Biomarker-Based Ovarian Carcinoma Typing: A Histologic Investigation in the Ovarian Tumor Tissue Analysis Consortium

Martin Köbel1, Steve E. Kalloger2, Sandra Lee1, Maire A. Duggan1, Linda E. Kelemen2, Leah Prentice3, Kimberly R. Kalli4, Brooke L. Fridley9, Daniel W. Visscher5, Gary L. Keeney5, Robert A. Vierkant6, Julie M. Cunningham7, Christine Chow3, Roberta B. Ness10, Kirsten Moysich11, Robert Edwards12,13,14, Susan J. Ramus18, and Ellen L. Goode8, on behalf of the Ovarian Tumor Tissue Analysis consortium

Abstract

Background: Ovarian carcinoma is composed of five major histologic types, which associate with outcome and predict therapeutic response. Our aim was to evaluate histologic type assessments across the centers participating in the Ovarian Tumor Tissue Analysis (OTTA) consortium using an immunohistochemical (IHC) prediction model.

Methods: Tissue microarrays (TMA) and clinical data were available for 524 pathologically confirmed ovarian carcinomas. Centralized IHC was conducted for ARID1A, CDKN2A, DKK1, HNF1B, MDM2, PGR, TP53, TFF3, VIM, and WT1, and three histologic type assessments were compared: the original pathologic type, an HIC-based calculated type (termed TB_COSPv2), and a WT1-assisted TMA core review.

Results: The concordance between TB_COSPv2 type and original type was 73%. Applying WT1-assisted core review, the remaining 27% discordant cases subdivided into unclassifiable (6%), TB_COSPv2 error (6%), and original type error (15%). The largest discordant subgroup was classified as endometrioid carcinoma by original type and as high-grade serous carcinoma (HGSC) by TB_COSPv2. When TB_COSPv2 classification was used, the difference in overall survival of endometrioid carcinoma compared with HGSC became significant [RR 0.60; 95% confidence interval (CI), 0.37–0.93; P = 0.021], consistent with previous reports. In addition, 71 cases with unclear original type could be histologically classified by TB_COSPv2.

Conclusions: Research cohorts, particularly those across different centers within consortia, show significant variability in original histologic type diagnosis. Our IHC-based reclassification produced more homogeneous types with respect to outcome than original type.

Impact: Biomarker-based classification of ovarian carcinomas is feasible, improves comparability of results across research studies, and can reclassify cases which lack reliable original pathology. Cancer Epidemiol Biomarkers Prev; 22(10); 1677–86. ©2013 AACR.

Note: Supplementary data for this article are available at Cancer Epidemiology, Biomarkers & Prevention Online (http://cebp.aacrjournals.org/).

M. Köbel and S.E. Kalloger contributed equally to this work.

Corresponding Author: Susan J. Ramus, Department of Preventive Medicine, University of Southern California/Norris Comprehensive Cancer Center, 1450 Biggy Street, Los Angeles, CA 90033. Phone: 323-442-8110; Fax: 323-442-7787; E-mail: sramus@usc.edu

doi: 10.1158/1055-9965.EPI-13-0391

©2013 American Association for Cancer Research.
Introduction

Studies in recent years have revealed that ovarian carcinoma is not a single disease entity and that histologic type is a reasonable first stratification (1–3). The 5 major histologic/morphologic types of ovarian carcinoma are: high-grade serous carcinoma (HGSC) accounting for 68%, clear cell carcinoma (CCC) for 12%, endometrioid carcinoma for 11%, mucinous carcinoma for 3%, and low-grade serous carcinoma (LGSC) for 3% (4). Histologic types of ovarian carcinoma are characterized by distinct precursor lesions, such as the recently identified serous tubal intraepithelial carcinoma for HGSC and atypical endometriosis for CCC and endometrioid carcinoma, which are associated with distinct molecular alterations (5–7). Because carcinomas of different histologic types originate from different precursor cells, they retain their cell lineage characteristics, which, together with the acquired molecular alterations during oncogenesis, result in specific gene and biomarker expression profiles, as well as a distinct morphologic phenotype (1, 3, 8). Histologic type is a prognostic marker independent of stage (HGSC is associated with the lowest 5-year survival rate) and is a predictive marker for response to standard platinum/paclitaxel chemotherapy, as well as targeted therapies (9–15). Recent studies also provide evidence that several epidemiologic and inherited risk factors are specific to histologic types (16–19).

