Multisite Validation Study to Determine Performance Characteristics of a 92-Gene Molecular Cancer Classifier

Sarah E. Kerr1, Catherine A. Schnabel2, Peggy S. Sullivan3, Yi Zhang2, Veena Singh2, Brittany Carey4, Mark G. Erlander2, W. Edward Highsmith1, Sarah M. Dry3, and Elena F. Brachtel4

Abstract

Purpose: Accurate tumor classification is essential for cancer management as patient outcomes improve with use of site- and subtype-specific therapies. Current clinicopathologic evaluation is varied in approach, yet standardized diagnoses are critical for determining therapy. While gene expression–based cancer classifiers may potentially meet this need, imperative to determining their application to patient care is validation in rigorously designed studies. Here, we examined the performance of a 92-gene molecular classifier in a large multi-institution cohort.

Experimental Design: Case selection incorporated specimens from more than 50 subtypes, including a range of tumor grades, metastatic and primary tumors, and limited tissue samples. Formalin-fixed, paraffin-embedded tumors passed pathologist-adjudicated review between three institutions. Tumor classification using a 92-gene quantitative reverse transcriptase polymerase chain reaction (RT-PCR) assay was conducted on blinded tumor sections from 790 cases and compared with adjudicated diagnoses.

Results: The 92-gene assay showed overall sensitivities of 87% for tumor type [95% confidence interval (CI), 84–89] and 82% for subtype (95% CI, 79–85). Analyses of metastatic tumors, high-grade tumors, or cases with limited tissue showed no decrease in comparative performance (P = 0.16, 0.58, and 0.16). High specificity (96%–100%) was showed for ruling in a primary tumor in organs commonly harboring metastases. The assay incorrectly excluded the adjudicated diagnosis in 5% of cases.

Conclusions: The 92-gene assay showed strong performance for accurate molecular classification of a diverse set of tumor histologies. Results support potential use of the assay as a standardized molecular adjunct to routine clinicopathologic evaluation for tumor classification and primary site diagnosis. Clin Cancer Res; 1–9. ©2012 AACR.

Introduction

Tumors of uncertain origin, where a presumptive diagnosis is made but is not definitive, are a common and important clinical problem that poses both diagnostic and therapeutic challenges. A majority of the approximately 600,000 annual cancer deaths in the United States are attributed to metastatic spread from a primary tumor site (1, 2), the identity of which remains equivocal in a significant number of patients (3, 4). As patient outcomes continue to improve with the use of site-specific and molecular-targeted therapies (5), the requirement for diagnostic certainty about tumor type and site of origin in an evidenced-based approach is increasingly vital.

A common challenge is to distinguish recurrent, metastatic disease from a new primary in patients with a known history of cancer. Predictive tests for response to treatment have revolutionized patient management, yet biomarker activity is specifically linked to cellular context or tissue site (6). For instance, the targeted RAF inhibitor, vemurafenib, is effective against melanomas with activated BRAF (BRAF-V600E), but not against colorectal cancers harboring the same mutation (7, 8). In other difficult to diagnose scenarios, clinicians face a range of possible differential diagnoses, all of which may have distinct optimal therapeutic approaches. At the other end of the diagnostic spectrum are Cancers of Unknown Primary (CUP). These account for 3% to 5% of all malignancies and are rendered when a primary tumor site cannot be identified after detailed (3, 9–13), and often expensive (12, 14–16) diagnostic
Translational Relevance

Standardized methods for accurate tumor classification are of critical importance for diagnosis and patient treatment, particularly in diagnostically challenging cases where site-directed therapies are an option. Molecular profiling assays for tumor classification have been proposed as complementary approaches to clinicopathologic evaluations. In this study, we characterize the performance of a 92-gene molecular cancer classifier by comparing assay results to adjudicated reference diagnoses in a large multi-institutional cohort. This research has direct application to evidence-based cancer treatment, which has been revolutionized by patient stratification with predictive biomarkers and application of targeted therapies—both of which rely fundamentally on precise determination of the tumor type and primary site. Molecular cancer classifiers have potential clinical utility to increase diagnostic specificity and optimize therapeutic strategies as standardized analytic correlates.

