Biomarker-based ovarian carcinoma typing: a histological investigation in the Ovarian Tumor Tissue Analysis consortium


1Department of Pathology and Laboratory Medicine, University of Calgary and Calgary Laboratory Services, Calgary, AB, Canada
2Department of Pathology and Laboratory Medicine, University of British Columbia, Vancouver, BC, Canada
3Department of Population Health Research, Alberta Health Services-Cancer Care and Department of Medical Genetics, University of Calgary, Calgary, AB, Canada
4Department of Medical Oncology, Mayo Clinic, Rochester, MN, USA
5Biostatistics and Informatics Shared Resource, University of Kansas Medical Center, Kansas City, KS, USA
6Department of Laboratory Medicine and Pathology, Division of Anatomic Pathology, Mayo Clinic, Rochester, MN, USA
7Department of Health Science Research, Division of Biomedical Statistics and Informatics, Mayo Clinic, Rochester, MN, USA
8Department of Laboratory Medicine and Pathology, Division of Experimental Pathology and Laboratory Medicine, Mayo Clinic, Rochester, MN, USA
9University of Texas School of Public Health, Houston, TX, USA
10Roswell Park Cancer Institute, Buffalo, NY, USA
11Women’s Cancer Research Center, Magee-Womens Research Institute and University of Pittsburgh Cancer Institute, Department of Obstetrics, Gynecology and Reproductive Sciences,
Running Title: Biomarker based ovarian carcinoma typing, an OTTA study.

Key Words: ovarian cancer, type, biomarker, immunohistochemistry, typing

Correspondence to:
Susan J. Ramus, Department of Preventive Medicine, University of Southern California / Norris Comprehensive Cancer Center, Harlyne Norris Research Tower, 1450 Biggy Street, Office 2517K, Los Angeles, CA 90033, USA.
Phone: 323-442-8110, Fax: 323-442-7787, sramus@usc.edu

Grant support: This study was supported by a Calgary Laboratory Service research grant RS11-508, R01-CA122443, P50-CA136393, the Mayo Foundation, the Fred C. and Katherine B. Andersen Foundation, K07-CA80668, DAMD17-02-1-0669, NIH/National Center for Research Resources/General Clinical Research Center grant MO1-RR000056, The Eve Appeal (The Oak
Foundation), and the National Institute for Health Research University College London Hospitals Biomedical Research Centre.

**Disclosure of Potential Conflict of Interest:** Usha Menon has a financial interest through UCL Business and Abcodia Ltd in the third party exploitation of clinical trials biobanks which have been developed through the research at UCL. The other authors declare no conflict of interest.
Abstract

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

Methods: Tissue microarrays (TMAs) and clinical data were available for 524 pathologically confirmed ovarian carcinomas. Centralized IHC was performed for ARID1A, CDKN2A, DKK1, HNF1B, MDM2, PGR, TP53, TFF3, VIM, and WT1, and three histological type assessments were compared: the original pathologic type, an IHC-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 (EC) 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 EC compared to HGSC became significant (RR 0.60, 95% 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 centres within consortia, show significant variability in original histological 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.
Introduction

Studies in recent years have revealed that ovarian carcinoma is not a single disease entity and that histological type is a reasonable first stratification (1-3). The five major histological/morphological types of ovarian carcinoma are: high-grade serous (HGSC) accounting for 68%, clear cell (CCC) for 12%, endometrioid (EC) for 11%, mucinous (MC) for 3%, and low-grade serous (LGSC) for 3% (4). Histological 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 EC, which are associated with distinct molecular alterations (5-7). Because carcinomas of different histological 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 morphological phenotype (1, 3, 8).

Histological 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 histological types (16-19).