Although misclassification of histologic type would not affect current clinical management, it is important to achieve a robust histologic classification for ovarian carcinomas according to histologic type for a number of reasons: principally for research into the etiology and prognostic factors in ovarian carcinoma. If pathologists are trained to use refined criteria, interobserver agreement for histologic type with available full slide sets can attain Cohen’s $k$ values of 0.89 (3). However, without specific training, which better reflects current pathology practice, the interobserver agreement is only moderate with Cohen’s $k$ varying between 0.54 and 0.67 (20). A major diagnostic shift has occurred in recent years affecting the classification of carcinomas with glandular architecture and high-grade nuclear features that were formerly classified as high-grade endometrioid carcinoma to now be diagnosed as HGSC (Supplementary Fig. S1; refs. 21, 22). This shift is justified by the fact that those carcinomas are molecularly indistinguishable from the morphologically typical HGSC (23).

We recently suggested an alternative approach to standard morphology-based typing using a nine-marker immunohistochemical (IHC) model termed Calculator for Ovarian Subtype Probability (COSP; ref. 24). COSP incorporated IHC-derived protein expression data, from formalin-fixed paraffin-embedded tissue (FFPE) assembled on tissue microarrays (TMA). Nine markers (CDKN2A, DKK1, HNF1B, MDM2, PGR, TFF3, TP53, VIM, and WT1) were used to predict ovarian carcinoma type in 2 cohorts with differences in tissue handling (24). One COSP model was developed for an archival cohort (A_COSP) assembled with nonmissing clinicopathologic, IHC, and outcome data from the previous FFPE tumor bank cohort (24) as a training set and then used OTTA cases as a testing set.
Local research ethics committees approved all aspects of this study.

**Training set analysis**

We supplemented the previously published tumor bank cohort (24) with 32 additional cases identified from a similar resource using the consultation files of one author (M.A.D.). These additional cases were composed of: CCC \((N = 7)\), endometrioid carcinoma \((N = 6)\), HGSC \((N = 13)\), LGSC \((N = 4)\), and mucinous carcinoma \((N = 2)\). These cases were diagnosed during the same time period (2001–2008), and fixation procedures can be expected to adhere to the same standards as the original tumor bank cohort. As in the original, only cases in which 2 gynecological pathologists independently agreed on histologic type were included, and this histologic type assignment was considered the "reference standard" for the training set. The new training set consisted of 253 cases representing the 5 major histologic types (Table 1).

**Testing set analysis**

The testing set was composed of cases from 3 studies participating in OTTA, including cases enrolled at the Mayo Clinic (MAY; Rochester, MN; \(N = 544\); ref. 30), in the UK Ovarian Cancer Population Study (UKO, \(N = 119\); ref. 31), and the Hormones and Ovarian Cancer Prediction study \((N = 49); \text{ref. 32}\) totaling 712 cases. As with the training set, only cases with nonmissing clinicopathologic and IHC data were included, which resulted in the exclusion of 117 cases which failed IHC for at least one marker (described below; Supplementary Table S1). A further 71 cases with uncertain original histology were excluded from comparisons between IHC-based prediction models and original type. This resulted in a final testing set of 524 cases, for which demographics and outcome data are shown in Table 1 and Supplementary Table S1.

**Immunohistochemistry**

TMAs containing duplicate to quadruplicate 0.6 mm tissue cores were used; 10 sections (4 microns in thickness) were stained for ARID1A, CDKN2A, DKK1, HNF1B, MDM2, PGR, TFF3, TP53, VIM, and WT1 (which were the nine markers from the previously published COSP panel; ref. 24) to which we added ARID1A because of its high specificity for clear cell and endometrioid carcinomas (6, 33). Centralized immunohistochemistry was conducted using the Ventana XT platform (Ventana Medical Systems) according to the standard procedures (protocol details are given in Supplementary Table S2), and scoring was conducted by a single pathologist (M. Köbel). The highest score for a given case was used for analysis. Guidelines for categorizing staining as positive or negative are given in Supplementary Table S2, with some refinement made for scoring cutoffs for HNF1B, TFF3, and VIM compared with the previous study (24). All markers were categorized as positive and negative with the exception of TP53, which was kept as 3 tiers: complete absence, wild-type pattern, and overexpression (3). These data were used to create TB_COSPv2 (described below). The testing set was subjected to A_COSP using the publicly available online calculation (under: http://www.gpec.ubc.ca/index.php?content=papers/ovcasubtype.php) as well as TB_COSPv2. A_COSP and TB_COSPv2-assigned probabilities for each of the 5 major histologic types. The histologic type with the highest probability, even if it was less than 50%, was assigned as the predicted type by A_COSP and TB_COSPv2.

