Profiling Studies in Ovarian Cancer: A Review

RUDOLF S. N. FEHRMANN, a–c XIANG-YI LI, a,b ATE G. J. VAN DER ZEE, a STEVEN DE JONG, b GERARD J. TE MEERMAN, c ELISABETH G. E. DE VRIES, b ANNE P. G. CRINS a,b

Departments of aGynecological Oncology, bMedical Oncology, and cGenetics, University Medical Center Groningen, Groningen, The Netherlands

Key Words. Gene-expression profiling • Microarray analysis • Review • Ovarian neoplasms • Methods

LEARNING OBJECTIVES

After completing this course, the reader will be able to:

1. Discuss the present status of treating ovarian cancer.
2. Summarize the research in microarray gene profiling.
3. Describe the weakness in microarray gene profile research.
4. Identify the potential applications of gene profiling.

ABSTRACT

Ovarian cancer is a heterogeneous disease with respect to histopathology, molecular biology, and clinical outcome. In advanced stages, surgery and chemotherapy result in an approximately 25% overall 5-year survival rate, pointing to a strong need to identify subgroups of patients that may benefit from targeted innovative molecular therapy. This review summarizes: (a) microarray research identifying gene-expression profiles in ovarian cancer; (b) the methodological flaws in the available microarray studies; and (c) applications of pathway analysis to define new molecular subgroups. Microarray technology now permits the analysis of expression levels of thousands of genes. So far seven studies have aimed to identify a genetic profile that can predict survival/clinical outcome and/or response to platinum-based therapy. To date, the clinical evidence of prognostic microarray studies has only reached the level of small retrospective studies, and there are other issues that may explain the nonreproducibility among the reported prognostic profiles, such as overfitting, technical platform differences, and accuracy of measurements. We consider pathway analysis a promising new strategy. The accumulation of small differential expressions within a meaningful molecular regulatory network might lead to a critical threshold level, resulting in ovarian cancer. Microarray technologies have already provided valuable expression data for classifying ovarian cancer and the first clues about which molecular changes in ovarian cancer could be exploited in new treatment strategies. Further improvements in technology as well as in study design, combined with pathway analysis, will allow us to detect even more subtle tumor expression differences among subgroups of ovarian cancer patients. The Oncologist 2007;12:960–966

Disclosure of potential conflicts of interest is found at the end of this article.
INTRODUCTION
Ovarian carcinoma is the leading cause of death from gynecological malignancies in the western world [1]. Its high death rate is a result of the fact that most patients (>60%) are diagnosed at an advanced stage of disease [2]. Debulking surgery followed by platinum-based chemotherapy schemes is considered standard care for these patients [3]. However, despite an initial response rate of 65%–80% to first-line chemotherapy, most ovarian carcinomas relapse. Acquired resistance to further chemotherapy is generally responsible for treatment failure, resulting in an overall 5-year survival rate of only 10%–30% for late-stage ovarian cancer [2]. Classic clinicopathological factors, such as age, stage, residual tumor after first laparotomy, differentiation grade, histiotype, and response to chemotherapy, are important prognostic markers [3], but it is not possible to select optimal chemotherapy for an individual patient based on these factors.

Ovarian cancer is thought to result from an accumulation of genetic changes [4]. When these changes are better understood, it should be possible to identify new subgroups of ovarian cancer patients. This may identify patients benefiting in the future from novel first-line experimental approaches when having a high chance of resistance to standard first-line chemotherapy. Furthermore, insights might be provided into tumor aggressiveness and drug resistance, yielding potential new molecular targets for therapy. However, to date, there exist no such therapies that are proven and used in ovarian cancer, and it is still speculative that there will be [5, 6]. The speed of elucidating this insight is hampered by the relatively low incidence of and the substantial molecular heterogeneity among ovarian cancer subtypes [2, 4]. The prognostic and/or predictive value of several single molecular markers has been evaluated in ovarian cancer [5, 7]. To date, gene-expression patterns have also been used to classify ovarian carcinomas into clinically relevant subtypes (Fig. 1). However, none of these single molecular markers or predictor sets of genes have been clinically implemented, mainly because their reliability and validity have not yet been well established.

