Molecular Biomarkers of Residual Disease after Surgical Debulking of High-Grade Serous Ovarian Cancer

Susan L. Tucker1, Kshipra Gharpure2, Shelley M. Herbrich1, Anna K. Unruh1, Alpa M. Nick2, Erin K. Crane6, Robert L. Coleman3, Jamie Guenthoer6, Heather J. Dalton2, Sherry Y. Wu2, Rajeshu Rupaimoole2, Gabriel Lopez-Berestein3,5, Bulent Ozpolat3, Cristina Ivan2, Wei Hu2, Erin K. Crane2, Robert L. Coleman3, Jamie Guenthoer6, Heather J. Dalton2, Sherry Y. Wu2, Rajeshu Rupaimoole2, Gabriel Lopez-Berestein3,5, Bulent Ozpolat3, Cristina Ivan2, Wei Hu2, Keith A. Baggerly1, and Anil K. Sood2,4,5

Abstract

**Purpose:** Residual disease following primary cytoreduction is associated with adverse overall survival in patients with epithelial ovarian cancer. Accurate identification of patients at high risk of residual disease has been elusive, lacking external validity and prompting many to undergo unnecessary surgical exploration. Our goal was to identify and validate molecular markers associated with high rates of residual disease.

**Methods:** We interrogated two publicly available datasets from chemonaive primary high-grade serous ovarian tumors for genes overexpressed in patients with residual disease and significant at a 10% false discovery rate (FDR) in both datasets. We selected genes with wide dynamic range for validation in an independent cohort using quantitative RT-PCR to assay gene expression, followed by blinded prediction of a patient subset at high risk for residual disease. Predictive success was evaluated using a one-sided Fisher exact test.

**Results:** Forty-seven probe sets met the 10% FDR criterion in both datasets. These included FABP4 and ADH1B, which tracked tightly, showed dynamic ranges >16-fold and had high expression levels associated with increased incidence of residual disease. In the validation cohort (n = 139), FABP4 and ADH1B were again highly correlated. Using the top quartile of FABP4 PCR values as a prespecified threshold, we found 30 of 35 cases of residual disease in the predicted high-risk group (positive predictive value = 86%) and 54 of 104 among the remaining patients (P = 0.0002; OR, 5.5).

**Conclusion:** High FABP4 and ADH1B expression is associated with significantly higher risk of residual disease in high-grade serous ovarian cancer. Patients with high tumoral levels of these genes may be candidates for neoadjuvant chemotherapy. Clin Cancer Res; 20(12): 3280–8. ©2014 AACR.

Introduction

Surgery is widely accepted as a critical component of treatment for high-grade serous ovarian cancer (HGSOC), although the optimal timing of surgery has been debated.

Residual disease following a primary cytoreductive effort is negatively associated with clinical metrics, including response to adjuvant chemotherapy, progression-free survival, and overall survival. Initial reports found improved survival among women left with <1.5 cm residual disease compared with larger volume disease (1–3). Since then, the yardstick for defining “optimal” debulking has varied between <2 and 0 cm (complete resection or R0). Recent reports suggest that the best metric is the distinction between R0 and any residual disease (3, 4), which is more reliably assessed and reproducible (external validity).

Accurate predictors of surgical outcome could substantially impact management of ovarian cancer by guiding patients with highest likelihood of having residual disease to neoadjuvant chemotherapy with the potential for interval cytoreductive surgery later. While baseline rates of leaving some extent of residual disease at upfront surgery are high (65%–75%), identifiable factors are needed to indicate whether a specific patient is at significantly higher risk. Some studies have attempted to define predictors of optimal cytoreduction, most often defined as <1 cm residual disease (5). However, given the considerable variability...
Biomarkers of Residual Disease in Ovarian Cancer

