Microarray Analysis of Early Stage Serous Ovarian Cancers Shows Profiles Predictive of Favorable Outcome

Andrew Berchuck,1,2 Edwin S. Iversen,3 Jingqin Luo,3 Jennifer P. Clarke,4 Hisani Horne,5 Douglas A. Levine,7 Jeff Boyd,7 Miguel A. Alonso,8 Angeles Alvarez Secord,1 Marcus Q. Bernardini,1 Jason C. Barnett,1 Todd Boren,8 Susan K. Murphy,1 Holly K. Dressman,4 Jeffrey R. Marks,6 and Johnathan M. Lancaster9

Abstract

Purpose: Although few women with advanced serous ovarian cancer are cured, detection of the disease at an early stage is associated with a much higher likelihood of survival. We previously used gene expression array analysis to distinguish subsets of advanced cancers based on disease outcome. In the present study, we report on gene expression of early-stage cancers and validate our prognostic model for advanced-stage cancers.

Experimental Design: Frozen specimens from 39 stage I/II, 42 stage III/IV, and 20 low malignant potential cancers were obtained from four different sites. A linear discriminant model was used to predict survival based upon array data.

Results: We validated the late-stage survival model and show that three of the most differentially expressed genes continue to be predictive of outcome. Most early-stage cancers (38 of 39 invasive, 15 of 20 low malignant potential) were classified as long-term survivors (median probabilities 0.97 and 0.86). MAL, the most differentially expressed gene, was further validated at the protein level and found to be an independent predictor of poor survival in an unselected group of advanced serous cancers ($P = 0.0004$).

Conclusions: These data suggest that serous ovarian cancers detected at an early stage generally have a favorable underlying biology similar to advanced-stage cases that are long-term survivors. Conversely, most late-stage ovarian cancers seem to have a more virulent biology. This insight suggests that if screening approaches are to succeed it will be necessary to develop approaches that are able to detect these virulent cancers at an early stage.

Ovarian cancer is highly lethal because a majority are high-grade invasive serous cancers that have metastasized widely in the abdomen prior to diagnosis. Advances in cytoreductive surgery and chemotherapy have improved median survival, but few are cured (1–3). Less than 5% of invasive serous ovarian cancers are confined to the ovary at diagnosis, but these cases usually are cured surgically (4). In view of this, it is thought that screening and early detection approaches might improve survival. An alternative view is that this would have little impact on ovarian cancer mortality because cancers with the most favorable biology will be preferentially detected by screening. In this regard, screening detects some ovarian cancers (5), but a reduction in mortality has not yet been shown.

Microarray analysis is a powerful genomic tool with the potential to elucidate relationships between clinical features of cancers and the underlying causative biological alterations. Patterns of gene expression have been identified with microarrays that correlate with ovarian cancer features, including tumor behavior (invasive versus low malignant potential; refs. 6–8), histologic type (9, 10), stage (11), and survival (12–14). The median survival of women with advanced ovarian cancer is about three to four years and only a small minority are long-term survivors or are permanently cured. We previously used Affymetrix U133A microarrays to develop a linear discriminant model that distinguishes between advanced-stage serous ovarian cancers from short-term (<3 years) and long-term (>7 years) survivors with 85% accuracy (12). Eleven early-stage invasive serous ovarian cancers were also predicted to be long-term survivors using this gene expression model of outcome. This suggested that cancers currently detected at an early stage are not representative of the entire spectrum of the disease, and that most ovarian cancers have a more virulent underlying biology and are likely much less amenable to early detection.
In the present study, we have validated the microarray-based model for long- versus short-term survival and confirmed that early-stage invasive and borderline serous ovarian cancers have gene expression profiles predictive of favorable outcome. In addition, our group has developed microarray-based signatures reflective of overexpression of the mcy, E2F1, E2F3, ras, src, B-catenin, AKT, PI3K and p63 oncogenic pathways (15, 16). Analyses of these oncogenic pathway signatures in the spectrum of invasive and low malignant potential serous ovarian cancers also provides insight into their behavior and clinical outcome.