While misclassification of histological type would not affect current clinical management, it is important to achieve a robust histologic classification for ovarian carcinomas according to histological 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, inter-
observer agreement for histological type with available full slide sets can attain Cohen’s $\kappa$ values of 0.89 (3). However, without specific training, which better reflects current pathology practice, the inter-observer agreement is only moderate with Cohen’s $\kappa$ 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 EC to now be diagnosed as HGSC (Figure S1) (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) (24). COSP incorporated IHC derived protein expression data, from formalin fixed paraffin embedded tissue (FFPE) assembled on tissue microarrays (TMAs). Nine markers (CDKN2A, DKK1, HNF1B, MDM2, PGR, TFF3, TP53, VIM, and WT1) were used to predict ovarian carcinoma type in two cohorts with differences in tissue handling (24). One COSP model was developed for an archival cohort (A_COSP) assembled from samples collected from 1984-2000 when tissue fixation procedures were less standardized than today; this showed very good agreement with expert reviewed morphological histotype (Cohen’s $\kappa = 0.85$) (24) and was subsequently validated on archival clinical trial material (25). Another COSP model was developed on the corresponding FFPE tissue of a tumor bank cohort (TB_COSP) consisting of more contemporarily-handled samples diagnosed between 2001-2008; this also showed substantial agreement with expert reviewed morphological histotype (Cohen’s $\kappa = 0.78$) (24). Performance of the latter TB_COSP suffered from a low sensitivity to detect EC, CCC and
Despite this shortcoming, the utility of TB_COSP is greater for modern research cohorts since most contemporary cases are processed according to standardized tissue handling and fixation procedures that are common across pathology departments (26).

The recently formed international Ovarian Tumor Tissue Analysis (OTTA) consortium includes a variety of tissue-based ovarian cancer research studies combined with the goal of understanding factors related to etiology and outcome of the disease. Recently, we have shown that reclassification of histological type in an effort to improve disease homogeneity can strengthen risk associations for some ovarian cancer histological types (27). This improved classification may be useful for future research efforts, as current studies often suffer from heterogeneity in diagnostic criteria, adherence to an outdated WHO standard (28), and do not reflect modern classification systems of the five major histological types (3). The main objective of this study was to reclassify TMA cohorts in OTTA using a combined biomarker and morphology approach. Without having access to the review of full pathology slide sets, we were limited to the TMA resource, where we assessed a 10 marker immunohistochemical classifier (TB_COSP) and a morphological review of TMA cores, which was done with the knowledge of WT1 expression status (WT-assisted core review). The serous cell lineage WT1 marker was selected because it represents the most informative sole biomarker for the most anticipated problem of distinguishing EC from HGSC (21). The specific aims of the study were: i) refine TB_COSP, ii) compare the internal validity of the newly refined TB_COSP version 2 (TB_COSPv2) with the previously reported TB_COSPv1; iii) evaluate agreement of the original type with TB_COSPv2; iv) arbitrate disagreement of TB_COSPv2 with original type using
WT1-assisted core review, and v) compare associations with overall survival across type assessment strategies.
Material and Methods

Study Design

In order to refine TB_COSP, we used an expanded set of cases with non-missing clinico-pathologic, IHC, and outcome data from the previous FFPE tumour 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 with 32 additional cases identified from the same resource (24) using the consultation files of one author (MAD). These additional cases were composed of: CCC (N=7), EC (N=6), HGSC (N=13), LGSC (N=4), and MC (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 two gynaecological pathologists independently agreed on histological type were included, and this histological type assignment was considered the ‘reference standard’ for the training set. The new training set consisted of 253 cases representing the five major histological types (Table 1).

Test Set Analysis

The test set was composed of cases from three studies participating in OTTA, including cases enrolled at the Mayo Clinic (MAY, N=544) (29, 30) and the Hormones and Ovarian Cancer Prediction study (HOP, N=49) (31) totalling 712 cases. As with the training set, only cases with non-missing clinico-pathologic and IHC data were included which resulted in the exclusion of 117 cases which failed IHC for at least one marker (described below) (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 Tables 1 & S1.