evaluation. Despite advances in tumor imaging and pathology, the primary tumor site remains unknown or uncertain in approximately 30% of these cases, leaving a substantial fraction of patients with potentially suboptimal treatment and poorer prognosis (17, 18). In such cases, improved survival has been shown when the primary source of cancer is identified and site-specific therapies are instituted (19, 20). Therefore, precise determination of primary tumor type remains a key diagnostic element to optimal treatment selection across many clinical contexts.

Gene expression signatures for tumor classification have recently been used as analytic complements to standard clinicopathologic evaluation (21–28). The lack of large-scale validation studies to comprehensively define performance characteristics and clinical use present an ongoing challenge to clinical adoption of these tests. Thus, the objective of the current study was to conduct a comprehensive validation study to determine the accuracy of a 92-gene cancer classification assay (CancerTYPE ID, bioTheranostics Inc.) for tumor classification. Although in clinical practice the 92-gene assay would be used to aid in the diagnosis of tumors of uncertain or unknown primary origin, characterization of diagnostic accuracy in known tumors with established reference diagnoses using a blinded series of representative cases is a critical step to showing clinical validity. In this study, validation was conducted on a large and comprehensive range of tumor types and subtypes in an adjudicated, multi-institutional cohort.

Materials and Methods

Study approval was obtained from the Institutional Review Board at each study site [Mayo Clinic (Mayo, Rochester, MN), Massachusetts General Hospital (MGH, Boston, MA), University of California Los Angeles (UCLA, Los Angeles, CA)].

Tumor specimens and case adjudication

Case selection targets were approximately 50% metastatic tumors of any grade with the remainder composed of moderately to poorly differentiated primary tumors, and approximately 10% limited tissue specimens from cytologic preparations [fine needle aspiration (FNA) cell block] or small biopsies (i.e., needle core biopsies). Inclusion criteria were: (i) formalin-fixed, paraffin-embedded (FFPE) tumors processed less than 6 years from time of testing, (ii) diagnosis contained within the assay panel, (iii) at least 40% tumor available in a markable area on the hematoxylin and eosin (H&E) slide, and (iv) minimal necrosis. Decalcified cases and cytology cases other than FNA cell blocks were excluded.

Table 1 shows patient and specimen characteristics. Histologic grading for primary tumors was based on standard criteria for each organ system. Study site pathologists reviewed slides, pathology reports, and clinical history and selected an H&E slide for adjudication by a second pathologist at a different institution. Slides and pathology reports were digitally scanned and uploaded (Spectrum & ImageScope, Aperio Technologies, Inc.). Figure 1 shows the case selection process. In total, 790 cases were included in the final analysis.