**WTI-assisted core review**

To address what has been previously found to be the predominant misclassification in the typing of ovarian carcinoma, the difference between endometrioid carcinoma and HGSC (21), the testing set was assessed by WTI-assisted core review. Tumor cores on TMA slides that were stained for hematoxylin and eosin (H&E) were reviewed by a single pathologist (M. Köbel) in combination with a corresponding TMA slide that was stained for WTI, and were assigned to one of the 5 major histologic types. In cases with discrepant WTI-assisted core review, resolution was achieved using the WTI-assisted core.

---

**Table 1. Characteristics of study populations**

<table>
<thead>
<tr>
<th>Characteristic</th>
<th>Training set</th>
<th>Testing set</th>
<th>(P)</th>
</tr>
</thead>
<tbody>
<tr>
<td>(N)</td>
<td>253</td>
<td>524</td>
<td>N/A</td>
</tr>
<tr>
<td>Age, year Mean (range)</td>
<td>60.2 (28–99)</td>
<td>61.5 (21–93)</td>
<td>0.17</td>
</tr>
<tr>
<td>Original histologic type</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>HGSC</td>
<td>176 (69%)</td>
<td>336 (64%)</td>
<td></td>
</tr>
<tr>
<td>EC</td>
<td>30 (12%)</td>
<td>97 (18%)</td>
<td>0.080</td>
</tr>
<tr>
<td>CCC</td>
<td>31 (12%)</td>
<td>47 (9%)</td>
<td></td>
</tr>
<tr>
<td>MC</td>
<td>7 (3%)</td>
<td>22 (4%)</td>
<td></td>
</tr>
<tr>
<td>LGSC</td>
<td>9 (4%)</td>
<td>22 (4%)</td>
<td></td>
</tr>
<tr>
<td>Stage</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>I/II</td>
<td>71 (28%)</td>
<td>147 (28%)</td>
<td>0.97</td>
</tr>
<tr>
<td>III/IV</td>
<td>182 (72%)</td>
<td>377 (72%)</td>
<td></td>
</tr>
</tbody>
</table>

Abbreviations: EC, endometrioid carcinoma; MC, mucinous carcinoma.
review for the majority. For example, if 2 of the 3 cores were called HGSC, the WT1-assisted core review was HGSC. A sixth category, "other," was used for cases that could not be histologically classified on the core, for cases where a common assessment could not be reached on a majority of the cores assessed for any single case.

**Evaluation of agreement between type assessments**

Given the lack of a clear gold standard, we created 2 internal references for histologic type. The first was based on the assumption that, in the cases we studied, agreement between TB_COSPv2 and the original type is likely correct. Second, WT1-assisted core review was used as a "tie-breaker" to arbitrate the cases in the testing set with disagreement between original type and TB_COSPv2. For each case, there were 3 possible outcomes: (i) WT1-assisted core review agreed with original type (TB_COSPv2 error assumed), (ii) WT1-assisted core review agreed with TB_COSPv2 (original type diagnosis error assumed), (iii) WT1-assisted core review did not agree with either (declared as "unclassifiable"). Thus, by combining the 3 methods (original type, TB_COSPv2 and WT1-assisted core review), each case was either assigned to a certain histologic type by at least 2 out of the 3 methods or was unclassifiable because the 3 methods disagreed (Fig. 1).