**Immunohistochemistry (IHC)**

TMA’s containing duplicate to quadruplicate 0.6 mm tissue cores were used; ten 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) (24) to which we added ARID1A because of its high specificity for clear cell and endometrioid carcinomas (6, 32). Centralized IHC was performed using the Ventana XT platform (Ventana Medical Systems, Tuscan Arizona) according to standard procedures (protocol details are given in Table S2) and scoring was conducted by a single pathologist (MK). The highest score for a given case was used for analysis. Guidelines for categorizing staining as positive or negative are given in Table S2, with some refinement made for scoring cut-offs for HNF1B, TFF3 and VIM compared to the previous study (24). All markers were categorized as positive and negative with the exception of TP53, which was kept as three 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](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 five major histological types. The histological 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

In order to address what has been previously found to be the predominant misclassification in the typing of ovarian carcinoma, the differential between EC & HGSC (21), the testing set was assessed by WTI-assisted core review. Tumor cores on TMA slides that were stained for H&E were reviewed by a single pathologist (MK) in combination with a corresponding TMA slide that was stained for WTI, and were assigned to one of the five major histological types. In cases with discrepant core assignments, resolution was achieved using the core assignment for the majority. For example, if two of the three cores were called HGSC, the core review assessment 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 two internal references for histological 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. Secondly, WTI-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 three possible outcomes: i) WTI-assisted core review agreed with original type (TB_COSPv2 error assumed), ii) WTI-assisted core review agreed with TB_COSPv2 (original type diagnosis error assumed), iii) WTI-assisted core review did not agree with either (declared as ‘unclassifiable’). Thus, by combining the three methods (original type, TB_COSPv2 and WTI-assisted core review), each case was either assigned to a certain histological type by at least two out of three methods or was unclassifiable because the three methods disagreed (Figure 1).
Statistical Methods

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 histological type category was calculated. The model predictions were then tested for external validity by application to the test set. To determine agreement between the three assessments of histological type, Cohen’s κ statistics were calculated (33). Qualitatively, a κ 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 (33). Two-way unsupervised hierarchical clustering was performed utilizing the Wald algorithm for histological type assignments by different methods. To evaluate histological type assessments with overall survival, survival curves were generated using the Kaplan-Meier Method and compared with the Wilcoxon test. We used the Cox proportional hazards model to estimate the hazard ratios (HR) and 95% confidence intervals (CI) for overall survival, accounting for left truncation. The covariates included in the Cox model were study site (MAY versus UKO), FIGO stage (stage I and II versus stage III and IV), and age (older than median versus median and younger). All statistical analyses were computed with JMP version 10.0 (SAS Institute, Cary, NC U.S.A.). This study adhered to the REMARK guidelines for the reporting of biomarker studies (34).
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 IHC based type prediction by adding ARID1A and refining the scoring categories of HNF1B, TFF3 and VIM, the details of which are shown in Table S2. The new prediction algorithm (TB_COSPv2) was compared to both of the previous algorithms (TB_COSPv1 & A_COSP). Table S3 demonstrates the overall superiority of the TB_COSPv2 algorithm compared to the previous algorithms produced from our earlier efforts as quantified by ROC AUC. This internal validation demonstrated 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, histological type and stage distribution. The mean age of the tumor blocks of the training set was 2006, which was significantly older by one year compared to the training set (p=0.021). Still, all training set cases were diagnosed with the time range of the testing set cases suggesting no differences in tissue handling due to overlapping time periods. Table S4 shows the categorized expression results for each marker by histological type. Statistically significant differential marker expression across the training and testing sets was seen for two markers: DKK1 and MDM2. Table 2 shows cross-tabulations and agreement between original type and TB_COSPv2.
The overall agreement with original diagnosis was 73% but varied across types. This agreement was highest for HGSC with 86% and lowest for MC and LGSC with 36% respectively. Most reclassifications by TB_COSPv2 were made in EC’s which were reclassified primarily to HGSC (HGSC N=38, non-HGSC N=10) or were reclassified from other histological types to EC (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%, Table S5). In order to objectively assess which prediction equation was superior (A_COSP versus TB_COSPv2) in the testing set, we chose to use the prognostic difference between HGSC & EC 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 EC (N = 9). Table S6 shows the hazard ratios for histological types in univariate analysis. The hazard ratio for EC compared to the HGSC reference was smaller for TB_COSPv2 (HR=0.35, 95% confidence interval 0.22-0.52) compared to A_COSP (HR=0.43, 95% confidence interval 0.28-0.64). Because of the superior survival difference, the following analysis 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 (Figure 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) (Figure 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 EC (N=16/34), which were equally distributed across the types and predicted as HGSC (N=4), LGSC (N=4), MC (N=4) and CCC (N=4) by the other two 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 to 73% (N=382/524) for TB_COSPv2 alone because in 34 cases were added where the WT1-assisted core review agreed with original type. The agreement rate was highest for HGSC with 93% and intermediate for CCC and LGSC with 80%. EC, however, showed the lowest level of agreement with 59%, which was due to a large group (N=32) with systematic disagreement between original EC type and HGSC by both other methods (TB_COSPv2 and WT1-assisted core review).