Table 1. Patient and tumor characteristics cases (N = 790) examined in the study

<table>
<thead>
<tr>
<th>Characteristic</th>
<th>N (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Gender</strong></td>
<td></td>
</tr>
<tr>
<td>Male</td>
<td>385 (49)</td>
</tr>
<tr>
<td>Female</td>
<td>405 (51)</td>
</tr>
<tr>
<td><strong>Age, y (±SD)</strong></td>
<td></td>
</tr>
<tr>
<td>&lt;50</td>
<td>203 (26)</td>
</tr>
<tr>
<td>50–64</td>
<td>271 (34)</td>
</tr>
<tr>
<td>≥65</td>
<td>316 (40)</td>
</tr>
<tr>
<td><strong>Tumor</strong></td>
<td></td>
</tr>
<tr>
<td>Primary</td>
<td>441 (56)</td>
</tr>
<tr>
<td>Grade I</td>
<td>35 (6)</td>
</tr>
<tr>
<td>Grade II</td>
<td>87 (20)</td>
</tr>
<tr>
<td>Grade III</td>
<td>189 (43)</td>
</tr>
<tr>
<td>Not gradeda</td>
<td>130 (29)</td>
</tr>
<tr>
<td>Metastatic</td>
<td>349 (44)</td>
</tr>
<tr>
<td><strong>Specimen types</strong></td>
<td></td>
</tr>
<tr>
<td>Limited tissue biopsy</td>
<td>109 (14)</td>
</tr>
<tr>
<td>Excision/resection</td>
<td>681 (86)</td>
</tr>
<tr>
<td><strong>Samples by site</strong></td>
<td></td>
</tr>
<tr>
<td>Mayo</td>
<td>401 (51)</td>
</tr>
<tr>
<td>UCLA</td>
<td>191 (24)</td>
</tr>
<tr>
<td>MGH</td>
<td>198 (25)</td>
</tr>
</tbody>
</table>

Abbreviation: N, number of samples.

aGrading is not traditionally conducted in certain tumors, such as pheochromocytomas and gastrointestinal stromal tumors. Metastatic tumors of all grades were enrolled.
Specimen processing and assay protocol

Study sites prepared 1 H&E and 3 unstained slides labeled with Study ID only. Laboratory personnel were blinded to all information except biopsy site and patient gender. Tumor cells were enriched by either macrodissection or laser microdissection (LMD 6000, Leica Microsystems). The 92-gene assay (real-time RT-PCR) was conducted on isolated total RNA as previously described (22) and used a prespecified computational algorithm to generate probabilities for candidate tumor types based on the degree of similarity of the queried sample to the reference tumor database. Cases exceeding the PCR analytic cutoff value for internal controls (cycling threshold >30) were considered quality control failures (Fig. 1).

Tumor classification in the 92-gene assay is structured as a 2-level labeling scheme: a tier 1 class or main type (lung) and a tier 2 class or subtype (lung adenocarcinoma). Highest probabilities for tier 1 and tier 2 classes were compared with the adjudicated diagnosis. In addition, main types with >5% probability (i.e., rank order predictions with significant similarity) and tumor types with a combined probability <5% (rule out types) were calculated. A predetermined threshold of less than 85% for the top ranked probability was used in the study for test results to be considered unclassifiable.

Study design and statistical analyses

In this prospectively defined, blinded, multi-institutional study, power calculations were based on an estimated overall sensitivity of 85%. A sample size of 620, or at least 22 samples per main tumor type, was determined to be the minimum number of samples needed to detect a 5% reduction in overall performance with 95% power at the level of 0.05 significance. Minimum sample requirements were 25 per main tumor type and 10 per tumor subtype. The primary objective of the study was to determine diagnostic accuracy of the 92-gene assay.

Overall sensitivity (i.e., overall diagnostic accuracy) was calculated as the number of cases with an assay diagnosis that was concordant with the adjudicated reference diagnosis, divided by the total number of cases classifiable by the assay. Sensitivity, specificity, positive predictive value (PPV) and negative predictive value (NPV) for each tumor class were calculated as previously described (29). Diagnostic OR as a measure of test effectiveness was calculated as previously described (30). Diagnostic accuracy between clinical subsets was compared using the Fisher exact test; the Cochran–Mantel–Haenszel test was used to compare accuracy across study sites adjusting for the proportion of metastatic cases. Analysis of covariance (ANCOVA) was used to assess performance differences due to tissue composition, adjusting for study sites and tumor enrichment methods. Logistic regression with biopsy sites and tumor content as covariates was used to compare accuracy between tumor enrichment methods.