**Translational Relevance**

Serous ovarian cancers account for a majority of cases, are usually diagnosed at an advanced stage, and cause most deaths. Because cases confined to the ovary are highly curable, a screening test that facilitates early detection might improve survival; this is based on the assumption, however, that the underlying biology and progression rate of most ovarian cancers are relatively similar. We have used microarrays to identify patterns of gene expression predictive of long-term versus short-term survival in ovarian cancer. We found that ovarian cancers presently detected at an early stage represent a subset with favorable underlying biology relative to the majority of ovarian cancers that present at an advanced stage. Therefore, screening strategies based upon the detection of stage I and II cancers may preferentially detect less virulent cancers. This casts doubt on whether early detection strategies for ovarian cancer can impact mortality.

**Patients and Methods**

**Patient samples.** Microarray analysis was done on 101 serous ovarian cancers collected from the primary ovarian site (42 advanced stage III/IV, 39 early stage I/II, 20 borderline) that were snap-frozen at initial surgery prior to chemotherapy under the auspices of Institutional Review Board–approved tissue collection protocols. This included 43 women treated at Duke University Medical Center (20 advanced, 3 early, 20 borderline), 22 advanced invasive cancers treated at H. Lee Moffitt Cancer Center in Tampa, 14 early invasive cancers treated at Memorial Sloan-Kettering Cancer Center, and 22 early invasive cancers from the Gynecologic Oncology Group Tumor Bank (Table 1). All of the surgeries were done by gynecologic oncologists. An additional 65 invasive serous ovarian cancers from our initial paper that reported a microarray-based predictor of long versus short survival (12) were also included in the present article in the analyses of oncogenic pathway signatures.

**Gene expression analysis.** Frozen tissue samples were embedded in optimal cutting temperature medium, and sections were cut and mounted on slides to assure that at least 60% of the cellular content comprised cancer cells. Microarray analyses were done using Affymetrix U133A GeneChip arrays as previously described (12). Analyses of the 42 advanced-stage cancers were done at H. Lee Moffitt Cancer Center (clinical characteristics of these cases are presented in Supplementary Table S1) and the other cases were analyzed at Duke University Medical Center. Genes whose correlation with survival was >0.4 in absolute value had been used previously to predict long (>7 y) versus short (<3 y) survival in 54 advanced-stage serous ovarian cancers from our institution using leave-one-out analyses (12). In the leave-one-out analyses, the discriminant functions were constructed from subsets of the 200 probe sets with the highest marginal association to survival and reflected models allowing for simple conditional dependencies among gene expression measures given survival category. The best multivariate linear discriminant functions had correctly classified 19 of 24 (79.2%) long-term survivors and 27 of 30 (90%) short-term survivors, achieving 85.2% out-of-sample classification accuracy (12). The best linear discriminant model based on all 54 advanced-stage samples included 7 probe sets (Table 2) and was used for out-of-sample predictions of long versus short survival in 11 early-stage samples (12). This 7-gene model was also used to generate predictions of long- versus short-term survival in cancers analyzed in the present study.

The linear discriminant model is retrospective: it is of the form Pr(X|Y) where X is the expression data and Y is binary survival category. In order to predict survival given expression data, i.e. Pr(Y|X), using the linear discriminant model, it is necessary to have an estimate of Pr(Y), the frequency of long-term survival in the target population. Because the evaluation set is a convenience sample enriched for long-term survivors, we generated predictions assuming Pr(Y) was equal to the observed frequency in the training data set (an overestimate) and when Pr(Y) was assumed to equal the empirical frequency in the evaluation set. We report the latter for predictions of the advanced-stage cases as, in practice, the marginal distribution of survival in a target clinical population will be known with some confidence. We predicted absolute risk for the early-stage and borderline cases using the late-stage survival model solely to determine if these cases share the late-stage survival signature. For these cases, we assumed Pr(Y) = 0.5. This corresponds to a uniform prior on whether or not a sample is a long-term survivor.

