Development of a Highly Specialized cDNA Array for the Study and Diagnosis of Epithelial Ovarian Cancer

G. Peter Sawiris, Cheryl A. Sherman-Baust, Kevin G. Becker, Chris Cheadle, Diane Teichberg, and Patrice J. Morin

ABSTRACT

Ovarian cancer is a major cause of cancer death in women. Unfortunately, the molecular pathways underlying ovarian cancer progression are poorly understood, making the development of novel diagnostic and therapeutic strategies difficult. On the basis of our previous observations obtained from serial analysis of gene expression, we have constructed a specialized cDNA array for the study of ovarian cancer. Small, specialized arrays have several practical advantages and can reveal information that is lost in the “noise” generated by irrelevant genes present in larger arrays. The array, which we named Ovachip, contains 516 cDNAs chosen from our serial analysis of gene expression and cDNA array studies for their relevance to ovarian cancer. The gene expression patterns revealed with the Ovachip are highly reproducible and extremely consistent among the different ovarian specimens tested. This array was extremely sensitive at differentiating ovarian cancer from colon cancer based on expression profiles. The Ovachip revealed clusters of coordinately expressed genes in ovarian cancer. One such cluster, the IGF2 cluster, is particularly striking and includes the insulin-like growth factor II, the cispalatin resistance-associated protein, the checkpoint suppressor 1, the cyclin-dependent kinase 7, and a protein tyrosine phosphatase receptor. We also identified a cluster of down-regulated genes that included the cyclin-dependent kinase 7 and cyclin H. Thus, the Ovachip allowed us to identify previously unidentified clusters of differentially expressed genes that may provide new paradigms for molecular pathways important in ovarian malignancies. Because of the relevance of the arrayed genes, the Ovachip may become a powerful tool for investigators in the field of ovarian cancer and may facilitate progress in understanding the etiology of this disease and in its clinical management.

INTRODUCTION

EOC is the most lethal of gynecological malignancies in the United States and other westernized countries. This is because of difficulties of early diagnosis (1) and frequent resistance of these tumors to chemotherapy (2). In fact, a majority of EOC patients eventually die with drug-resistant cancers. Unfortunately, the specific molecular pathways important for the development of this disease have remained elusive. Alterations in p53, K-Ras, HER-2/neu, c-Myc, and many other genes have been reported in ovarian cancer, but the prevalence of these alterations depends greatly on the cohort and subtype, and do not represent ovarian-specific alterations (3, 4). Gains and losses of various chromosomal regions are common in EOC (5), but the target genes have remained elusive.

Recently, gene expression profiling techniques have provided many insights into the complexity of gene regulation in ovarian cancer.

MATERIALS AND METHODS

Tumors Specimens and Cell Lines. The ovarian tumors were obtained from the Collaborative Human Tissue Network, Gynecological Oncology Group (Children’s Hospital, Columbus, OH). All of the ovarian specimens were diagnosed as stage III or stage IV papillary serous adenocarcinoma. Colorectal carcinoma specimens (stage T3) were a gift of Dr. Bert Vogelstein (Howard Hughes Medical Institute, Baltimore, MD). Cell lines used in this study have been described elsewhere (9). JUSE29EC was generously provided by Dr. Nelly Auersperg (University of British Columbia, Vancouver, British Columbia, Canada; Ref. 12).

Construction of a Specialized Ovarian cDNA Microarray. The selection of genes for the specialized ovarian array (the Ovachip) was made using two datasets. First, genes up-regulated or down-regulated in ovarian cancer were chosen based on our SAGE data (9). Genes that exhibited at least 5-fold up-regulation or down-regulation in at least one ovarian tumor were included. In addition, a preliminary analysis with a 15,000-gene array revealed many genes differentially expressed in ovarian cancer. The most highly differentially expressed genes identified with this array were also included on the Ovachip.

A total of 516 cDNA clones were selected for constructing the Ovachip, and the entire list of genes is available online. The expressed sequence tag clones were obtained from Research Genetics (Huntsville, AL). After amplification, the PCR products were analyzed by agarose gel electrophoresis to determine the quality and specificity of the PCR products. In addition, over 300 of the cDNA clones were sequenced to confirm their identity and their positions on the array. PCR products were denatured and spotted in triplicate onto nylon membrane (Schleicher & Schuell, Keene, NH) using a GMS 417 Arrayer (Affymetrix, Santa Clara, CA). Although a nylon membrane was used for these experiments, the reference set of cDNAs can in principle be used with a variety of platforms. Indeed, we are currently investigating the use glass slides for Ovachip construction. For sample analysis using the Ovachip, total RNA from...
the various tissues was isolated from guanidium isothiocyanate cell lysates by centrifugation on CsCl (9). cDNA synthesis, probe preparation, and hybridizations to the array were done essentially as described (13).