In order to provide justification for the reclassification, outcome analysis were performed with four type assessments: original type (100% of cases), TB_COPSv2 (100%), agreement between original type and TB_COSPv2 (73%) and agreement between 2 out of 3 methods of assessment (94% of cases, Figure 1). Kaplan-Meier curves are displayed in Figure 2 and the 5-year overall survival rates (5y-OS) for each histological type are shown in Table S7. The 5y-OS of HGSC, the largest group, varied only slightly across type assessments, probably because only a small
fraction of HGSC’s were reclassified by TB_COSPv2. The 5y-OS rates for CCC were also relatively stable across the methods. In contrast, EC, the second largest group, showed an improved 5y-OS rate of 87% when both the original type and TB_COSPv2 prediction agreed, compared to only 59% when the original type was used. Reclassified EC 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 EC histology. Results of the Cox proportional hazards models adjusted for stage, age, and study site are depicted in Table 4. The survival difference between EC 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, e.g. TB_COSPv2 (HR 0.60, 95% CI 0.37-0.93; p=0.021), suggesting improved group homogeneity.

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

Finally, 71 cases with uncertain original diagnoses were included in TB_COSPv2 predictions (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 EC (N=8, 22%).
Almost all ‘other’ or undifferentiated carcinomas classified as HGSC. Mixed carcinoma split into HGSC, EC and CCC. The predicted types showed the expected 5y-OS rates of 39% for HGSC and 71% for EC, providing an example of the viability of using IHC when the original pathology is unclear.

Discussion

This study demonstrates the feasibility of using IHC classifiers such as the TB_COSPv2 prediction model for improved classification of histological type in research utilizing TMA cohorts. TMA technology has become popular for biomarker interrogation and can overcome the remarkable heterogeneity in histological 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 histological classification purposes. However, in light of the recent acceptance that ovarian carcinoma types are essentially distinct diseases, and that misclassification of histological type can confound results (1), it is important to achieve a standardized and reproducible system/method of sub-classification. Morphological review of slides suffers from high retrieval costs for slides, constraints of pathologists’ time and inter-observer variation between pathologists. Therefore, an approach that is able to directly utilize 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 morphological ‘gold standard’ type in the training set and a 73% concordance with the original type in the OTTA testing set. In the training 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 to 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 (35). Nevertheless, WT1-assisted core review increased the agreement rate with the original type from 73% to 85% of cases can be 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. An example for the minor pattern of systematic disagreement was identified as the tendency for pathologists to overcall EC in the original type that were consequently reclassified as HGSC by TB_COSPv2 and WT1-assisted core review. Reclassification of EC to HGSC is an expected result, which has been shown previously in a morphological review of a large population-based series in Canada (21). The differential diagnosis of HGSC and EC, particularly in higher 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, 36-39). Hence, the criteria for diagnosing HGSC versus EC 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 EC showed WT1 expression, which decreased to 6% (p<0.0001) when any two of the three assessment methods were in agreement to predict EC histology. Although this would be expected, since WT1 expression was a component
of both the TB_COSPv2 algorithm and the WT1-assisted core review, a similar shift was observed for reclassified EC based on morphology only (21, 40). Furthermore, reclassified EC 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 outcome as an objective outcome parameter to provide validity for the specific reclassification of EC to HGSC. The 5y-OS rate for the original type of EC was only 59%, and this survival became statistically significantly different compared to HGSC following the reclassification of HGSC’S to EC’s by TB_COSPv2. Survival based outcome measures are highly stage-dependent and almost half of the original diagnoses of EC were high stage compared to ~27% when any two of the three assessment methods agreed. One could argue our findings for EC are due to the exclusion of high stage disease; however, associations remained statistically significant even when adjustment was made for stage in multivariate models. Overall, we believe that these outcome changes justified our approach.