**Statistical analysis**

Nominal logistic regression modeling was used to generate prediction equations, as previously described (24), using the 10-marker panel on the training set. For model predictions, a receiver operator characteristic area under the curve (ROC AUC) for each histologic type category was calculated. The model predictions were then tested for external validity by application to the testing set. To determine agreement between the 3 assessments of histologic type, Cohen’s k statistics were calculated (34). Qualitatively, a k value of 0.2 to 0.4 indicates minimal agreement, 0.4 to 0.6 indicates moderate agreement, 0.6 to 0.8 indicates substantial agreement, and 0.8 to 1.0 indicates excellent agreement (34). Two-way unsupervised hierarchical clustering was conducted using the Wald algorithm for histologic type assignments by different methods. To evaluate histologic type assessments with overall survival, survival curves were generated using the
Results

Training set

In an effort to improve upon our previous work in ovarian carcinoma histologic type classification, we examined an expanded version of a previously used training set (24) and improved the internal validity of the immunohistochemistry-based type prediction by adding ARID1A and refining the scoring categories of HNF1B, TFF3, and VIM, the details of which are shown in Supplementary Table S2. The new prediction algorithm (TB_COSPv2) was compared with both of the previous algorithms (TB_COSPv1 and A_COSP). Supplementary Table S3 shows the overall superiority of the TB_COSPv2 algorithm compared with the previous algorithms produced from our earlier efforts as quantified by ROC AUC. This internal validation showed a 98% concordance between TB_COSPv2 and original type, yielding only 5 misclassifications within the training set.

Testing set

External validity was assessed by the application of TB_COSPv2 to the OTTA testing set. The testing did not significantly differ from the training set with respect to patient age, histologic type, and stage distribution. The mean diagnosis year of the tumor blocks of the training set was 2006, which was significantly older by one year compared with the testing set (P = 0.021). Still, all training set cases were diagnosed within the time range of the testing set cases suggesting no differences in tissue handling due to overlapping time periods. Supplementary Table S4 shows the categorized expression results for each marker by histologic type. Statistically significant differential marker expression across the training and testing sets was seen for two markers in HGSC: DKK1 and MDM2; and for MDM2 alone in EC. Table 2 shows cross-tabulations and agreement between original type and TB_COSPv2. The overall agreement with original diagnosis was 73% but varied across the types. This agreement was highest for HGSC with 86% and lowest for mucinous carcinoma and LGSC both with 36%. Most reclassifications by TB_COSPv2 were made in endometrioid carcinomas, which were reclassified primarily to HGSC (HGSC, N = 38; non-HGSC, N = 10) or were reclassified from other histologic types to endometrioid carcinoma (N = 40). We also applied the previous A_COSP to the OTTA training set, and cross-tabulations of A_COSP with original type showed a similar agreement rate with original diagnosis (75%; Supplementary Table S5). To objectively assess which prediction equation was superior (A_COSP vs. TB_COSPv2) in the testing set, we chose to use the prognostic difference between HGSC and endometrioid carcinoma as a measure of proper classification based on the fact that the latter should have a superior prognosis to the former. In the examination of the prognostic significance, we had to exclude cases derived from the Hormones and Ovarian Cancer Prediction study due to a lack of outcome data. This exclusion resulted in the removal of 10 cases with original type of CCC (N = 1) and endometrioid carcinoma (N = 9). Supplementary Table S6 shows the HRs for histologic types in univariate analysis. The HR for endometrioid carcinoma compared with the HGSC reference was smaller for TB_COSPv2 (HR = 0.35; 95% CI, 0.22–0.52) compared with A_COSP (HR = 0.43; 95% CI, 0.28–0.64). Because of the superior survival difference, the following analyses were restricted to TB_COSPv2.

TB_COSPv2 had a 73% agreement with the original type. For the 27% (N = 142) discordant cases, the WT1-assisted core review was used as an arbiter to resolve the discrepancy (Fig. 1). By doing so, 54% of cases discordant between the original type and TB_COSPv2 were determined to have an error in original type (N = 76 or 15% of all cases). Twenty-four percent of discordant cases were...
found to have an error for TB_COSPv2 prediction (N = 34 or 6% of all cases), and 22% of discordant cases were declared as "unclassifiable" (N = 32 or 6% of all cases; Fig. 1). The probabilities derived by TB_COSPv2 were not significantly different for cases with original type error, TB_COSPv2 error, or unclassifiable (P = 0.51). The most common TB_COSPv2 error was the prediction of endometrioid carcinoma (N = 16/34), which were equally distributed across the types and predicted as HGSC (N = 4), LGSC (N = 4), mucinous carcinoma (N = 4), and CCC (N = 4) by the other 2 methods.