**Real Time RT-PCR.** Total RNA (1 μg) from selected ovarian and colon samples was used to generate cDNA using the Taqman Reverse Transcription Reagents (PE Applied Biosystems, Foster City, CA). Similarly, cDNA was prepared using RNA from OV16, OVT8, and IOSE29EC, and included as control. The SYBR Green I assay and the GeneAmp 5700 Sequence Detection System (PE Applied Biosystems) were used for detecting real-time PCR products as described previously (9). Primers for four candidate genes (TAL-1, EFB, PSMA1 β subunit, and ApoE) and glyceraldehyde-3-phosphate dehydrogenase as control were designed to cross intron-exon boundaries to distinguish PCR products generated from genomic versus cDNA template. The primer sequences are available from the authors on request.

**Computational Analyses.** Results from three independent hybridizations were obtained for each probe and analyzed by Array-Pr Analyzer software version 4.0 (Media Cybernetics, L.P., Des Moines, IA). Expression data for all of the specimens were compared with the nonmalignant sample IOSE29EC as reference. After normalization, the signal intensity values were fed into the Cluster program (Version 3.1), which arranged the genes hierarchically using complete linkage clustering, and then visualized with Treeview Version 1.6.6 (14). Biological expression patterns of the differentially expressed genes were processed for presentation using GeneSpring software (version 4.0.7; Silicon Genetics, San Carlos, CA). All of the MDS analyses were performed using BRB ArrayTools developed previously (15).

**RESULTS**

**Analysis of Gene Expression Using the Ovachip.** The genes arrayed on the Ovachip were chosen because of their importance in EOC according to our previous SAGE study (9). The complete list of the genes present on the Ovachip is available online.3 A typical hybridization of the Ovachip is shown in Fig. 1A. The entire set of 516 spots is present in triplicate on each membrane. The reproducibility within a membrane was remarkable as shown by the scatterplots and high level of correlation between the different replicates (Fig. 1B). The Ovachip was used to analyze a series of 11 stage III-IV ovarian tumors, 4 colon tumors, and various cell lines (Fig. 2). Visual inspection of the clustered data indicates that the ovarian tumors exhibit a distinct gene expression signature when analyzed with the Ovachip. This is evidenced by the presence of a highly correlated node containing all of the ovarian tumors. The ovarian cancer specimens are additionally subdivided into two main branches distinguished mostly by the behavior of a middle cluster of genes that appears to be highly elevated in specimens ov11, ov8, ov2, and ov7 but not in the others. The colon tumors as well as the various cell lines cluster separately from the ovarian tumors and from each other, and exhibit quite heterogeneous gene expression patterns. The similarity between overall patterns of gene expression as measured with the Ovachip can be quantitated using $P_{ov}$. Fig. 3 shows the average $P_{ov}$ values for all of the pair-wise comparisons of the specimens included in this study. It is apparent that when measured with the Ovachip the ovarian tumors are typically very similar to each other with an average $P_{ov}$ between different ovarian tumors of 0.89 (Fig. 3). On the other hand, the ovarian specimens exhibited little similarity to colon tumors (average $P_{ov}$ of only 0.11). Thus, the Ovachip is extremely sensitive at distinguishing colon from ovarian tumors. It is noteworthy that the average $P_{ov}$ between the various colon cancers is only 0.73. It is important to emphasize that these correlation coefficients reflect gene expression as measured by the Ovachip and do not represent global gene expression. The fact that ovarian tumors appear more homogeneous than colorectal tumors is likely to be a consequence of our choice of ovarian-specific genes for the Ovachip. The similarity of gene expression signature of tumors (ovarian or colon) and the various cell lines is essentially nonexistent ($P_{ov}$ close to 0), and the cell line expression patterns are unrelated to the tissue of origin (Fig. 3). This finding is similar to what we observed previously and suggests that cell lines represent a poor model for the study of tissue-specific gene expression (9). Cell lines exhibited some similarity to each other (average $P_{ov}$ = 0.64) probably reflecting the similarities in culture conditions. MDS has also been used to investigate gene expression relationships between different specimens (16, 17). Fig. 4 shows a MDS three-dimensional plot of all of the specimens included in this study. In this depiction of the data, the correlation coefficients are represented in a three-dimensional space, with samples exhibiting similar patterns of gene expression clustering closer together than samples with dissimilar patterns. Clearly, MDS was also extremely sensitive at differentiating ovarian tumors from colorectal tumors and cell lines.