To evaluate classifier performance in discriminating primary versus metastatic tumors, cases from a designated primary site were grouped to include multiple histologies; lung included all non–small cell, squamous, large cell, and neuroendocrine lung tumors; brain included both gliomas and meningiomas; pleura/peritoneum included...
mesothelioma, ovary included epithelial, and sex cord stromal tumors; and liver included hepatocellular carcinoma and intrahepatic cholangiocarcinoma. Performance was examined on the basis of biopsy site and calculated from a 2 × 2 table by considering "positives" as a correctly predicted primary tumor and "negatives" as metastatic tumors to the respective biopsy site (i.e., correct prediction of a lung biopsy as metastatic renal cell carcinoma would be scored as a true negative, whereas a correct prediction of lung adenocarcinoma would be scored as a true positive).

**Results**

### Overall performance of the 92-gene assay for tumor classification and subclassification

The 92-gene assay showed an overall sensitivity of 87% [95% confidence interval (CI), 84–89] for 28 main tumor types (Table 2). Specificities for main type classes ranged from 98% to more than 99%. PPVs ranged from 61% (intestine) to 100% (brain, endometrium, GIST, meningioma, mesothelioma, prostate, sex cord stromal, skin basal cell, and thymus). PPVs for the most prevalent cancers of breast, prostate and lung were 89%, 100%, and 94%, respectively, and strong precision (≥80%) was showed across the majority of the classifier. Figure 2 shows a matrix showing the relationship of test results compared with the reference diagnoses. Test sensitivity improved to 91% and 94%, respectively, if the assay's second and third predicted diagnoses were included. The reference diagnosis incorrectly was ruled out by the assay in 5% of cases. Forty-seven cases (5.9%) were unclassifiable by the assay (Fig. 1, Table 2). Lung, lymph node, brain, and liver were the most common biopsy sites in the study (Supplementary Fig. S1).

For tumor subtyping of 50 histologies, the overall sensitivity was 82% (95% CI, 79–85) with subtype specificities ranging from 98% to more than 99% (Supplementary Table S1). Diagnostic ORs for all the main types and subtypes were significantly more than 1, indicating each class and subclass reported by the 92-gene assay provides significant discrimination and performance (data not shown).