Table 3 shows the patterns of agreement between the original type and after the WT1-assisted core review was applied to resolve discrepancies between original type and TB_COSPv2. After excluding the 32 unclassifiable cases, the agreement rate with original type increased to 85% (N = 416/492) compared with 73% (N = 382/524) for TB_COSPv2 alone because 34 cases were added where the WT1-assisted core review agreed with original type. The agreement rate was highest for HGSC with 94% and intermediate for CCC and LGSC with 80%. Endometrioid carcinoma, however, showed the lowest level of agreement with 59%, which was due to a large group (N = 32) with systematic disagreement between original endometrioid carcinoma type and HGSC by both the other methods (TB_COSPv2 and WT1-assisted core review).

To provide justification for the reclassification, outcome analysis were conducted with 4 type assessments: original type (100% of cases), TB_COSPv2 (100%), agreement between original type and TB_COSPv2 (73%), and agreement between 2 out of 3 methods of assessment (94% of cases; Fig. 1). Kaplan–Meier curves are displayed in Fig. 2 and the 5-year overall survival rates (5y-OS) for each histologic type are shown in Supplementary Table S7.
## Table 4. Multivariate Cox model in the testing set using different methods of type assessment

<table>
<thead>
<tr>
<th>Method of Type Assessment</th>
<th>Original Type</th>
<th>TB_COSPv2</th>
<th>Original Type</th>
<th>TB_COSPv2</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Site</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>MAY versus UKO</td>
<td>1.02 (0.76-1.34)</td>
<td>1.02 (0.75-1.30)</td>
<td>1.05 (0.73-1.50)</td>
<td>1.02 (0.75-1.30)</td>
</tr>
<tr>
<td>Age Older than median versus median or younger</td>
<td>1.45 (1.14-1.87)</td>
<td>1.42 (1.13-1.82)</td>
<td>1.16 (1.10-1.23)</td>
<td>1.00 (0.90-1.12)</td>
</tr>
<tr>
<td>Histologic type</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>HGSC (Reference)</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>LGSC</td>
<td>0.85 (0.40-1.75)</td>
<td>0.86 (0.40-1.84)</td>
<td>0.95 (0.45-2.06)</td>
<td>0.50 (0.24-1.04)</td>
</tr>
<tr>
<td>CCC</td>
<td>1.25 (0.72-2.14)</td>
<td>1.23 (0.67-2.21)</td>
<td>1.11 (0.64-1.92)</td>
<td>0.86 (0.45-1.70)</td>
</tr>
<tr>
<td>MC</td>
<td>1.46 (0.42-3.85)</td>
<td>1.46 (0.43-3.97)</td>
<td>1.46 (0.46-3.85)</td>
<td>1.00 (0.46-2.14)</td>
</tr>
<tr>
<td><strong>NOTES</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Abbreviations: EC, endometrioid carcinoma; MC, mucinous carcinoma.</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

The 5y-OS of HGSC, the largest group, varied only slightly across type assessments, probably because only a small fraction of HGSCs were reclassified by TB_COSPv2. The 5y-OS rates for CCC were also relatively stable across the methods. In contrast, endometrioid carcinoma, the second largest group, showed an improved 5y-OS rate of 87% when both the original type and TB_COSPv2 prediction agreed, compared with only 59% when the original type was used. Reclassified endometrioid carcinoma by the TB_COSPv2 prediction was associated with a 71% 5y-OS rate, and this improved to 80% when agreement between 2 out of 3 methods of assessment predicted the endometrioid carcinoma histology. Results of the Cox proportional hazards models adjusted for stage, age, and study site are depicted in Table 4. The survival difference between endometrioid carcinoma and HGSC was statistically insignificant with the original type (HR 0.85; 95% CI, 0.57–1.23; \( P = 0.41 \)) but attained statistical significance when the other methods of type assessment were applied, for example, TB_COSPv2 (HR 0.60; 95% CI, 0.37–0.93; \( P = 0.021 \)), suggesting improved group homogeneity.