Batch-to-batch variability in expression array data can be significant (17, 18). We tested for differences between the data reported in our original publication (12) and the current sample of cancers by calculating T statistics associated with the 2-sample test for a batch location effect implemented in the dChip software package where it is referred to as the batch effect using "standardized separators" (19). The method employs a gene-by-gene location and scale adjustment. In doing this analysis, we focused on the 200 genes used to construct the linear discriminant model used for out-of-sample predictions and allowed for batch-specific variances. We found that 170 of the 200 probe sets showed significant (at the 5% level) differential expression between the initial and the new batches of early-stage tumors; we would expect to see about 10 probes with this property if there was no batch effect. In addition, we found 121 of the 200 to be significantly different between the initial and new batches of late stage cases and 39 of 200 to be different between the Duke and Moffitt subsets of the new advanced-stage subjects. This suggests the presence of a batch effect that results in a location shift and possibly a scale shift.

Analysis of expression data for the oncogenic pathway signatures for the mcy, ras, src, β-catenin, and E2F1, E2F3, p63, AKT, and PI3K oncogenes was done as previously described (15). The oncogenic pathway signatures were developed on the Affymetrix Hu133 plus 2.0 arrays, and the ovarian data were generated on the Affymetrix Hu133A array; thus, only those probes that overlap between both were used in developing the signatures. Because of the potential for batch variation, the batch corrected data were used in the analysis. Each signature is a metagene that summarizes its constituent genes as a single expression profile, and is derived as the first principal component of that set of genes as determined by a singular value decomposition. In predicting the pathway activation of cancer samples, gene selection and identification is based on the training data, and then metagene values are computed using the principal components of the training data and additional expression data from the samples. Bayesian fitting of binary probit regression models to the training data then permits an assessment of the relevance of the metagene signatures in within-sample classification, and estimation and uncertainty assessments for
the binary regression weights mapping metagenes to probabilities of relative pathway status. Predictions of the relative pathway status of the validation tumor samples are then evaluated, producing estimated relative probabilities – and associated measures of uncertainty – of activation/deregulation across the validation samples. Differences in mean pathway expression between tumor types were evaluated using Mann-Whitney \( t \) tests.

**MAL immunostaining.** Formalin-fixed, paraffin-embedded tissues were serially sectioned in 4- to 5-μm-thick sections, deparaffinized in three changes of xylene, and then rehydrated in graded alcohols. The slides were quenched for endogenous peroxidase with an aqueous solution of 3% hydrogen peroxide for 10 min. The sections were rinsed in three washes of PBS, preincubated in Background Terminator (Biocare Medical) for 5 min, then incubated in a humidity chamber with the mouse monoclonal antibody raised against a synthetic MAL peptide (AA 114-123, ref. 20) at room temperature for 1 h. Detection was accomplished with the two-step Heuristic Pattern Reduction (HPR) method of detection, using the Universal 4plusHPR horseradish peroxidase kit and the chromogen diaminobenzidine (Biocare Medical). Slides were counterstained with hematoxylin, and dehydrated and coverslipped. A gynecologic pathologist evaluated an adjacent tissue section stained with H&E to confirm correct histology and that viable cancer was present. Staining results were expressed on a scale of 0 to 4 based on the sum of the products of the fraction (0-1.0) of cells stained at different intensities (0-4). Survival was evaluated using Kaplan-Meier [Log-rank (Mantel-Cox) Test]. Lower MAL expression was associated with improved survival no matter what staining cutoff was used.

**Multivariate analysis of MAL immunostaining with clinical and demographic parameters was done using the Cox proportional hazards model on the 122 advanced-stage cases for which complete data were available.**

### Results

Microarray analysis was done on 101 ovarian cancers using U133A gene chips (Table 1). To correct for batch effect, these newly analyzed cases were normalized to samples of the corresponding type from our prior report (12). The original batch comprises the 54 previously reported late stage cases that were used to build the predictive model of long- versus short-term survival, as well as 5 low malignant potential (borderline) cases and 8 early-stage invasive cases. We analyzed four additional sample batches defined according to time and location of tissue processing (see supplemental materials). For each sample type within each batch, we linearly transformed each expression variable to have mean and variance matching batch 1 cases of the corresponding type. With the samples thus transformed, we treated the new samples as a “test set” and predicted whether they were likely to be long- or short-term survivors using the previously developed model (12).

