showed that the two distinct clinical stages could be correctly classified based on their expression profiles (Fig. 1C). Furthermore, we found that only 4 of 100 genes in this analysis overlapped with those identified from the above-mentioned analysis between normal cervix and cervical cancer (data not shown).

We next asked whether the expression profiles could predict treatment response in the cervical cancer patients who had been given radiotherapy. On further stratification of the 26 patients (Table 1), 13 patients who were given radiotherapy as primary treatment were selected for further analysis for radiotherapeutic sensitivity (responder) or resistance (nonresponder). The 13 patients were stratified into two groups based on treatment outcome. Patients who received radiotherapy but died from the disease were considered resistant to treatment, and patients who are alive with no evidence of disease 36 months after treatment were considered sensitive to radiotherapy. Patients who did not respond to radiotherapy had a mean survival time of 22.2 months, whereas the group sensitive to radiotherapy had a mean survival time of 66.5 months. All of the patients in the radiotherapy-resistant group had since died from the disease, whereas patients in the radiotherapy-sensitive group were all still alive at the time this study was conducted.

Supervised clustering analysis was used to analyze our samples of 13 patients. At the beginning of data analysis, one responsive patient and one resistant patient were reserved as the test samples. The other 11 training samples were used to build the model. The gene expression profile from these 11 patients’ tissue and their corresponding clinical treatment outcome records (i.e., responsive and nonresponsive) are the two sets of inputs used in the supervised clustering data analysis. These inputs were analyzed to identify the cause of their differences in treatment outcomes. The optimized result was then used to build the treatment prediction model. The four reserved blinded samples were then used to determine the accuracy of the prediction model. The genes used in the prediction model can correctly classify all 13 samples as either radiotherapy sensitive or radiotherapy resistant based on their gene expression profiles (Fig. 2).

The predictive performance of the model was evaluated by randomly dividing the samples into training and testing samples. Each time, all 13 patients’ microarray data were randomly re-separated into training and testing sets, and the entire training and testing processes were repeated. After repeating the procedure (test sample selection, data analysis, and testing) 100 times, the average prediction accuracy was determined to be 97%.

To further test the validity of our data analysis, our result was compared with randomized data sets by assigning each of the 13 samples arbitrarily to different clinical groups (radiotherapy-sensitive or radiotherapy-resistant), in which a responsive sample could be randomly assigned to either the radiotherapy-sensitive or the radiotherapy-resistant group. These artificially created groups were generated to have an approximately equal number of true positive and negative samples in each group and then subjected to supervised clustering analysis to create the best models to discriminate one group from another. The data were repeatedly created and analyzed >30 times. Our results in Table 2 showed that: (a) self-included and self-excluded results were similar to the true clinical data, however, self-excluded results were much worse than self-included results in randomized test; and (b) self-excluded prediction accuracy of true clinical data (96.0%) was 4.47 SDs better than the mean randomized results (52.2%), thus suggesting that our results did not occur stochastically (P < 0.00001).

To understand the prediction model, the top 100 genes used by the model were extracted for bioinformatics searches and found to represent a wide spectrum of cellular functions (Table 3) including genes with transcription, cell adhesion, membrane and cytoskeletal, and signal transduction functions.

**DISCUSSION**

DNA microarray has been used to profile gene expression in some human clinical cancer specimens (5–9, 23–27). Recent
reports further showed that monitoring breast cancer gene expression yielded signature expression profiles that could predict clinical prognostic outcome in these patients (5, 9). We showed in this study that expression profiling by DNA microarray in cervical cancer could be an important tool for molecular classification of the disease as well as for prediction of treatment responsiveness in cervical cancer, which has not been examined to date. The expression profiles of cervical cancer were distinguishable from those of normal cervix (Fig. 1B). With the expression profiles, we were also able to classify the specimens
correctly in their disease stages according to the FIGO system (Fig. 1C). Most importantly, our results further showed that response to radiotherapy could be predicted based on the signature expression patterns of cervical cancer.

A major caveat in this study is the small sample size. However, this is a pilot study, and efforts are ongoing to continue this research and acquire more samples for future studies to further validate the predictors that we identified in this report. Despite the above-mentioned drawback, our results, nevertheless, showed clearly that expression profiling by DNA microarray in cervical cancer provides a genomics platform for molecular classification and prognostication of treatment response for the disease.

Another potential concern of our prediction of treatment response is that the number of genes on the array (10,692 features) is much larger than the sample size (13 samples). Therefore, it is possible that a small number of samples may show correlation with gene expression stochastically. To address this issue, we tested our analytical methods by comparing the observed clinical data with randomized data. Our results demonstrated that self-excluded accuracy of the observed clinical radiotherapy outcome was better than randomized analysis (Table 2), thus confirming the validity of our analysis.

The signature expression profiles we identified in this report include genes that are differentially expressed between normal cervix and cervical cancer, which encompass a number of genes of diverse biological functions (data not shown). Another signature profile was also identified that differentiated the tumors between stage IB and stage IIB (data not shown). Interestingly, there are few overlaps in the genes between these two signature profiles. It is possible that the genes that are required to distinguish stage IB from stage IIB tumors are not necessarily the same genes that are differentially expressed between normal cervix and cervical cancer. Further research is required to increase the sample size so that tumors of different stages are equally represented in each stage to clearly define the signature profiles distinctly associated with each clinical stage for accurate classification.