<table>
<thead>
<tr>
<th>Tumor type</th>
<th>Tested (N)</th>
<th>Unc (N)</th>
<th>Sensitivity (95% CI)</th>
<th>Specificity (95% CI)</th>
<th>PPV</th>
<th>NPV</th>
</tr>
</thead>
<tbody>
<tr>
<td>Adrenal</td>
<td>25</td>
<td>0</td>
<td>0.96 (0.80–1.00)</td>
<td>0.99 (0.98–1.00)</td>
<td>0.83</td>
<td>1.00</td>
</tr>
<tr>
<td>Brain</td>
<td>25</td>
<td>0</td>
<td>0.96 (0.80–1.00)</td>
<td>1.00 (0.99–1.00)</td>
<td>1.00</td>
<td>1.00</td>
</tr>
<tr>
<td>Breast</td>
<td>25</td>
<td>5</td>
<td>0.80 (0.56–0.94)</td>
<td>1.00 (0.99–1.00)</td>
<td>0.89</td>
<td>0.99</td>
</tr>
<tr>
<td>Cervix adenocarcinoma</td>
<td>25</td>
<td>7</td>
<td>0.72 (0.47–0.90)</td>
<td>0.99 (0.99–1.00)</td>
<td>0.76</td>
<td>0.99</td>
</tr>
<tr>
<td>Endometrium</td>
<td>25</td>
<td>4</td>
<td>0.48 (0.26–0.70)</td>
<td>1.00 (0.99–1.00)</td>
<td>1.00</td>
<td>0.99</td>
</tr>
<tr>
<td>Gastroesophageal</td>
<td>25</td>
<td>5</td>
<td>0.65 (0.41–0.85)</td>
<td>0.99 (0.99–1.00)</td>
<td>0.76</td>
<td>0.99</td>
</tr>
<tr>
<td>Germ cell</td>
<td>25</td>
<td>2</td>
<td>0.83 (0.61–0.95)</td>
<td>1.00 (0.99–1.00)</td>
<td>0.86</td>
<td>0.99</td>
</tr>
<tr>
<td>GIST</td>
<td>25</td>
<td>0</td>
<td>0.92 (0.74–0.99)</td>
<td>1.00 (0.99–1.00)</td>
<td>1.00</td>
<td>1.00</td>
</tr>
<tr>
<td>Head-neck-salivary</td>
<td>25</td>
<td>1</td>
<td>0.88 (0.68–0.97)</td>
<td>0.99 (0.98–1.00)</td>
<td>0.78</td>
<td>1.00</td>
</tr>
<tr>
<td>Intestine</td>
<td>25</td>
<td>5</td>
<td>0.85 (0.62–0.97)</td>
<td>0.98 (0.97–0.99)</td>
<td>0.61</td>
<td>1.00</td>
</tr>
<tr>
<td>Kidney</td>
<td>30</td>
<td>0</td>
<td>0.97 (0.83–1.00)</td>
<td>1.00 (0.99–1.00)</td>
<td>0.91</td>
<td>1.00</td>
</tr>
<tr>
<td>Liver</td>
<td>25</td>
<td>0</td>
<td>0.96 (0.80–1.00)</td>
<td>1.00 (0.99–1.00)</td>
<td>0.96</td>
<td>1.00</td>
</tr>
<tr>
<td>Lung-adeno/large cell</td>
<td>25</td>
<td>2</td>
<td>0.65 (0.43–0.84)</td>
<td>1.00 (0.99–1.00)</td>
<td>0.94</td>
<td>0.99</td>
</tr>
<tr>
<td>Lymphoma</td>
<td>25</td>
<td>0</td>
<td>0.84 (0.64–0.95)</td>
<td>1.00 (0.99–1.00)</td>
<td>0.95</td>
<td>0.99</td>
</tr>
<tr>
<td>Melanoma</td>
<td>25</td>
<td>0</td>
<td>0.88 (0.69–0.97)</td>
<td>1.00 (0.99–1.00)</td>
<td>0.96</td>
<td>1.00</td>
</tr>
<tr>
<td>Meningioma</td>
<td>25</td>
<td>0</td>
<td>1.00 (0.80–1.00)</td>
<td>1.00 (0.99–1.00)</td>
<td>1.00</td>
<td>1.00</td>
</tr>
<tr>
<td>Mesothelioma</td>
<td>25</td>
<td>2</td>
<td>0.87 (0.66–0.97)</td>
<td>1.00 (0.99–1.00)</td>
<td>1.00</td>
<td>1.00</td>
</tr>
<tr>
<td>Neuroendocrine</td>
<td>50</td>
<td>0</td>
<td>0.98 (0.89–1.00)</td>
<td>1.00 (0.99–1.00)</td>
<td>0.94</td>
<td>1.00</td>
</tr>
<tr>
<td>Ovary</td>
<td>40</td>
<td>4</td>
<td>0.86 (0.71–0.95)</td>
<td>0.98 (0.97–0.99)</td>
<td>0.69</td>
<td>0.99</td>
</tr>
<tr>
<td>Pancreaticobiliary</td>
<td>30</td>
<td>6</td>
<td>0.88 (0.68–0.97)</td>
<td>0.99 (0.98–1.00)</td>
<td>0.72</td>
<td>1.00</td>
</tr>
<tr>
<td>Prostate</td>
<td>25</td>
<td>0</td>
<td>1.00 (0.80–1.00)</td>
<td>1.00 (0.99–1.00)</td>
<td>1.00</td>
<td>1.00</td>
</tr>
<tr>
<td>Sarcoma</td>
<td>60</td>
<td>0</td>
<td>0.95 (0.86–0.99)</td>
<td>0.99 (0.97–0.99)</td>
<td>0.85</td>
<td>1.00</td>
</tr>
<tr>
<td>Sex cord stromal tumor</td>
<td>25</td>
<td>0</td>
<td>0.80 (0.59–0.93)</td>
<td>1.00 (0.99–1.00)</td>
<td>1.00</td>
<td>0.99</td>
</tr>
<tr>
<td>Skin basal cell</td>
<td>25</td>
<td>0</td>
<td>1.00 (0.80–1.00)</td>
<td>1.00 (0.99–1.00)</td>
<td>1.00</td>
<td>1.00</td>
</tr>
<tr>
<td>Squamous</td>
<td>30</td>
<td>1</td>
<td>0.86 (0.68–0.96)</td>
<td>0.98 (0.97–0.99)</td>
<td>0.66</td>
<td>0.99</td>
</tr>
<tr>
<td>Thymus</td>
<td>25</td>
<td>0</td>
<td>0.72 (0.51–0.88)</td>
<td>1.00 (0.99–1.00)</td>
<td>1.00</td>
<td>0.99</td>
</tr>
<tr>
<td>Thyroid</td>
<td>25</td>
<td>0</td>
<td>0.96 (0.80–1.00)</td>
<td>1.00 (0.99–1.00)</td>
<td>0.96</td>
<td>1.00</td>
</tr>
<tr>
<td>Urinary bladder</td>
<td>25</td>
<td>3</td>
<td>0.64 (0.41–0.83)</td>
<td>0.99 (0.98–1.00)</td>
<td>0.67</td>
<td>0.99</td>
</tr>
<tr>
<td>Overall</td>
<td>0.87 (0.84–0.89)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Abbreviations: N, number of cases; Unc, unclassified by the 92-gene assay.
Effects of histologic and clinical variables