When we compared clinico-pathologic parameters among cases classified as endometrioid carcinoma according to the different methods of type assessment, those with agreement between the original type and TB_COSPv2 prediction, compared with the original type alone, had lower proportions of high-stage disease (20% vs. 46%; \( P = 0.0022 \)), grade 3 (25% vs. 50%; \( P = 0.050 \)), and WT1 expression (6% vs. 40%; \( P < 0.0001 \); Supplementary Table S8), which is consistent with expected endometrioid carcinoma characteristics (21).

Finally, 71 cases with uncertain original diagnoses were included in TB_COSPv2 predictions (Supplementary Table S9). Of the 35 cases known to be serous but which were ungraded (hence could not be assigned to HGSC or LGSC), 28 were predicted to be HGSC and 7 to be LGSC. Of the remaining 36, the most common predicted types were HGSC (\( N = 23, 64\% \)) and endometrioid carcinoma (\( N = 8, 22\% \)). Almost all "other" or undifferentiated carcinomas classified as HGSC. Mixed carcinoma split into HGSC, endometrioid carcinoma, and CCC. The predicted types showed the expected 5y-OS rates of 39% for HGSC and 71% for endometrioid carcinoma, providing an example of the viability of using immunohistochemistry when the original pathology is unclear.

### Discussion

This study shows the feasibility of using IHC classifiers such as the TB_COSPv2 prediction model for improved classification of histologic type in research using TMA cohorts. TMA technology has become popular for biomarker interrogation and can overcome the remarkable heterogeneity in histologic type assignment across cohorts that use combinations of original pathology report, local or central pathology review, or full slide or selected slide review. Currently, very few studies rely on tumor biomarkers for histologic classification purposes. However, in light of the recent acceptance that ovarian...
carcinoma types are essentially distinct diseases, and that misclassification of histologic type can confound results (1), it is important to achieve a standardized and reproducible system/method of subclassification. Morphologic review of slides suffers from high retrieval costs for slides, constraints of pathologists’ time, and interobserver variation between pathologists. Therefore, an approach that is able to directly use TMAs can be advantageous. Our results show that the application of IHC biomarker assessment can be highly accurate and able to uncover misclassified cases. The TB_COSPv2 model showed a 98% concordance with the morphologic “gold standard” type in the training set and a 73% concordance with the original type in the OTTA testing set. In the testing set, WT1-assisted core review revealed that most cases of disagreement (about 15% of all cases) were likely due to original type misclassification, whereas TB_COSPv2 error occurred only in 6% of cases. Another 6% of cases remained unclassifiable with this approach. We acknowledge that the amount of tissue available for WT1-assisted core review, being limited, may not be representative. But the amount of tissue in TMA cores is comparable with diagnostic material available from a cell block obtained by paracentesis or a core biopsy taken from an omental cake in current clinical practice before commencing neoadjuvant chemotherapy (36). Nevertheless, WT1-assisted core review increased the agreement rate with the original type from 73% to 85% of cases, and can therefore be used as a safeguard to avoid TB_COSPv2 errors in discordant cases.

By considering the final reclassification assessment in Table 3, the pattern of disagreement is mostly random. The minor pattern of systematic disagreement could be due to the tendency for pathologists to overcall endometrioid carcinoma in the original type that was consequently reclassified as HGSC by TB_COSPv2 and WT1-assisted core review. Reclassification of endometrioid carcinoma to HGSC is an expected result, which has been shown previously in a morphologic review of a large population-based series in Canada (21). The differential diagnosis of HGSC and endometrioid carcinoma, particularly in high-grade cases, is still a matter of controversy among pathologists (22). But many now advocate that the vast majority of cases that show high-grade nuclear atypia regardless of architectural features should be diagnosed as HGSC (2, 37–40). Hence, the criteria for diagnosing HGSC versus endometrioid carcinoma have evolved over time. This study uses biomarkers for reclassification; therefore, it is difficult to improve on tumor classification systems and to avoid circular reasoning. In the current study, 40% of original endometrioid carcinoma showed WT1 expression, which decreased to 6% (P < 0.0001) when any 2 of the 3 assessment methods were in agreement to predict endometrioid carcinoma histology. Although this would be expected, because WT1 expression was a component of both the TB_COSPv2 algorithm and the WT1-assisted core review, a similar shift was observed for reclassified endometrioid carcinoma based on morphology only (21, 41).