In this review we summarize: (a) microarray research identifying gene-expression profiles in ovarian cancer; (b) the methodological flaws in ovarian cancer microarray studies; and (c) applications of pathway analysis to define new molecular subgroups.

MICROARRAY STUDIES IN OVARIAN CANCER
Studies on the role of single molecular markers have shown ovarian carcinogenesis to be a complex process, multifactorial in nature, and associated with genetic abnormalities in multiple gene families. Microarray technology now permits analysis of expression levels of thousands of genes and is widely used to identify prognostic gene-expression profiles for all sorts of cancer. Several studies using DNA microarrays to determine gene-expression profiles in ovarian cancer cell lines or ovarian carcinomas have been reported. The specific aims of the separate studies are listed in online supplementary Table 1. Most have tried to identify diagnostic markers by comparing ovarian cancers or ovarian cancer cell lines with normal ovarian epithelium. Ovarian cancer cell line models were mostly used to identify genes related to drug resistance. Recently, we published a study on drug resistance in four ovarian cancer cell lines, which differed only in their degree of cisplatin resistance [8]. We identified 315 genes associated with cisplatin resistance. Many studies profiling ovarian tumors have searched for genes associated with classical clinical prognostic factors, or drug resistance, to identify molecular clues to prognosis and treatment efficacy. So far, seven microarray studies have looked at prognostic and/or predictive genes and gene profiles for predicting individual disease outcome in advanced ovarian cancer [9–15]. Three studies aimed to predict survival/clinical outcome. Spentzos et al. [9] used 34 tumor samples to identify a 115-gene profile. When applied to a...
validation set of 34 samples, the profile distinguished patients with a favorable overall survival (median, 30 months versus not yet reached; log-rank \( p = .004 \)). The profile consists of 70 genes overexpressed in the unfavorable outcome group and 45 genes underexpressed in the favorable group.

Berchuck et al. [10] tried to identify gene-expression patterns using classification tree and linear discriminant models that reflected patient survival by profiling tumors from 54 advanced-stage patients. Survival was <3 years in 30 cases and >7 years in 24 cases. They presented 26 genes used in tree and linear discriminant models of short-term versus long-term survival, including three genes involved in classification trees and 23 genes appearing in 15 or more leave-one-out linear discriminant models. They were also able to confirm that the identified set of genes associated with survival in their linear discriminant models held their prognostic value in the set of independent tumors profiled by Spentzos et al. [9].

Instead of studying overall survival using expression profiles, Hartmann et al. [11] tried to predict which women were at risk for early (<21 months) versus late (>21 months) relapse after initial chemotherapy. They identified a 14-gene profile predictive for early recurrence by profiling 79 advanced-stage, high-grade tumors. In addition to these three prognostic studies, four other studies tried to identify a profile predictive of response to platinum-based therapy. Spentzos et al. [12] discovered a 93-gene profile predictive of pathological complete response to chemotherapy by profiling 24 tumors from patients who had undergone second-look surgery. This profile was also able to distinguish between cancers in a validation series of patients with favorable versus unfavorable overall survival. Their prognostic 115-gene profile for overall survival, described above, was determined from the same data but, surprisingly, showed no gene overlap with the 93-gene profile associated with predicting response to chemotherapy. Combining both profiles yielded better prognostic discrimination than either profile alone. Hellemann et al. [13] identified a nine-gene profile using microarray and reverse transcription-polymerase chain reaction (RT-PCR) that was able to classify tumors from 24 patients according to their response to platinum-based therapy. Using RT-PCR, this nine-gene set predicted platinum resistance in an independent validation set of 72 patients. However, in contrast to Spentzos et al. [12], who assessed response pathologically, they assessed a clinical response according to World Health Organization criteria. They further defined nonresponders as patients whose tumors showed progression, whereas Spentzos et al. [12] considered all the patients whose tumors showed no pathological complete response as nonresponders. The differences in definition of response to chemotherapy may be one reason for the low number of genes shared by the two profiles. In a preliminary study, Crijns et al. [14] labeled samples from 157 patients with advanced ovarian cancer with Cy5 and Cy3. They identified a 53-gene Cy5 predictor set and a 76-gene Cy3 predictor set for response to chemotherapy. The two predictor sets had 32 genes in common. Misclassification rates for the original Cy5 and Cy3 datasets were validated with a phenotype randomization test and were significantly better than random, confirming the relevance of the new predictor sets for response to chemotherapy. Dressman et al. [15] developed an expression model with 1,727 genes that predicts response to platinum-based therapy using a training set of 83 advanced-stage serous ovarian cancers and tested it on a 36-sample validation set. A prediction accuracy of 78% in the validation set was achieved. In parallel, expression signatures defining the status of oncogenic signaling pathways were evaluated. In patients with platinum-resistant disease they identified expression signatures consistent with activation of Src and Rb/E2F pathways. Several genes reported in the prognostic and predictive profile studies described above (such as those encoding T-lymphocyte maturation–associated protein, heat shock protein 27, SRD5A1, lysophospholipase II, plasminogen activator inhibitor type 1, BAX, platelet-derived growth factor receptor, fibronectin, connective tissue growth factor) are known to affect the malignant phenotype [16–25]. This suggests that gene-expression profiling can indeed be used to define prognosis and yield mechanistic insights into ovarian carcinogenesis and chemoresistance.