By ANCOVA analysis, assay performance was not affected by any of the measured histologic variables or dissection method (Supplementary Table S2). Analysis of relevant clinical subsets including metastatic tumors, histologic grades, or cases with limited tissue showed no decrease in comparative performance (Table 3, $P = 0.16, 0.58, \text{and} 0.16$, respectively). In addition, performance across the study sites was not statistically different.

92-Gene assay prediction of primary versus metastatic tumors

A posthoc analysis was conducted to evaluate the ability of the 92-gene assay to discriminate primary from metastatic lesions in biopsies from common metastatic sites. In an analysis that included 205 metastatic and 147 primary cases, the 92-gene assay showed strong precision for accurate identification of a primary tumor, reported as PPVs of 100% for lung ($N = 99$), brain ($N = 84$), and pleura/peritoneum ($N = 73$), 92% for ovary ($N = 46$), and 80% for liver ($N = 65$; Table 4).

Discussion

In current practice, diagnostically challenging tumors are evaluated using a nonstandardized approach that integrates data from clinical history, radiology, and surgical findings, and morphologic and immunohistochemical analyses, which may be associated with considerable time and cost (12, 14–16). Immunohistochemistry is a cornerstone of traditional cancer classification, however, reported accuracies for primary site diagnosis in cases of metastatic tumors are 66% to 68% (10, 12). In cancers of uncertain or unknown origin, identification of a primary site remains equivocal in a significant numbers of cases, particularly when the clinical presentation is atypical and immunohistochemistry is noncontributory. Gene expression signatures offer a higher resolution and standardized approach. As a molecular complement to clinicopathologic evaluation, gene expression–based classifiers have the potential to significantly impact patient management through improving the accuracy and specificity of tumor classification.