Reclassification of histological 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 EC 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 histological type for all instances or in individual cases. The WT1-assisted core review was particularly helpful to increase the agreement rate of MC 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 histological 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 eight 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 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 test set cases, which did not have a diagnosis of one of the five 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, EC 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 immunohistochemical stains has been stressed in recent years. While our data show 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 two problem markers where more reliable antibodies are needed (MDM2 and DKK1).
This study sheds light on issues of heterogeneity in pathological subclassification of ovarian carcinomas. Our reclassification exercise shows that outcome of EC 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 IHC-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 towards a biological stratification of ovarian carcinomas.
Acknowledgments

We gratefully thank Shuhong Liu for constructing a training set tissue microarray, and Karin Goodman, Ashley Pitzer, and the Mayo Clinic Medical Genome Facility.
References

Figure Legends

Figure 1 Histological type assignments in the testing set

Figure 2 Overall survivals of the testing set shown in univariate Kaplan-Meier analysis.

Red is endometrioid carcinoma, Blue is high-grade serous carcinoma, yellow is clear cell carcinoma, green is mucinous carcinoma, light blue is low-grade serous carcinoma. P-values represent Wilcoxon test of survival difference across type assessments: A - original type, B - TB_COSPv2, C – agreement between original type and TB_COSPv2, D – agreement between 2 out of 3 methods of assessment (original type, TB_COSPv2 and WT1-assisted core review)
Table 1 Characteristics of study populations

<table>
<thead>
<tr>
<th></th>
<th>Training set</th>
<th>Testing set</th>
<th>p-value</th>
</tr>
</thead>
<tbody>
<tr>
<td>N</td>
<td>253</td>
<td>524</td>
<td>N/A</td>
</tr>
<tr>
<td>Diagnosis Year</td>
<td>Median (Range) 2001-2008</td>
<td>(2000-2009)</td>
<td>0.021</td>
</tr>
<tr>
<td>Age (years)</td>
<td>Mean (Range) 60.2 (28-99)</td>
<td>61.5 (21-93)</td>
<td>0.17</td>
</tr>
<tr>
<td>Original histological type</td>
<td></td>
<td></td>
<td>0.080</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></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>0.97</td>
</tr>
<tr>
<td>I/II</td>
<td>71 (28%)</td>
<td>147 (28%)</td>
<td></td>
</tr>
<tr>
<td>III/IV</td>
<td>182 (72%)</td>
<td>377 (72%)</td>
<td></td>
</tr>
</tbody>
</table>