**Validation of the Ovachip.** Although the global expression pattern obtained with this array were reproducible and consistent with our inclusion of ovarian-relevant genes, we wished to verify that individual gene expression measurements were accurate. Levels of expression for FR, SLPI, and IGFBP-2 were high in EOC as measured by the Ovachip (data not shown), which is consistent with our previous SAGE and RT-PCR results (10).

---

[Image: Typical example of the results of a hybridization of an ovarian sample to the Ovachip. The 516 genes constituting the Ovachip are spotted in triplicate on a single membrane as described in “Materials and Methods.” Scatter plots show a high level of reproducibility between the replicates with R values of 0.98 and 0.99.]
Typically, cDNA array data are validated by Northern or RT-PCR analysis of the genes demonstrating the most remarkable patterns of gene expression. For example, genes found to be highly up-regulated or down-regulated compared with a reference sample are often examined more closely using these techniques. Because we believe this may not be a completely fair validation of an array, we also wished to examine genes with unremarkable expression patterns as measured by the Ovachip. We chose the genes TAL-1, TCEB1, PMSA1, and ApoE for their variety in Ovachip expression patterns. Fig. 5 compares the ability of the Ovachip and real-time RT-PCR to measure expression of these four genes in ovarian and colon tumors. It is clear that, whereas the results are not completely concordant, the general trend of gene expression is consistent.
expression as measured by the array is generally verified by the highly accurate RT-PCR method. For example, both TAL-1 and TCEB1 are low in ovarian tumors and increased in colon tumors, with TAL-1 higher than TCEB1 in colon tumors. Similarly, ApoE is elevated both in colon and ovarian tumors, a trend observed with both techniques. These results prove the ability of the Ovachip to accurately measure levels of individual transcripts in cancer.

**Identification of Differentially Expressed Genes.** Using GeneSpring to analyze the hybridization data, 25 genes were found to be significantly up-regulated in ovarian cancer compared with nonmalignant ovarian surface epithelial cells, whereas 11 genes were found down-regulated (Table 1). A total of 26 genes were also up-regulated in colon cancer specimens, but only 1 gene was commonly elevated in ovarian and colon tumors. Of the 11 genes that were down-regulated in ovarian cancer, only 1 was also decreased in colon tumors (Table 1). Again, this is a likely consequence of our choice of genes relevant to ovarian tumorogenesis for inclusion on the Ovachip. Interestingly, BRCA1 was found to be decreased using the Ovachip. BRCA1 has been reported to be down-regulated at the mRNA level in the majority of sporadic ovarian cancers (18) by mechanisms that may involve promoter hypermethylation (19).

**Coordinately Expressed Clusters of Genes.** Many interesting clusters of genes with similar gene expression patterns can be identified from our study. Probably the most striking and interesting cluster was the IGF2 cluster, named after the gene with the highest level of up-regulation in ovarian cancer, IGF-II (Fig. 6). IGF-II is an imprinted gene expressed in the normal ovary and has been reported to be up-regulated in ovarian cancer, possibly through LOI (20, 21). The IGF2 cluster included 7 genes coordinately elevated in addition to IGF-II, including TGFβ-2, 2 genes previously shown to be elevated in ovarian cancer (Ref. 22; Table 2). These genes may all be related to an important function in ovarian cancer development.

Another interesting cluster was composed of genes found to be coordinately down-regulated specifically in the ovarian carcinomas (Fig. 6). This cluster was named the CAK cluster because it included the cdk7 as well as its regulatory subunit cyclin H (Fig. 6; Table 2). The CAK is a trimer consisting of cdk7, cyclin H, and MAT1 (not on the Ovachip) and phosphorylates a conserved threonine residue on cdkks that control cell cycle progression (23). The consequences of down-regulation of this activity are unclear but would clearly affect cell cycle control and possibly response to DNA damage and genomic integrity. The CAK cluster comprises a total of 10 genes, including the FGR oncogenes, the selenium binding protein 1 and filamin A (Table 2).