Most significantly, our results showed that signature expression profile derived from supervised clustering analysis can be predictive of response to radiotherapy in these patients, thus suggesting that patients whose expression profiles predicted them to be nonresponders should be given alternative treatments. Comprising these predictors is also a large number of genes with diverse physiological functions (Table 3), which include transcription factors, proteins with cytoskeletal, membrane, and cell structural functions. Interestingly, the gene damage-specific DNA binding-protein 1 (28) seemed to be generally expressed in higher levels in patients who did not respond to radiation therapy (Table 3). Clearly, increased expression of one gene in the DNA damage response and repair pathway is probably insufficient to account for resistance to radiotherapy observed in these patients. It is unclear at present whether the other genes may also have roles in conferring radiotherapeutic resistance. Further analysis of a larger group of cervical cancer patient samples treated with radiotherapy will be required to

<table>
<thead>
<tr>
<th>Analysis of clinical data</th>
<th>Analysis of randomly grouped data</th>
</tr>
</thead>
<tbody>
<tr>
<td>Self-included prediction accuracy</td>
<td>Self-excluded prediction accuracy</td>
</tr>
<tr>
<td>Cervical cancer</td>
<td></td>
</tr>
<tr>
<td>99.4</td>
<td>96.0</td>
</tr>
<tr>
<td>Mean value</td>
<td>SD</td>
</tr>
<tr>
<td></td>
<td></td>
</tr>
<tr>
<td>Cervical cancer</td>
<td></td>
</tr>
<tr>
<td>99.4</td>
<td>96.0</td>
</tr>
<tr>
<td>Mean value</td>
<td>SD</td>
</tr>
</tbody>
</table>

**Table 2** Prediction accuracy of array data analysis by supervised clustering methods compared with randomized data sets.

Fig. 2 Discriminant analysis and prediction of response to radiotherapy in cervical cancer. Analysis of array data was performed by supervised pattern-matching methods. The results for the 300 genes that most correctly classified the samples into radiotherapy-resistant and radiotherapy-sensitive groups were further subjected to cluster analysis and are displayed as dendrogram. Mean survival time after radiation treatment for the two groups of patients: radioresistant, 22.2 months; radiosensitive, 66.5 months.
confirm our finding and to further elucidate the direct role of these genes in resistance to radiotherapy, or, indirectly, how their expression may be associated with such resistance. In addition, the utility of these genes as specific markers for predicting responsiveness to radiotherapy will also need to be confirmed in future studies.

The current one-size-fits-all approach in the clinical treatment of cancers does not take into account whether patients may or may not respond to radiotherapy. It is certain that expression profiling by DNA microarray in human cancer will be a genom-ics approach that can provide pharmacogenomics information that can more accurately predict outcome, thereby bringing an end to the practice of trial and error and the one-size-fits-all approach in cancer therapy today. Our results further imply that patients whose tumor expression profiles exhibit a pattern associated with resistance to radiotherapy should be given alternative or supplementary modality of treatment that may result in improved responsiveness or cure, thus personalizing treatment for individuals based on their gene expression patterns.

REFERENCES

Table 3 Genes differentially expressed between patients who responded (sensitive) or failed to respond (resistant) to radiotherapy

| Gene name & category | Accession no.* | Ratio
<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>DNA damage response</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Damage-specific DNA-binding protein 1</td>
<td>AA608557</td>
<td>2.5</td>
</tr>
<tr>
<td>Transcription factors</td>
<td></td>
<td></td>
</tr>
<tr>
<td>T-box 19</td>
<td>AI630980</td>
<td>2.4</td>
</tr>
<tr>
<td>Zinc finger protein 33a (KOX 31)</td>
<td>NS7658</td>
<td>2.1</td>
</tr>
<tr>
<td>Serum response factor</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Nuclear receptor corepressor 1</td>
<td>AA487973</td>
<td>2.7</td>
</tr>
<tr>
<td>Immune modulatory</td>
<td></td>
<td></td>
</tr>
<tr>
<td>IFN-regulatory factor 2</td>
<td>AA416883</td>
<td>3.2</td>
</tr>
<tr>
<td>IFN-α-inducible protein (clone IFI6-16)</td>
<td>AA448478</td>
<td>2.3</td>
</tr>
<tr>
<td>Interleukin 10 receptor β</td>
<td>AA486939</td>
<td>2.3</td>
</tr>
<tr>
<td>Actin-associated and cytoskeletal proteins</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Highly similar to kininogen L</td>
<td>H69834</td>
<td>2.3</td>
</tr>
<tr>
<td>KIAA0220 protein, actin cytoskeleton organization</td>
<td>R91822</td>
<td>2.7</td>
</tr>
<tr>
<td>KIAA0336 gene product, similar to mouse myosin heavy chain</td>
<td>R00035</td>
<td>2.5</td>
</tr>
<tr>
<td>KIAA0514 gene product, proline-rich actin-binding protein family</td>
<td>AI139146</td>
<td>2.1</td>
</tr>
<tr>
<td>KIAA0966 protein, similar to SAC1, suppressor of actin mutations</td>
<td>R69354</td>
<td>2.3</td>
</tr>
<tr>
<td>Ras family of proteins</td>
<td></td>
<td></td>
</tr>
<tr>
<td>N-Ethylmaleimide-sensitive factor</td>
<td>H38086</td>
<td>2.9</td>
</tr>
<tr>
<td>RAB7</td>
<td>AI202933</td>
<td>2.4</td>
</tr>
<tr>
<td>Similar to mouse putative Rho/Rac GEF</td>
<td>AI042352</td>
<td>2.5</td>
</tr>
</tbody>
</table>

* Accession numbers are from GenBank database.


Expression Genomics of Cervical Cancer: Molecular Classification and Prediction of Radiotherapy Response by DNA Microarray

Yick Fu Wong, Zachariah E. Selvanayagam, Nien Wei, et al.