HGSC-high-grade serous carcinoma, EC-endometrioid carcinoma, CCC-clear cell carcinoma, MC-mucinous carcinoma, LGSC-low-grade serous carcinoma based on histology.
<table>
<thead>
<tr>
<th>TB_COSPv2 prediction</th>
<th>HGSC</th>
<th>EC</th>
<th>CCC</th>
<th>MC</th>
<th>LGSC</th>
<th>Total</th>
<th>Concordance rate</th>
<th>K (95% CI)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Original type</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>HGSC</td>
<td>288</td>
<td>17</td>
<td>7</td>
<td>9</td>
<td>15</td>
<td>336</td>
<td>86%</td>
<td>0.497 (0.432-0.561)</td>
</tr>
<tr>
<td>EC</td>
<td>38</td>
<td>49</td>
<td>5</td>
<td>3</td>
<td>2</td>
<td>97</td>
<td>51%</td>
<td></td>
</tr>
<tr>
<td>CCC</td>
<td>7</td>
<td>7</td>
<td>29</td>
<td>1</td>
<td>3</td>
<td>47</td>
<td>62%</td>
<td></td>
</tr>
<tr>
<td>MC</td>
<td>4</td>
<td>8</td>
<td>1</td>
<td>8</td>
<td>1</td>
<td>22</td>
<td>36%</td>
<td></td>
</tr>
<tr>
<td>LGSC</td>
<td>5</td>
<td>8</td>
<td>1</td>
<td>0</td>
<td>8</td>
<td>22</td>
<td>36%</td>
<td></td>
</tr>
<tr>
<td>Total</td>
<td>342</td>
<td>89</td>
<td>43</td>
<td>21</td>
<td>29</td>
<td>524</td>
<td>73%</td>
<td></td>
</tr>
</tbody>
</table>

N’s are shown in each cell, except where % is indicated. Bold indicates cases with agreement; HGSC-high-grade serous carcinoma, EC-endometrioid carcinoma, CCC-clear cell carcinoma, MC-mucinous carcinoma, LGSC-low-grade serous carcinoma
Table 3 Pairwise agreement of histological types in the testing set between original type and agreement between 2 out of 3 methods of agreement (original type, TB_COSPv2 and WT1-assisted core review)

<table>
<thead>
<tr>
<th></th>
<th>Agreement between 2 out of 3 methods of assessment</th>
<th>Concordance rate</th>
<th>K (95% CI)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>HGSC</td>
<td>EC</td>
<td>CCC</td>
</tr>
<tr>
<td>Original type</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>HGSC</td>
<td>300</td>
<td>6</td>
<td>5</td>
</tr>
<tr>
<td>EC</td>
<td>32</td>
<td>54</td>
<td>3</td>
</tr>
<tr>
<td>CCC</td>
<td>5</td>
<td>3</td>
<td>37</td>
</tr>
<tr>
<td>MC</td>
<td>2</td>
<td>4</td>
<td>0</td>
</tr>
<tr>
<td>LGSC</td>
<td>3</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>Total</td>
<td>342</td>
<td>67</td>
<td>45</td>
</tr>
</tbody>
</table>

N’s are shown in each cell, except where % is indicated. Bold indicates cases with agreement; HGSC-high-grade serous carcinoma, EC-endometrioid carcinoma, CCC-clear cell carcinoma, MC-mucinous carcinoma, LGSC-low-grade serous carcinoma.
Table 4 Multivariate Cox model in the testing set using different methods of type assessment.