The best linear discriminant model that we had developed previously for predicting long versus short survival in advanced ovarian cancer included seven genes (Table 2). To further validate the predictive model for long- versus short-term survival in advanced ovarian cancer, microarray analysis of 42 advanced-stage cases from Duke and Moffitt was done at Moffitt. This included 33 short-term survivors who lived <3 years after initial diagnosis and 9 long-term survivors who lived >7 years. Using a threshold probability of 0.5 to classify these samples, 81% (34 of 42) of the new test set of advanced-stage

---

**Table 1. Demographic and clinical characteristics of 101 serous ovarian borderline, early invasive and advanced invasive cases**

<table>
<thead>
<tr>
<th>Stage</th>
<th>Borderline (n = 20)</th>
<th>Early invasive (n = 39)</th>
<th>Advanced invasive (n = 42)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>#/%</td>
<td>#/%</td>
<td>#/%</td>
</tr>
<tr>
<td>IA</td>
<td>6 (30%)</td>
<td>4 (10.3%)</td>
<td></td>
</tr>
<tr>
<td>IB</td>
<td>1 (5%)</td>
<td>2 (5.1%)</td>
<td></td>
</tr>
<tr>
<td>IC</td>
<td>5 (25%)</td>
<td>7 (17.9%)</td>
<td></td>
</tr>
<tr>
<td>IIIA</td>
<td>3 (15%)</td>
<td>1 (2.6%)</td>
<td></td>
</tr>
<tr>
<td>IIIB</td>
<td>3 (15%)</td>
<td>18 (46.2%)</td>
<td></td>
</tr>
<tr>
<td>IIIC</td>
<td>3 (15%)</td>
<td>18 (46.2%)</td>
<td>32 (76%)</td>
</tr>
<tr>
<td>III</td>
<td>2 (10%)</td>
<td>10 (24%)</td>
<td></td>
</tr>
<tr>
<td>IV</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Histologic grade</th>
<th>Borderline</th>
<th>Early invasive</th>
<th>Advanced invasive</th>
</tr>
</thead>
<tbody>
<tr>
<td>Well differentiated</td>
<td>20 (100%)</td>
<td>3 (7.7%)</td>
<td>2 (5%)</td>
</tr>
<tr>
<td>Moderately differentiated</td>
<td>8 (20.5%)</td>
<td>15 (36%)</td>
<td></td>
</tr>
<tr>
<td>Poorly differentiated</td>
<td>28 (71.8%)</td>
<td>25 (59%)</td>
<td></td>
</tr>
</tbody>
</table>

**Table 2. Genes used in prediction of long-term versus short-term survival in advanced ovarian cancers**

<table>
<thead>
<tr>
<th>ProbeID</th>
<th>Gene</th>
<th>New evaluation set fold difference</th>
<th>Original training set fold difference</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>Short/long P</td>
<td>Short/long P</td>
</tr>
<tr>
<td>204777_s_at</td>
<td>MAL</td>
<td>0.34 0.015</td>
<td>0.33 &lt;0.001</td>
</tr>
<tr>
<td>218140_x_at</td>
<td>APMCF1</td>
<td>1.06 0.644</td>
<td>1.39 &lt;0.001</td>
</tr>
<tr>
<td>203401_at</td>
<td>PRPS2</td>
<td>1.23 0.365</td>
<td>1.58 0.0001</td>
</tr>
<tr>
<td>216076_at</td>
<td>L3MBTL</td>
<td>0.76 0.008</td>
<td>0.73 0.0004</td>
</tr>
<tr>
<td>204112_s_at</td>
<td>HNMT</td>
<td>1.26 0.289</td>
<td>1.65 0.0002</td>
</tr>
<tr>
<td>202582_s_at</td>
<td>RANBP1</td>
<td>1.06 0.716</td>
<td>1.48 &lt;0.001</td>
</tr>
<tr>
<td>214198_s_at</td>
<td>DGCR2</td>
<td>1.32 0.026</td>
<td>1.37 0.0001</td>
</tr>
</tbody>
</table>
cancers were predicted correctly using the previously developed 7-gene linear discriminant model (Fig. 1). Correct predictions were obtained in 82% (27 of 33) of short-term survivors and 78% (7 of 9) of long-term survivors. The median and mean predicted probabilities of long-term survival for the advanced-stage samples were 0.1 and 0.21 (SD, 0.31) for short-term survivors and 0.65 and 0.66 (SD, 0.33) for long-term survivors. Among the short-term survivors, correct predictions were obtained in 13 of 17 cases from Duke and 14 of 16 from Moffitt.