Furthermore, reclassified endometrioid carcinoma was also less likely to be categorized as grade 3 and FIGO stage III or higher, suggesting improved homogeneity and consistency with expected patterns.

We considered survival as an objective outcome parameter to provide validity for the specific reclassification of endometrioid carcinoma to HGSC. The 5y-OS rate for the original type of endometrioid carcinoma was only 59%, and this survival became statistically significantly different compared with HGSC following the reclassification of HGSCs to endometrioid carcinomas by TB_COSPv2. Survival-based outcome measures are highly stage dependent and almost half of the original diagnoses of endometrioid carcinoma were high stage compared with approximately 27% when any 2 of the 3 assessment methods agreed. One could argue that our findings for endometrioid carcinoma are due to the exclusion of high-stage disease; however, associations remained statistically significant even when the adjustment was made for stage in multivariate models. Overall, we believe that these outcome changes justified our approach.

Reclassification of histologic types is much more random with respect to the other types. TB_COSPv2 shifted 34% of original CCC into other categories (mostly HGSC) and reclassified mostly HGSC and endometrioid carcinoma to compose 35% of reclassified CCC. The majority of these reclassifications were confirmed by WT1-assisted core review. This shift resulted in no relevant change in the 5y-OS rate (56%–65%), probably due to the intermediate outcome of CCC and the random nature of reclassification events. Therefore, outcome may not be used as a surrogate for histologic type for all instances or in individual cases. The WT1-assisted core review was particularly helpful to increase the agreement rate of mucinous carcinoma and LGSC from 36% each by TB_COSPv2 alone to 68% or 80%, respectively. Further research in larger TMA collections of these minority types is needed.

This exercise not only shows the feasibility of retrospective reclassification of histologic type for accuracy but offers economic efficiencies as well. With an average cost of $40 per IHC slide, the 10-marker assay would cost approximately CAD 400 per TMA block plus scoring and analytic time at a typical medical research institution. For this study, cases were assembled on 8 TMA resulting in a total assay cost of CAD 3,200 (or CAD 6 per case). There are also limitations to this method, including case loss due to the requirement for a complete biomarker dataset for TB_COSPv2. We lost 117 (16%) of the 712 testing set cases for classification. An imputation of missing data would be very desirable to avoid case loss, as would use of triplicate cores and optimization of TMA design to minimize core drop out. On the other hand, 71 (10%) of the test set cases, which did not have a diagnosis of 1 of the 5 major histological types, could be reclassified by TB_COSPv2. Notably, TB_COSPv2 could differentiate between high-grade and low-grade serous, reclassify almost all “other” or undifferentiated carcinomas classified as HGSC, and assign mixed carcinoma as HGSC,
endometrioid carcinoma, or CCC. Another limitation of TMAs is with regard to mixed carcinomas. Because it is often not feasible to take multiple cores from different tumor components of a mixed carcinoma, full sections will be necessary to accurately classify mixed carcinomas. Reproducibility of IHC stains has been stressed in recent years. Although our data show that the majority of IHC markers used in this study showed a constant expression rate within types across sets (some markers even within 1% range), we identified 2 problem markers where more reliable antibodies are needed (MDM2 and DKK1).

This study sheds light on issues of heterogeneity in pathologic subclassification of ovarian carcinomas. Our reclassification exercise shows that outcome of endometrioid carcinoma is largely driven by accurate diagnosis. The issues of heterogeneous type assignment could have a major effect on large-scale tissue-based biomarker studies, particularly within the uncommon types. This study presents a research tool that can be used for immunohistochemistry-based ovarian carcinoma typing. Further advances that will increase robustness of the classification include digital pathology enabling assessment of H&E morphology with simultaneous assessment of biomarker expression on a single screen, multiplex immunofluorescence methods, and use of additional molecular markers such as somatic mutation status. Application of these approaches could provide a more objective frame of reference toward a biologic stratification of ovarian carcinomas.

Disclosure of Potential Conflicts of Interest

M.A. Duggan has commercial research support from Hologic and is a consultant/advisory board member of BD. U. Menon has expert testimony from Abxodia Ltd. No potential conflicts of interest were disclosed by the other authors.

References