<table>
<thead>
<tr>
<th></th>
<th>Original type HR (95% CI), p-value</th>
<th>TB_COSPv2 HR (95% CI), p-value</th>
<th>Agreement between original type and TB_COSPv2 HR (95% CI), p-value</th>
<th>Agreement between 2 out of 3 methods of original type, TB_COSPv2 and WT1-assisted core review HR (95% CI), p-value</th>
</tr>
</thead>
<tbody>
<tr>
<td>FIGO stage</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Stage III/IV versus I/II</td>
<td>4.39 (2.81-7.12), p&lt;0.0001</td>
<td>3.83 (2.50-5.99), p&lt;0.0001</td>
<td>4.47 (2.47-8.67), p&lt;0.0001</td>
<td>4.82 (3.00-8.08), p&lt;0.0001</td>
</tr>
<tr>
<td>Site</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>MAY versus UKO</td>
<td>1.02 (0.70-1.53), p=0.91</td>
<td>1.06 (0.75-1.60), p=0.76</td>
<td>1.24 (0.76-2.14), p=0.40</td>
<td>1.07 (0.73-1.64), p=0.71</td>
</tr>
<tr>
<td>Age</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Older than median versus median or younger</td>
<td>1.45 (1.14-1.87), p=0.0026</td>
<td>1.42 (1.11-1.82), p=0.0053</td>
<td>1.42 (1.06-1.81), p=0.016</td>
<td>1.49 (1.16-1.93), p=0.0019</td>
</tr>
<tr>
<td>Histological type*</td>
<td>HGSC (Reference) p-value overall</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>1.00, p=0.67</td>
<td>1.00, p=0.054</td>
<td>1.00, p=0.026</td>
<td>1.00, p=0.032</td>
</tr>
<tr>
<td>EC</td>
<td>0.85 (0.57-1.23), p=0.41</td>
<td>0.60 (0.37-0.93), p=0.021</td>
<td>0.36 (0.14-0.80), p=0.0098</td>
<td>0.56 (0.29-0.99), p=0.046</td>
</tr>
<tr>
<td>CCC</td>
<td>1.25 (0.72-2.07), p=0.41</td>
<td>1.06 (0.62-1.70), p=0.82</td>
<td>1.00 (0.46-1.92), p=1.00</td>
<td>1.31 (0.78-2.06), p=0.29</td>
</tr>
<tr>
<td>MC</td>
<td>1.46 (0.42-3.83), p=0.50</td>
<td>0.95 (0.45-1.75), p=0.87</td>
<td>1.30 (0.20-5.60), p=0.73</td>
<td>1.51 (0.52-3.47), p=0.40</td>
</tr>
<tr>
<td>LGSC</td>
<td>0.85 (0.40-1.58), p=0.64</td>
<td>0.57 (0.31-0.97), p=0.037</td>
<td>0.31 (0.05-0.98), p=0.045</td>
<td>0.50 (0.21-0.99), p=0.046</td>
</tr>
</tbody>
</table>

OTTA Testing Set (N = 524)  
Original Diagnosis

- HGSC 64% (N = 336)
- EC 19% (N = 97)
- CCC 9% (N = 47)
- MC 4% (N = 22)
- LGSC 4% (N = 22)

OTTA Testing Set (N = 524)  
TB_COSPv2 Predictions

- HGSC 65% (N = 342)
- EC 17% (N = 89)
- CCC 8% (N = 43)
- MC 4% (N = 21)
- LGSC 6% (N = 29)

Agreement between Original & TB_COSPv2

Yes

73% (382/524)

No

27% (142/524)

WT1 Assisted Core Review

Agreement with TB_COSPv2 15% (76/524)

Agreement with Original Type

Disagreement with Original & TB_COSPv2

TB_COSPv2 Error 6% (34/524)

"Unclassifiable" 6% (32/524)

Agreement between 2 out of 3 methods of assessment 94% (492/524)
Figure 2

A. Original type TB_COSPv2

B. TB_COSPv2

C. Agreement

D. Majority

Survival Plot

Time to event: Time after diagnosis (months)

Surviving

Time after diagnosis (months)

p=0.0057

p=0.0004

p=<0.0001

p=0.0003

Original type

TB_COSPv2

Agreement

Majority

Downloaded from cebp.aacrjournals.org on October 13, 2017. © 2013 American Association for Cancer Research.
Biomarker-based ovarian carcinoma typing: a histological investigation in the Ovarian Tumor Tissue Analysis consortium

Martin Koebel, Steve Kalloger, Sandra Lee, et al.

Cancer Epidemiol Biomarkers Prev  Published OnlineFirst July 23, 2013.

Updated version
Access the most recent version of this article at:
doi:10.1158/1055-9965.EPI-13-0391

Supplementary Material
Access the most recent supplemental material at:
http://cebp.aacrjournals.org/content/suppl/2013/07/25/1055-9965.EPI-13-0391.DC1

Author Manuscript
Author manuscripts have been peer reviewed and accepted for publication but have not yet been edited.

E-mail alerts
Sign up to receive free email-alerts related to this article or journal.

Reprints and Subscriptions
To order reprints of this article or to subscribe to the journal, contact the AACR Publications Department at pubs@aacr.org.

Permissions
To request permission to re-use all or part of this article, contact the AACR Publications Department at permissions@aacr.org.