The out-of-sample predictive model trained on all advanced-stage cases from the model published previously (12) depends on expression as measured by seven probes. Of these, three [204777_s_at (MAL), 216076_at (L3MBTL), and 214198_s_at (DGCR2)] continued to show significant differences – by $t$-test – in the advanced-stage cancers from the new out-of-sample test set (Table 2). For each of these three probe sets, the estimated fold difference was very similar in both magnitude and direction between the original training and present evaluation sets. As in the training set, the largest fold change between long- and short-term survivors in the evaluation set was in the MAL gene. Expression of this gene was again found to be 3-fold higher in short-term survivors compared with long-term survivors and was the most important gene in the predictive model. Likewise, expression of MAL was much higher (22-fold) in short-term survivors with advanced disease compared with the early-stage cases discussed below.

Because MAL continued to be one of the most predictive transcripts for distinguishing aggressive biology, we evaluated levels of the protein in a series (partially overlapping with cancers subjected to array analysis) of 144 late-stage serous cancers with known outcome. Immunohistochemical analysis of formalin-fixed paraffin-embedded samples using a mouse monoclonal antibody raised against a synthetic peptide (AA 114-123 of the human MAL protein; ref. 20) revealed highly variable staining in the malignant epithelium (examples shown in Fig. 2A). Tissues were categorically assigned a staining value from 0 to 4+ based upon the intensity multiplied by the percentage of positive tumor cells. Of the 144 cases stained, we had previously obtained expression microarray data on 52. From this overlapping set, we observed a high degree of concordance ($r^2 = 0.66; P < 0.0001$) between RNA and protein levels of MAL. Using 1+ staining as a cutoff (95 MAL low versus 49 high), we examined whether MAL protein levels had prognostic value in this cohort. Overall survival was significantly ($P = 0.0004$) reduced in patients with tumors that showed high levels of MAL protein (Fig. 2B). In multivariate analysis, MAL staining retained significance with response to chemotherapy (Table 3). Other cut points for MAL staining yielded similar statistically significant results. MAL is also highly expressed in benign fallopian tube with accentuated staining at the apical surface of the serous epithelial consistent with its role in transport and signaling. Benign ovarian surface epithelial cells were negative for MAL expression using this antibody (not shown).

All but one of the 39 early-stage invasive cancers was predicted to be a long-term survivor (Fig. 1). The mean prediction was 0.92 (SD, 0.11) and the median was 0.96. The median predictions for the invasive early-stage cancers were similar among institutions (Duke = 0.98, Memorial Sloan-Kettering Cancer Center = 0.98, Gynecologic Oncology Group = 0.94). Among the borderline tumors, 15 of 20 (75%) were predicted to be long-term survivors with a mean probability of 0.72 (SD, 0.27) and a median probability of 0.86. Among the 17 patients with early-stage invasive cancers from Duke and Sloan Kettering in whom follow up was available, 1 of 6 with stage I disease has died (92-month survivor) and only 2 of 11 with stage IIC disease have died. Median follow-up of these 17 cases is 56 months (range, 15-180 months). The one patient who was predicted to be a short-term survivor (prediction 0.36) had stage IIC disease and is alive without recurrence. Information on treatment and outcome for the 22 early-stage invasive cancers obtained from the Gynecologic Oncology Group was requested, but is not available. All patients with borderline tumors are alive with a median follow-up of 76 months (range, 11-190 months).