Translational Relevance
Primary debulking surgery followed by adjuvant chemotherapy has been the standard for treatment of women with high-grade serous ovarian cancer. Residual disease following primary cytoreduction is known to be associated with poor overall and progression-free survival, and patients at high risk of having residual disease may be candidates for neoadjuvant chemotherapy. Our study used publicly available microarray data to identify 2 biomarkers, FABP4 and ADH1B, which were subsequently validated in an independent clinical cohort to identify patients at high risk of residual disease after primary surgery. Use of these biomarkers could help clinicians personalize treatment options for women suspected of having ovarian cancer. Our study has led to ongoing development of a CLIA-based assay for use of these biomarkers in a prospective validation trial.

in definition and assessment of debulked disease, such predictors have not been particularly reliable. Other efforts to spare unnecessary primary debulking surgery by pre- or intraoperative assessment have abounded (6–8). Unfortunately, none have reached the level of external validity necessary for incorporation into general practice.

A predictor of residual disease with high sensitivity is unlikely to exist, as incomplete resection sometimes occurs owing to tumor location near critical organ structures. In many cases, though, residual disease is a consequence of biologic tumor characteristics, with wide dissemination of disease throughout the pelvis. We hypothesized that in the latter case, a high likelihood of residual disease might be predictable based on biomarkers assessed from tumor tissue. The goal of the present study, therefore, was to identify molecular markers associated with a high likelihood of residual disease. We used 2 publicly available microarray datasets that included residual disease information to discover candidate gene markers and subsequently validated our biomarkers in an independent clinical cohort. In the validation cohort, blinded predictions of residual disease risk based on candidate biomarker gene expression assayed using quantitative real time polymerase chain reaction (qRT-PCR) were compared with actual surgical outcomes.

Materials and Methods
Here, we briefly outline our materials, experimental procedures, and methods of analysis. Full details (including computer code) are given in the Supplementary Appendix and at http://bioinformatics.mdanderson.org/Supplements/ResidualDisease.

Data for exploratory studies
For biomarker discovery, we used 2 large publicly available Affymetrix microarray datasets involving patients with HGSOCs and providing associated clinical information, including residual disease status. The first of these was the ovarian cancer cohort from The Cancer Genome Atlas (TCGA; ref. 9). We downloaded CEL files (level 1 data) for the ovarian samples (Affymetrix HT HG-U133A arrays, \( n = 598 \)) on September 2, 2012; these represent the TCGA update that was current as of June 24, 2011 (revision 1007). We downloaded the associated clinical data \( (n = 576) \) on September 14, 2012. We omitted 4 samples marked for exclusion by TCGA and performed additional sample filtration based on clinical annotation. We excluded samples if they were from recurrent tumor, omental tumor, or normal tissue. When there were multiple primary tumor samples per patient, we retained data from just one sample. We also excluded cases if there was no information about residual disease status, if the tumor was not high grade, or if the patient received neoadjuvant chemotherapy. The second dataset was from the study of Tothill and colleagues (10). We downloaded CEL files (Affymetrix U133+2 arrays, \( n = 285 \)) and clinical data from the Gene Expression Omnibus (GSE9891) on September 13, 2012. We excluded cases from this dataset if tumor samples were low grade, of low malignant potential, non-serous histology, or non-ovarian or peritoneal origin. Cases were also excluded if the patient received neoadjuvant chemotherapy or if residual disease status was not provided. Accordingly, biomarker discovery was performed using only data from primary tumors of chemo naïve patients with HGSOCs. Following initial biomarker discovery, we considered 2 additional datasets for qualitative checks on patterns of expression for genes of interest (Bonome and colleagues and The Cancer Cell Line Encyclopedia; see Supplementary Appendix for details; refs. 11, 12). We note that although the study of Bonome and colleagues includes clinical data, we could not use it for validation because in that dataset residual disease is scored as optimal versus suboptimal, whereas our method was developed to predict risk of any residual disease versus R0.

For each of the 4 datasets described above, we quantified expression at the probeset level with R statistical software (version 2.15.1) using the robust multi-array average procedure (13) as implemented in the subroutine justRMA in the R “affy” package (version 1.34.0). We only considered probesets common to the array platforms used in the first 2 datasets. Except for the 4 samples marked for exclusion by TCGA, our sample filtrations described above were performed subsequent to the quantification step.