**Methodological Difficulties of Prognostic Ovarian Cancer Microarray Studies**

So far, the clinical evidence of prognostic microarray studies (as well as prognostic studies on single markers) has only reached level 3 (small retrospective studies) on a scale of 5 (low) to 1 (high) [26]. These studies therefore encounter the same methodological difficulties as prognostic studies on single molecular markers. However, in ovarian cancer microarray studies, there are other reasons that may explain the limited overlap between the prognostic profiles.

**Overfitting**

In ovarian cancer microarray studies, the problem of sample size becomes even more important, as overfitting, that is, finding a discriminatory pattern by chance, may occur when large numbers of potential predictors are used to discriminate among a small number of outcome events [27, 28]. A complex multivariate analysis may very well result in a perfect prognostic profile for the dataset at hand, but perform poorly on independent datasets. Because the num-
bers of samples included in the seven prognostic and/or predictive microarray studies are all of limited size (range, 24–157) in comparison with the numbers of features on the microarrays, they are all at risk for overfitting. The Hartmann et al. [11] study, identifying a 14-gene profile, is a striking example where overfitting might have occurred. They proposed that their final model classified 40 of 51 samples correctly (78%; \( p < .001 \)). This high classification accuracy was probably achieved by using all 51 samples to fit the model. They also reported a classification accuracy of 86% \( (p < .05) \) for their model on an independent test set of 28 samples, but we feel this should be interpreted with caution. The authors called this test set independent, but it was actually used to select the best-performing prediction model out of 100 and thus introduced additional degrees of freedom for fitting a model.

**Multiple Microarray Platforms**

Multiple array platforms vary in the kind of probes they use (short-oligonucleotide, long-oligonucleotide, cDNA, etc.), the production method (in situ polymerization, spotting, microbeads, etc.), and the labeling method. For microarrays to be reliable tools, they must possess probes that hybridize with high sensitivity and specificity. The seven studies to find prognostic and/or predictive gene profiles for predicting individual disease outcome in advanced ovarian cancer used different types of microarray platforms. One major advantage of oligonucleotide arrays over cDNA arrays is that there is less possibility of probe mix-ups [29, 30]. Oligonucleotide arrays can be synthesized in situ or synthesized ex situ and attached to a derivative substrate. Affymetrix (Affymetrix, Santa Clara, CA) and Agilent (Agilent Technologies, Palo Alto, CA) are both commercial platforms that rely on the in situ synthesis of probes. Affymetrix oligonucleotide arrays consist of 25-mer probes contrasting with the longer 60-mer probes employed by Agilent. Short oligonucleotides should, in theory, provide the greatest discrimination between related sequences, but the longer 60-mer probes are more tolerant of sequencing mismatches and thus more suitable for the analysis of highly polymorphic regions. Furthermore, when 60-mers instead of 25-mers are used, the sensitivity is enhanced, partly because of the larger area available for hybridization. Only recently, a novel array platform was developed on the basis of randomly assembled arrays of beads in wells [31]. Following random assembly, the location and identity of each bead, bearing a gene-specific probe sequence, were determined via a sequential decoding process. Advantages of this approach are the dense packing that can be achieved and that there are multiple copies of each sequence-specific bead in an array. Randomness minimizes the effects of spatially localized artifacts and redundancy increases the precision and robustness of measurements. These factors combine to increase accuracy. Another advantage is that only 100 ng of input RNA is necessary.