Analyses of groups of genes comprising functional pathways have also revealed important information about the molecular changes present in malignancy. Previously defined oncogenic pathway signatures (15) were thus evaluated in both the 65 invasive serous ovarian cancers from our prior publication (54 advanced, 11 early; ref. 15) and the new set of 101 cases (42 advanced, 39 early, 20 borderline; Fig. 3). Pathway signatures associated with proliferation such as myc, E2F1 and E2F3, as well as $\beta$-catenin and pI3K, were higher in invasive
cancers than in borderline tumors ($P < 0.001$). Conversely, the mean $\text{ras}$ signature was significantly higher in borderline tumors (0.75) compared with that seen in early-stage (0.49) or short/long survivors with advanced disease (0.45, 0.42; $P < 0.001$). The AKT and p63 pathway signatures did not vary by tumor type. As previously described (15), the src pathway was most highly expressed in the least favorable subgroup, those who were short-term survivors ($P < 0.001$). Further details regarding the mean pathway signatures in each category are available in Supplementary Table S2.

**Discussion**

About 10% of ovarian cancers arise in women with germline mutations in $\text{BRCA1}$ and $\text{BRCA2}$, and most of these cases are high-grade serous cancers as are most advanced-stage sporadic cases. Invasive serous ovarian cancers have a propensity to present at an advanced stage and are responsible for most ovarian cancer deaths. Less than 5% of invasive serous ovarian cancers are confined to the ovary at diagnosis. Most early-stage serous ovarian cancers are low malignant potential borderline tumors. Genetic analyses have suggested that borderline serous tumors are not precursors to high-grade invasive cancers. Mutations of the $K\text{-ras}$ and $\text{BRAF}$ genes are frequent in borderline tumors, but rarely occur in invasive cases (21). Conversely, invasive cancers are characterized by frequent $\text{TP53}$ mutations (22), but the frequency of these mutations is similar

![Fig. 2. Immunodetection of MAL protein in primary tissues and relation to disease outcome. A. examples of immunohistochemical detection of the MAL protein in two advanced serous ovarian cancers and a benign fallopian tube with negative control (mouse IgG) and H&E-stained parallel sections. Examples of weak (1+) and strong (4+) expression are shown. For each staining run, a benign kidney was included that serves as both a positive (proximal tubules) and negative (glomeruli) control for the MAL protein. B. Kaplan-Meier survival curves for 144 late stage serous ovarian cancer patients stratified by MAL staining using +1 as a threshold to define high versus low expression ($P = 0.0004$).](image)

| Table 3. Cox multivariate analysis by survival in advanced ovarian cancer |
| Variable | Hazard ratio | $P$ |
| Clinical response (CR vs. other) | 0.15 | <0.0001 |
| MAL expression (≥1) | 1.9 | 0.004 |
| Stage (IV vs. III) | 1.5 | 0.13 |
| Race (black vs. white) | 0.71 | 0.24 |
| Debulking status (suboptimal vs. optimal) | 1.3 | 0.34 |
| Age (continuous) | 1.0 | 0.91 |

**Abbreviation:** CR, completed response.
in early- and advanced-stage invasive serous cases (23). Microarray studies also have highlighted global differences in gene expression between borderline and invasive serous ovarian tumors (6–8, 21).

Several groups have used microarrays as a tool to predict outcome of patients with advanced ovarian cancer (9–14, 24–26). We compared gene expression in patients who represent the extremes with respect to outcome, namely those who survived either <3 years or >7 years. The observation that no single gene was more than 3-fold differentially expressed validates the rationale for examining patterns of gene expression that may reflect underlying tumor biology. Spentzos et al. used microarrays to develop a 115-gene model that classified ovarian cancers into two groups that exhibited significantly different survival (13). The results are consistent with ours and together suggest that clinical differences in outcome are reflected in global patterns of gene expression that can be appreciated using microarrays.