Validation studies
Following identification of candidate genes having a wide dynamic range and with high expression levels associated with high risk of residual disease in exploratory analyses (see below), we performed validation studies in an independent cohort. Following Institutional Review Board approval, we obtained primary ovarian tumor samples from archived surgical material at The University of Texas MD Anderson Cancer Center (Houston, TX; \( n = 84 \)) and the Pacific Ovarian Cancer Research Consortium at the Fred Hutchinson Cancer Research Center (Seattle, WA;
n = 55). We also included 41 samples of omental tumor tissue from MD Anderson for comparison. All samples were selected from chemonaive tumors. Information about residual disease status in the validation cohort was extracted from patient medical records by clinical members of the team (A.M. Nick and A.K. Sood) and was scored as R0 versus any residual disease. This information was kept blinded from all other team members until after predictions were made concerning identification of a subset of patients at high risk for residual disease. These predictions were made on the basis of expression levels of a candidate gene of interest relative to 18S, assayed using qRT-PCR (ref. 14; see Supplementary Appendix for additional details).

Statistical methods

Exploratory data analyses. We used 2-sample t tests to compare expression levels by residual disease status (R0 vs. any residual disease) in the TCGA and Tothill datasets, separately. We analyzed the resulting collections of nominal P values using the method of Benjamini and Hochberg (15) to identify probesets significant at a 10% or 5% false discovery rate (FDR) in each data set separately. We selected probesets meeting the specified criterion in both data sets. We produced density plots for these probesets, as well as heatmaps using hierarchical clustering, to illustrate patterns of expression. We examined the resulting plots to identify probesets with (i) wide dynamic range and (ii) high levels of expression associated with particularly high incidence of residual disease. We examined similar plots for the selected probesets in the Bonome and CCLE datasets to see whether the qualitative expression patterns were also observed there.

Selecting a sample size and prediction threshold for the validation study. Following identification of genes for which high expression levels were associated with high risk of residual disease, we wished to select an a priori decision threshold for “calling” a patient in the validation cohort to be at high risk of residual disease based on elevated biomarker expression measured by qRT-PCR. Patients with expression of a candidate biomarker gene above the selected threshold would be in the predicted high-risk group for residual disease. To select a decision threshold, we first computed the positive predictive values (PPV) in the TCGA and Tothill datasets at thresholds defined by varying quantities of biomarker expression. We assumed that similar PPVs would apply to the validation cohort and performed numerical simulations to determine a decision threshold at which we would be maximally powered to detect a difference in incidence of residual disease in the validation set. We chose a threshold with an estimated power of at least 80% to detect a difference in incidence of residual disease significant at P < 0.05 using a 1-sided Fisher exact test. We used similar simulations for our initial sample size computations before collecting the validation set.

Statistical assessment of success. We constructed a 2 × 2 table showing our calls (high risk of residual disease vs. lower risk of residual disease) compared with actual surgical outcomes (residual disease vs. R0) in the validation cohort. We used a 1-sided Fisher exact test to test the null hypothesis of equivalent rates of residual disease in the 2 groups against the alternative of increased incidence of residual disease in the predicted high-risk group.

Results

Clinical outcomes in the exploratory cohorts

In the TCGA and Tothill datasets, 491 and 189 patients, respectively, met the criteria for inclusion in our analysis. Figure 1 shows overall survival in these cohorts by extent of residual disease after debulking surgery. Survival is significantly better for patients with complete resection (R0) compared with those with any residual disease, including those traditionally considered to be optimally debulked (1–10 mm residual disease). This result, reported previously for the TCGA data (9), is consistent with the findings from other studies as well (3, 4). The percentages of patients with any residual disease are 77% (n = 378) and 74% (n = 139) in the TCGA and Tothill cohorts, respectively.

Identification of genes associated with increased incidence of residual disease

There are 47 probesets, representing 38 different genes (Supplementary Table S1), with P values meeting a 10% FDR criterion in both datasets. Lowering the FDR to 5% results in 8 probesets common to both datasets. Of note, the additional filtration imposed by requiring success in both datasets is severe; the 8 common probesets represent the intersection of 149 probesets meeting the 5% FDR threshold for TCGA and 81 meeting the same criterion for the Tothill cohort. Heatmaps of standardized expression of the 8 probesets meeting the 5% FDR criterion in both data sets are shown in Fig. 2; heatmaps for the 47 probesets meeting the 10% FDR criterion in both data sets are included in the Supplementary Appendix.