The problem of nonreproducibility of prognostic profiles related to the technical differences described above has been addressed by several studies, which drew different conclusions [32–45]. Some studies found that the concordance of results across expression analysis platforms was low in contrast to other studies that concluded the opposite. A study comparing the Affymetrix and Illumina (Illumina Inc., San Diego, CA) arrays showed that agreement was very high, particularly for genes that were predicted to be differentially expressed between two tissue types [45]. That study further demonstrated that the agreement was strongly correlated with the level of gene expression and concordance was superior when probes on both platforms could be identified as being likely to target the same set of transcripts of a given gene.

**Signal-to-Noise Ratio**

The intensities observed from ovarian cancer microarray experiments do not simply represent the relative gene-expression level. They are composed of signal (gene-expression level) and experimental noise [46]. Each stage in the course of a microarray experiment is a source of experimental noise, for example, sample collection, sample labeling and hybridization, and the analysis and explanation of microarray data [29]. The number of genes expected to be differentially expressed within a single tissue type between two classes of interest is probably small, and the differences in expression may not be large in relation to experimental noise [47]. Because the samples used in the seven prognostic ovarian cancer profile studies described above were mainly composed of the serous histological type, it may be difficult to identify the relevant genes from among the numerous noise genes. As such, confirmation and validation of genome-wide microarray results using additional methods such as real-time quantitative RT-PCR remain critical steps.

**APPLYING PATHWAY ANALYSIS TO DEFINE NEW MOLECULAR SUBGROUPS**

So far, profiling of ovarian cancer tissue has enabled classification of discrete subsets of ovarian cancers to predict outcome or response to chemotherapy. Ultimately, one would like to use expression profiles to better understand the biological pathways involved in oncogenesis. Several novel strategies have recently been proposed to identify relevant pathways linked to expression signatures. Bild et al. [48] experimentally generated expression signatures that
reflect the activation of various oncogenic signaling pathways. Initially, they brought primary human mammary epithelial cell cultures (HMECs) to quiescence. Subsequently, the HMECs were infected with adenovirus expressing either human c-Myc, activated H-Ras, human c-Src, human E2F3, or β-catenin. Quiescent cells were used because many of the pathways of interest, such as apoptotic or proliferative pathways, would be inactive. The activation states of the five defined pathway signatures were simultaneously analyzed in advanced-stage ovarian carcinomas, among other tumor types. The ovarian cancers exhibited a dominant pattern of β-catenin and Src deregulation, either elevated or diminished. Notably, the coderegulation of Src and β-catenin identified a population of patients with very poor survival compared with other pathway clusters. They also showed that signatures for Ras and Src pathway activation accurately predicted the in vitro sensitivity of a broad range of human tumor cell lines to drugs targeting the mutational activated versions of these proteins. Potti et al. [49] used in vitro drug sensitivity data obtained in cell lines coupled with Affymetrix microarray data to develop gene-expression signatures that predict sensitivity to chemotherapeutic drugs (docetaxel, topotecan, doxorubicin, etoposide, 5-fluorouracil, paclitaxel, and cyclophosphamide). They integrated the chemotherapy response signatures with signatures of oncogenic pathway deregulation to identify new therapeutic strategies. Regression analysis showed a relationship between phosphatidylinositol 3-OH (PI3)-kinase pathway deregulation and docetaxel resistance, indicating a potential relevance of future use of a PI3-kinase inhibitor in this docetaxel-resistant subgroup. Several other studies generating expression signatures representing deregulated pathways made use of mouse tumor models or cells stably transformed by the loss of a tumor-suppressor gene or the expression of an oncogene [50–54]. These strategies have the advantage of being more reflective of the oncogenic phenotype compared with the acute expression of an oncogene with an adenovirus. On the other hand, the limitation of steady-state expression is the extent to which the steady-state obscures the direct effects of a single pathway and begins to merge several pathways. In contrast to experimentally derived signatures Subramanian et al. [54] describe a computational approach called gene set enrichment analysis. Genes that are differentially expressed between two phenotypic classes are compared according to rank against a large collection of pathway data and other categorizations (functional groupings). This is done to determine enrichment for one or more of these functional groupings. Functional groupings are made according to known pathway databases such as Kyoto Encyclopedia of Genes and Genomes (KEGG), cytogenetic loci, or other relevant functional groupings data. A drawback of this approach is the reliability of functional groups defined by these databases and the fragmented knowledge about them. For ovarian cancer, a similar approach was applied by Donninger et al. [55] using PathwayAssist (Iobion Informatics, LLC, La Jolla, CA), a commercially available program. Based on the expression results and PathwayAssist, a signaling pathway associated with tumor cell migration, spread, and invasion was identified as being activated in advanced ovarian cancer. Mougeot et al. [56] also performed microarray analysis with the aim of identifying molecular interaction networks defining tumor growth. Hierarchical clustering identified groups of tissues reflective of low malignant potential/early cancer onset and possible precancerous stages involving small-molecule, cytokine, and/or hormone-dependent feedback responses specific to the pelvic reproductive system and a priori–initiated tumor suppression mechanisms. They also established a protein/protein interaction database using the Biomolecular Interaction Network Database and Cytoscape 2.1 [57, 58]. This associated database was used to build and visualize molecular regulatory networks integrating small but significant changes in gene expression.