**DISCUSSION**

Ovarian cancer is a poorly understood disease at the molecular level. Despite a recent series of exciting advances in the field leading to a better characterization of ovarian-specific gene expression signatures and potential new tumor suppressor genes, the molecular pathways crucial for initiation and progression of EOC remain mostly unknown. In fact, the exact nature of the tissue of origin is still a matter of debate. Although EOC is generally thought to arise from the layer of mesothelial cells covering the ovaries, a convincing argument can be made that the actual precursor is a different tissue altogether, namely the secondary Mullerian system (22). Clearly, the elucidation of these puzzling questions will require coordinated efforts as well as novel tools and reagents. In this report, we have constructed a specialized cDNA array for the study of EOC. Although our previous SAGE study of ovarian malignancy revealed major gene expression patterns and many differentially expressed genes (9), SAGE does not allow high throughput analysis of many tumors, which may be required to identify relevant pathways. A small, focused array has several advantages compared with the large arrays currently widely used for gene expression analysis. First, a small, specialized chip includes only genes relevant to the biological process of interest, eliminating noise and allowing the identification of relationships that may otherwise be lost (11, 23). Secondly, it allows a larger level of...
replication for the same RNA input. For example, the Ovachip is entirely contained in triplicate on a single membrane (see Fig. 1) allowing three independent measurements from a single hybridization and increasing statistical power (13). Finally, the small chip greatly simplifies data analysis, and reduces time and computer power required to obtain expression information in final form. Again, all of these advantages are the results of the exclusion of genes that are not involved in the process under study. The construction of focused array requires previous knowledge of gene expression profiles for the tissue of interest. Considering the expanding number and size of public and commercial gene expression databases, this problem should not represent a major hurdle.

All of the ovarian tumors have similar Ovachip signatures, which are clearly different from the signatures obtained with colon tumors (Fig. 2). This is likely the result of our selection of genes relevant to EOC for selective inclusion on the array. This suggests that this array would be extremely sensitive at identifying EOC when histopathologically parameters may be ambiguous. Because hybridization can be performed with small amounts of starting RNA, the Ovachip could be used to distinguish ambiguous clinical specimens. It is also possible that small variations of this known pattern of gene expression may have relevant biological significance. For example, we have found that a low-malignant potential ovarian tumor exhibits a significantly different Ovachip signature compared with the typical serous ovarian carcinomas (data not shown). Thus, the Ovachip may be useful in identifying small but significant gene expression differences that are relevant to the clinical behavior of these tumors such as growth and invasive capabilities. In this context, it is also clear that the Ovachip may be used advantageously in ovarian cancer research. Indeed, as described above, this array may allow the identification of subtle differences that larger arrays may miss because of the presence of many irrelevant genes that increase the noise during data analysis (11). In our hands, small arrays have provided more accurate and reproducible gene expression data, which could be validated using real-time RT-PCR (Fig. 5). The genes that exhibit altered expression levels in ovarian cancer (Table 1) may all represent targets for diagnosis or mechanism-based therapy.

Using the Ovachip, we have found many genes differentially expressed in ovarian cancer (Table 1), including two major clusters of coordinately regulated genes that may be important in ovarian cancer (Fig. 6; Table 2). The IGF2 cluster includes seven genes apparently functionally unrelated. Relaxation of IGF2 LOI has been suggested to be at least partly responsible for its abnormal expression in ovarian cancer, and it will be interesting to see whether the other genes of the IGF2 cluster also exhibit abnormal methylation patterns of their promoters. It is intriguing that this cluster includes the CRA and several genes involved in cell cycle control. The IGF2 expression signature may be related to the modulation of cell cycle checkpoints and acquisition of drug resistance. In addition, this cluster includes FLT1, one of the VEGF receptors (26). FLT1 expression has been reported in tumor cells, endothelial cells, and stromal fibroblasts of various tumors and may be part of an important autocrine loop (27). Considering the importance of angiogenesis to ascites development and ovarian cancer metastasis, the identification of FLT1 as part of this clusters suggests that these genes may be related to advanced disease.

A cluster of highly coordinately expressed down-regulated genes
was also identified (Fig. 6). The main feature of this cluster is the presence of cdk7 and its regulatory subunit cyclin H, comprising two of the three necessary components that make up the CAK. The third component of this complex, MAT1 [also known as CAK assembly factor and MNAT1 (23)], was not present on our array. It is unclear why these genes would be down-regulated in ovarian cancer as they have previously been found to be moderately up-regulated in other cancers (28, 29). Changes in expression of these genes would clearly affect cell cycle control and progression and may influence important checkpoints in ovarian cancer. We are currently investigating genes from both the IGF2 cluster and the CAK cluster for their role in ovarian tumorigenesis.

In summary, we have constructed and tested a specialized ovarian cancer chip that we named the Ovachip. The advantages of this small, specialized array are numerous and enabled us to identify novel clusters of genes differentially expressed in ovarian cancer. It is our hope that the Ovachip will facilitate progress in understanding the etiology of this disease and in its clinical management.

ACKNOWLEDGMENTS

We thank Drs. Ellen Pizer, Leticia Rangel, Rachana Agarwal, and Ashani Weeraratna, for helpful comments on the manuscript.

REFERENCES

Development of a Highly Specialized cDNA Array for the Study and Diagnosis of Epithelial Ovarian Cancer

G. Peter Sawiris, Cheryl A. Sherman-Baust, Kevin G. Becker, et al.