Figure 3 shows density plots of expression levels for 2 of the top 8 candidate probesets of interest, LUM and FABP4 (plots for all 47 probesets identified at the 10% FDR level are included in the Supplementary Appendix). While both LUM and FABP4 show increased rates of residual disease at higher expression levels, the distributions of the 2 genes are qualitatively different. For FABP4, there is a level of expression above which nearly all patients have residual disease. For LUM, in contrast, there is no clearly defined cutoff point above which the majority of patients have residual disease.

An additional probeset exhibiting the same type of qualitative behavior as for FABP4 corresponds to ADH1B. For the FABP4 and ADH1B probesets, the distribution of expression values is bimodal, with a large cohort having low expression levels falling within a relatively narrow range (the lower limit of detection for the assay), and a smaller cohort with substantially (up to orders of magnitude) higher expression levels. Moreover, the
heatmaps in Fig. 2 show that the highest expression levels of FABP4 and ADH1B occur in a subset of patients with a very high incidence of residual disease. Figure 4A and B show the highly correlated levels of FABP4 and ADH1B in the TCGA and Tothill datasets, respectively. Using cutoffs of 3.5 on both TCGA axes, 97 of 107 (90.6%)
patients with high expression of both genes have residual disease versus 281 of 384 (73.2%) of the remaining patients. Using cutoffs of 5.25 and 4.5 for FABP4 and ADH1B, respectively, in the Tothill data, 59 of 63 (93.7%) patients with high expression of both genes have residual disease, versus 80 of 126 (63.5%) of the others.
In the cohort of Bonome and colleagues, we again see a subset of patients with very high levels of both FABP4 and ADH1B (Fig. 4C), but the correlation between the 2 is not as tight as for the TCGA and Tothill datasets, possibly as a result of the microdissection used in the study of Bonome and colleagues. In the cell line data from CCLE (Fig. 4D), joint overexpression of FABP4 and ADH1B does not occur among the ovarian cancer lines, although there is clear overexpression of each gene in some cell lines. This suggests the linkage we see in clinical samples may be due to some type of stromal interaction.

Validation studies

In general, levels of FABP4 in the validation studies were markedly higher in tumor samples from omentum than from primary tumor (Supplementary Appendix). Therefore, we restricted our attention to the samples from primary tumor for consistency with the datasets of TCGA and Tothill and colleagues. The paired FABP4 and ADH1B qRT-PCR values for the primary ovarian cancer samples are shown in Fig. 5. Expression levels of the 2 genes are highly correlated, and there is no difference in distribution between the values from samples contributed by the 2 institutions.
Although the correlations between the microarray values for \textit{FABP4} and \textit{ADH1B} in our discovery cohorts suggested that we would not need both for successful prediction of high residual disease risk, it was not certain, \textit{a priori}, that either assay would work. In addition, we were interested in exploring the nature of the association between \textit{FABP4} and \textit{ADH1B} further. Therefore, we assayed expression levels of both genes in our validation studies. However, after observing the results in Fig. 5, and in view of the lack of elevation of \textit{ADH1B} in the CCLE data, we proceeded to predict high risk of residual disease in our validation cohort based on \textit{FABP4} expression levels alone.

The bimodal distribution seen in the microarray values for \textit{FABP4} was not seen with the qRT-PCR values in the validation cohort, pointing to a dynamic range limitation for the former. Therefore, we chose to select a cutoff for predicting high risk of residual disease based on a specific quantile of \textit{FABP4} expression (Supplementary Appendix). Specifically, we selected a prediction threshold corresponding to the 25\% of patients with the highest levels of \textit{FABP4} expression; we predicted that these patients had a significantly higher incidence of residual disease than the remaining patients.