DISCUSSION
In ovarian cancer, the use and potential of conventional chemotherapeutic drugs seem to have reached a ceiling. Chemotherapeutic drugs used differently can still clearly improve survival. Armstrong et al. [59] recently showed that i.p. cisplatin and paclitaxel resulted in a higher survival rate, 48% versus 38%, at 60 months in stage III patients in comparison with i.v. paclitaxel. Although this is a clear improvement in this group of patients over i.v. treatment, it also shows that there is still ample room for improvement. In the future we would like to have more patient-tailored therapy, in which molecularly different subtypes are treated with specific drugs.

Microarray technologies have already provided valuable expression data for classifying ovarian cancers based on gene profiling. However, both at the level of study performance as well as at the level of technology used, there is room for further improvement. The low signal-to-noise ratio often seen in microarray experiments can generally be improved by synthesizing multiple copies of the (same) probe. Chou et al. [60] showed that accurate gene-expression measurement can be achieved with multiple probes per gene, and fewer probes are needed if longer probes rather than shorter probes are used. To further reduce error, standards for reporting microarray data have been established: minimum information about a microarray experiment (MIAME) [61]. This standard contains information required to easily interpret the microarray data so that the results de-
rived from its analysis can be independently verified. McShane et al. [62] recently formulated the REporting recommendations for tumor MARKer prognostic studies (REMARK) to provide relevant information on how to perform good prognostic studies and to encourage transparent and complete reporting so that the usefulness of the data can be judged by outsiders. As concluded in all prognostic microarray studies, the prognostic value of gene profiling in ovarian cancer must be further evaluated in additional, large prospective studies to reach levels 2 or 1 for clinical evidence [26].

However, microarray analyses or studies, mainly providing lists of significant genes with p-values, are insufficient to fully understand the etiology of ovarian cancer. For example, a single gene that is significantly up- or down-regulated does not necessarily have any physiological impact on the growth of ovarian cancer. Physiological effects will depend on the relative distribution of the gene product and its interaction with other genes. If the progression of ovarian cancer depends on the accumulation of small significant changes in expression, it will be necessary to place these significant genes in gene regulatory networks to reveal whether molecular pathways are deregulated. Such an analysis should provide insight into which molecular pathways contribute to the development and clinical outcome of ovarian cancer. Moreover, it may indicate which components of these associated pathways can be used as therapeutic targets. In the future, it will be necessary to extend and correct our knowledge of molecular regulatory networks to be able to identify the molecular pathways involved in tumor progression with greater reliability.

ACKNOWLEDGMENTS

This review was supported by grant UMCG 2002–2681 from the Dutch Cancer Society.

DISCLOSURE OF POTENTIAL CONFLICTS OF INTEREST

The authors indicate no potential conflicts of interest.

REFERENCES