In view of the potential for false discovery, we previously confirmed that expression of the genes that comprise our predictive model for survival held prognostic value in the independent set of tumors used in the Spentzos study that were

---

**Fig. 3.** Scatter plots of oncogenic pathway signatures for myc, E2F1, E2F3, ras, src, β-catenin, AKT, PI3K and p63 in borderline and invasive ovarian cancers. A, pathway signatures that were higher in all invasive cancers (early, long-term survivors, short-term survivors) relative to borderline tumors. All \( P < 0.0001 \) except β-catenin (borderline vs early \( P = 0.21 \), long-term \( P = 0.02 \), short \( P < 0.0001 \)). B, additional pathway signatures. Ras pathway is higher in borderline cases relative to invasive cases (early, long-term survivors, short-term survivors; all \( P < 0.001 \)). Expression of the AKT and p63 pathway signatures did not differ between borderline and invasive cancers. The src pathway was most highly expressed in short-term survivors relative to long-term survivors \( (P = 0.001) \).
analyzed at another institution using a different microarray platform (12). In the present study, we have shown an 81% predictive accuracy for long- versus short-term survival in a new set of 42 advanced-stage serous cancers for which the microarray analysis was done at another institution. These findings not only validate the predictive model, but also show that they can be replicated when samples are collected and subjected to independent microarray analysis (in a laboratory at another institution). These predictions were essentially driven by three of the seven gene probes in the original model that maintained significance in the new evaluation set of cancers. Once again, the MAL gene had the highest fold difference between short- and long-term survivors (3-fold), and was more than 20-fold higher in short-term survivors than in early-stage cases. The other two significant probe sets correspond to L3MBTL (U133A: 216076_at) and DGCR2 (U133A: 214198_s_at).

The availability of a specific monoclonal antibody that can detect the MAL protein in situ allowed us to further investigate expression of this gene with respect to ovarian cancer outcome. Staining a series of cancers lends further support for the expression of this gene as a prognostic indicator. From the set of specimens for which expression array data were available, we observed excellent concordance between protein and mRNA levels ($r^2 = 0.66$). By staining a larger set of advanced serous cancers unselected for disease outcome, we found that high-level expression of the MAL protein is associated with shorter survival. We also show that benign fallopian tube epithelium have abundant expression of the MAL protein whereas ovarian surface epithelial cells have no detectable protein by immunohistochemistry. These results further validate the expression array analysis and provide an interesting gene target for additional mechanistic studies.

In our initial study we found that patterns of gene expression in advanced-stage serous cancers from long-term survivors, which represent only a small fraction of advanced-stage serous cases, were shared by 11 early-stage serous cancers (12). In the present study, we were able to validate this finding using an independent set of cancers obtained from three tumor banks. We found that all but 1 of the 39 invasive early-stage serous cancers and 75% of borderline cases were predicted to be long-term survivors using the 7-gene model. These data provide compelling evidence that the favorable clinical outcome of both long-term survivors with advanced-stage disease and early-stage cases is attributable to a shared underlying biology. Conversely, a more virulent biology likely underlies the majority of ovarian cancers that are detected at an advanced stage and have worse survival.

Whether these differences in biology are related to the natural disease course or intrinsic response to therapy cannot be discriminated with these data. Virtually all of the advanced cancers in this study were treated in a similar fashion, providing no “control” arm. Although some of the early-stage cancers were treated with chemotherapy, we have incomplete data both on treatment and outcome from this cohort. At least for the MAL protein, we have independent data from breast cancer indicating that expression of this gene is a prognostic factor related to chemotherapeutic treatment (27). In the current study, high MAL expression does correlate with reduced clinical response after surgery ($t$-test, $P = 0.04$) further supporting this possibility.