After unblinding the clinical findings concerning residual disease in the validation cohort, we observed that among the 35 patients predicted to be at high risk for residual disease, 30 (PPV 86\%) did have residual disease. In contrast, only 54 of the 104 patients with \textit{FABP4} values below the decision threshold (52\%) had incomplete resections. This difference is highly significant using a 1-sided Fisher exact test ($P = 0.0002$) and corresponds to an OR of 5.5. Our predictive method based on \textit{FABP4}, therefore, correctly identified a cohort of patients with significantly increased rates of residual disease in an independent test set. Further examination of the \textit{ADH1B} results (performed after unblinding) showed that predictions using either \textit{ADH1B} alone or in combination with \textit{FABP4} would have produced similar results (Fig. 5 and Supplementary Appendix).

To obtain insight into the potential biological role of \textit{FABP4} and \textit{ADH1B}, we used the reverse-phase protein array (RPPA) data available from TCGA to investigate which proteins, if any, had levels correlated with \textit{FABP4} and \textit{ADH1B} expression in the TCGA samples. While all correlations were low, there were nonetheless more significant associations than could be explained by chance (Supplementary Appendix). Many of the positively associated proteins are related to metastasis or proliferation, and among the proteins negatively associated with both genes is E-cadherin, which might imply greater levels of epithelial-mesenchymal transition and would also be prometastatic.

### Discussion

We have shown that high expression levels of \textit{FABP4} and \textit{ADH1B} in primary tumor specimens are associated with high incidence of residual disease after primary debulking surgery of high-grade serous ovarian cancer. If confirmed in the biopsy setting, this discovery lays the foundation for customized surgical treatment approaches based on unbiased and objective measures from the tumor itself. Triage to neoadjuvant chemotherapy versus primary debulking could be prospectively assessed to provide an optimized treatment algorithm based on intrinsic tumor biology. For such use, biopsy from the primary versus metastatic sites would also have to be carefully considered. Consistent with a report of Nieman and colleagues (16), we observed higher levels of \textit{FABP4} in samples from the omentum compared with primary tumor and in 4 patients with samples available from both sites, \textit{FABP4} levels were higher in metastatic tissue (Supplementary Appendix).

These results suggest that if omental levels of \textit{FABP4} expression are also associated with risk of residual disease, the \textit{FABP4} threshold for calling patients to be at high risk of residual disease might depend on whether primary or metastatic tissue is sampled.

Our identification of \textit{FABP4} and \textit{ADH1B}, which encode for fatty acid–binding protein 4 and alcohol dehydrogenase 1B, respectively, leveraged large, publicly available datasets, providing further evidence that such datasets (with clinical annotation) may be relevant for discovering approaches to tailor therapy. This assertion, however, comes with several caveats. First, results from almost all such studies can be tainted by batch effects (17). To minimize the potential for such effects, we sought associations occurring in 2 different datasets and looked for confirmation of expression patterns in at least one additional publicly available dataset, before proceeding to
validation. Second, by selecting genes whose expression levels exhibited a wide dynamic range, we focused on differences large enough to survive migration to a different assay. This is in keeping with the use of “bimodal” genes for classification (18, 19) and is similar in spirit to the “barcode” approach (20). Third, to keep the test practical, we restricted attention to a very small number of genes (here 2). Fourth, we used a clinical endpoint that should be consistent across centers (i.e., R0 resection contrasted with any residual disease).

Although the extent of cytoreduction is still most commonly scored as optimal (<1 cm) versus suboptimal (≥1 cm) resection, we elected to seek biomarkers of any residual disease, instead of suboptimal debulking, for 2 reasons. First, there is growing evidence that there is a larger difference in overall survival between patients with R0 resection versus any residual disease than between optimal versus suboptimal resection, as shown in our Fig. 1 and elsewhere (3, 4, 9). Second, the distinction between R0 resections and any residual disease is expected to be more reliably assessed and reproducible between institutions. Having elected to investigate the distinction between residual disease and no residual disease, we sought biomarkers associated with residual disease instead of optimal R0 because there is a subset of patients in whom complete resection is not achievable because of widely disseminated disease. This suggests the existence of a subset of patients at high risk of residual disease because of biologic tumor characteristics, whereas failure to achieve R0 may occur simply because of tumor location near critical organs, and other medical reasons unrelated to biologic characteristics of the tumor. We, therefore, expect that a biologic predictor for likelihood of R0 disease may not achieve high predictive accuracy. Consistent with this hypothesis, although we find a continuous trend in the risk of residual disease as a function of FABP4 expression levels (Supplementary Appendix), the incidence of R0 resection is not very high (~60%) among patients with the lowest levels of FABP4 in their primary tumors.

The present study did not address the biologic mechanisms that might underlie the ability of FABP4 and/or ADH1B to serve as biomarkers for high risk of residual disease. The study of Nieman and colleagues suggested that FABP4 plays a role in widely metastatic disease (16), which could explain the associated high risk of residual disease, and several studies have reported a role of FABP4 in angiogenesis (21, 22). We found that protein levels measured by RPPA in the TCGA data and correlated with high FABP4 or ADH1B expression also suggest increased metastasis and proliferation. We further note that several collagen genes are included among the 47 genes meeting the 10% FDR in our discovery analyses (Supplementary Appendix); this might indicate an ability of the tumor to remodel the local environment and promote metastases. Additional studies investigating both genes are currently underway in our laboratory.

While current front-line treatment for patients with HGSC consists of tumor debulking surgery followed by platinum- and taxane-based chemotherapy (23), some studies point to comparable outcome with neoadjuvant chemotherapy followed by interval debulking surgery (24). An accurate predictor could allow “personalized” surgical therapy, whereby patients unlikely to have complete debulking would forego upfront surgery and would receive neoadjuvant chemotherapy instead. With such an approach, patients with a high likelihood of residual disease at the upfront surgery could be spared from the morbidity associated with such procedures. Previous attempts to predict resectability have included use of computerized tomography, which was not found to be highly predictive (25) and preoperative serum levels of CA-125, which was found to be a significant predictor of suboptimal disease but with an OR lower than the value found here for prediction of residual disease using FABP4 (26). Consistent with our observation that low FABP4 expression does not accurately identify patients without residual disease, the latter study found that CA-125 lacks the ability to predict optimal cytoreduction. An ongoing study at our institution is investigating the utility of a laparoscopy-based assessment of the potential to achieve R0 resection, which could potentially complement the biomarkers proposed here for detecting high risk of residual disease. Development of a CLIA-compliant assay for FABP4 is currently underway to enable prospective validation of our proposed molecular biomarkers. In addition, ongoing biologic studies of FABP4 and ADH1B may elucidate the mechanisms behind some of the most refractory cases of HGSC and point to new avenues for therapeutic intervention.

Disclosure of Potential Conflicts of Interest
No potential conflicts of interest were disclosed.

Authors’ Contributions

Acknowledgments
The authors thank Drs. Charles Drescher and Muneech Tewari at the Fred Hutchinson Cancer Research Center for helpful discussions and for providing tumor samples.

www.aacrjournals.org
3287
Grant Support
This work was supported in part by the National Cancer Institute at the NIH (CA016672, CA101642, CA109298, CA128797, P50CA038363, P50CA098238, U54CA151668, U54CA36300, U54CA36297, CA177909, and U24 CA143883); the Cancer Prevention Research Institute of Texas (RP110595); the Department of Defense (BC055265, OC073399, W81XWH-10-1-0158); the Ovarian Cancer Research Fund [Program Project Development Grant]; the Betty Ann Asche Murray Distinguished Professorship; the Meyer and Ida Gordon Foundation, the Judi A. Rees Ovarian Cancer Research Fund, the Marcus Foundation; the RGK Foundation; the Gilder Foundation; and the Blanton-Davis Ovarian Cancer Research Program.

The costs of publication of this article were defrayed in part by the payment of page charges. This article must therefore be hereby marked advertisement in accordance with 18 U.S.C. Section 1734 solely to indicate this fact.

Received February 22, 2014; revised April 4, 2014; accepted April 8, 2014; published OnlineFirst April 22, 2014.

References