These findings have implications for screening as an approach to decreasing ovarian cancer mortality. They suggest that serous cancers presently diagnosed at an early stage are a small subset with the most favorable biology rather than being representative of the full spectrum of the disease. The underlying biology of these cancers may confer slower growth and a longer interval of time during which these cancers exist prior to the development of disseminated metastasis. Conversely, the observation that only one early-stage invasive serous ovarian cancer exhibited patterns of gene expression predictive of short-term survival in advanced disease suggests that the most highly lethal serous ovarian cancers may not be easily amenable to early detection. To be effective in reducing mortality due to high-grade serous ovarian cancer, screening approaches must have the ability to detect cancers that have a less favorable biology. This represents a considerable challenge, as these cancers likely are more rapidly progressive. Screening will have little impact on ovarian cancer mortality if it merely identifies cancers already destined to have a favorable outcome when diagnosed clinically. The utility of computerized tomography screening to reduce lung cancer mortality has been questioned on similar grounds.

In addition to defining gene expression patterns that predict outcome, our group has also developed genomic signatures associated with overexpression of specific oncopgenes (15). These signatures were developed by transfection and overexpression of specific oncogenes in normal mammary epithelial cells. The level of expression of these oncogenic pathway signatures in cell lines has been shown to be predictive of response to biological therapies that target these pathways (15). In principle, these signatures could be used to direct the use of biological therapies that target these pathways specifically (28).

In this article we examined the pathway signatures in the spectrum of serous ovarian cancers from borderline tumors to early-stage invasive cases and advanced-stage cases with poor and favorable outcomes. Expression of the $myc$ and $E2F1$ and $E2F3$ pathway signatures, which reflect cell cycle progression, was higher in invasive cancers of all stages than in borderline tumors. Likewise, in another study that used microarrays to examine gene expression in serous ovarian cancers, Bonome et al. found that expression of cellular proliferation genes was higher in invasive cancers relative to borderline tumors, whereas invasive cancers had activation of pathways involved in metastasis and chromosomal instability (6). These findings are also consistent with older studies that have shown that proliferation rates are higher in invasive cancers relative to borderline tumors (29). The $\beta$-catenin and p38K pathway signatures also were notably higher in invasive cancers compared with borderline cases, and likely reflect differences in the molecular pathogenesis of these tumor types.

Expression of the ras pathway signature was much higher in borderline tumors relative to early and advanced invasive cancers. This is consistent with the prior observation that activating mutations in codon 12 of $K-ras$ occur in a significant fraction of serous borderline tumors, but rarely in invasive cancers (21). These findings parallel our prior microarray studies of lung cancer (15). High expression of the ras pathway signature in lung cancers was noted to be characteristic of adenocarcinomas, which also frequently have $K-ras$ mutations (15). In contrast, the average ras pathway signature was strikingly lower in squamous lung cancers, which do not have...
K-ras mutations. Interestingly, activation of the ras pathway has a more favorable clinical outcome in both ovarian (borderline) and lung (adenocarcinoma) cancers. However, the level of ras pathway activation was not a predictor of outcome among invasive serous ovarian cancers.

Previously we reported that the src and β-catenin signatures were associated with poor outcome in advanced serous ovarian cancers (15), and that the src pathway was particularly predictive of poor survival in patients who had an incomplete response to primary platinum-based chemotherapy (16). In addition, our group (15) and others (30) have shown that signatures of src pathway activation are predictive of response to biological therapies directed at this pathway. In the present study, both the src and β-catenin pathway signatures were more highly expressed in advanced-stage ovarian cancers that were short-term survivors relative to more favorable subgroups. This provides further evidence of the more virulent biology underlying ovarian cancers with high expression of these pathways. The mean p38k pathway expression, but not AKT or p63, was also higher in invasive cancers than in borderline tumors.

In summary, both the model for long versus short survival in advanced ovarian cancer and the oncogenic pathway signatures suggest that early- and advanced-stage serous ovarian cancers are fundamentally different at a biological level. These data suggest that the excellent clinical outcome of early-stage cases is not attributable to fortuitous diagnosis prior to metastatic dissemination. Rather, clinically detectable early-stage cancers are diagnosed before they spread because they are inherently less virulent.

References