![Confusion matrix by tumor type. Reference diagnoses are shown along the left-hand column, and 92-gene assay predictions are shown across the top row. The matrix shows the direct relationship between each adjudicated reference diagnosis versus the molecular classifier prediction, including reproducible patterns of classification and misclassification.](image-url)
Clinical variables % Sensitivity \( P \)
---
Disease type
- Metastatic 44 85% 0.157
- Primary 56 88%
Histologic grade
- 1 4 91% 0.577
- 2 10 89%
- 3 24 89%
- Not graded 62 85%
Specimen type
- Limited tissue biopsy 14 91% 0.161
- Excision 86 86%

NOTE: Metastatic, high-grade, and limited tissue samples did not show a statistically significant decrease in assay performance compared with primary, low-grade, and larger sample types, respectively.

Kerr et al. date and proposes several key advancements to previous efforts. Unlike prior validation studies, rigorous adjudication was used to establish diagnosis in all cases: previous reports of other gene expression–based classifiers did not include independent validation of the diagnoses used to assess accuracy (24, 31, 32). In tumor bank studies of archival specimens, corresponding pathology reports may contain inaccurate diagnoses, leading to over- or underestimation of test performance (33, 34). The increased rigor provided by peer adjudication allows a more precise characterization of performance, and therefore a more substantiated clinical indication for the 92-gene assay.

Molecular tests must carry out well on limited diagnostic material as trends continue to shift to minimally invasive biopsies. However, prior studies of microarray-based classifiers have not conducted a blinded validation of performance in limited biopsy specimens, although feasibility assessments have been conducted (35). In fact, studies of other microarray platforms included a significant proportion of cases that were evaluated from electronic files of gene expression data and excluded a tissue-processing component within the study (24, 36). Approximately 14% of the total cases tested in the current study were limited tissue biopsies and consisted of 48% core needle biopsies, 34% small biopsies, and 18% cell blocks. Accuracy within this subgroup was 91% and comparative analysis of excisional versus limited tissue biopsies showed no statistical difference in performance (\( P = 0.16 \)), supporting the feasibility of using small tissue biopsy specimens with the 92-gene assay.

Definitive analytic validation with sufficient statistical power to characterize performance is a requisite to enabling increased adoption of molecular cancer classifiers. Results from the current study represent an important extension to previous analyses of the 92-gene assay performance, and in addition, close the gap on several limitations. Leave-one-out cross-validation studies are a standard preliminary estimate of accuracy, however, as results by definition are generated with the training database, they may be subject to overfitting and potential overestimation of true performance. Further validation is often conducted on an independent test set of tumors. However, if class representation and sample requirements are not sufficient, and cases are not tested in a blinded fashion, true performance remains unsubstantiated. In the current study of 790 blinded cases with a minimum of 25 cases tested per tumor type, the 92-gene assay was conducted with 87% accuracy for identification of 28 main tumor types. These findings were consistent with previous studies of the 92-gene assay showing accuracies of 83% in an independent test set of 187 tumors and 87% in leave one out cross-validation. In addition, results compare favorably with those reported for other gene expression–based classifiers, which ranged from 75% to 89%, and with current standard-of-care immunohistochemical classification, which ranged from 66% to 88% (10, 12). Notably, the 92-gene assay showed no statistical difference in overall accuracy within relevant clinical subsets of metastatic versus primary tumors (\( P = 0.16 \)) or across tumor grades (\( P = 0.6 \)).

The analytic strength of the 92-gene assay established in this study is attributed to several specific test characteristics. Because tumor classification is carried out by quantifying the molecular similarity of the gene expression profile of a sample tissue to a reference database of known tumors, the quality and scope of the reference database is integral for accurate tumor classification. The reference database of the 92-gene assay is composed of more than 2,000 independently validated tumor specimens which cover more than

Table 4. Clinical use of the 92-gene assay in discrimination of primary versus metastatic tumors

<table>
<thead>
<tr>
<th>Biopsy site</th>
<th>Lesion type</th>
<th>Performance for prediction of primary Tumor</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Primary ( n )</td>
<td>Metastatic ( n )</td>
</tr>
<tr>
<td>Lung</td>
<td>34</td>
<td>62</td>
</tr>
<tr>
<td>Liver</td>
<td>8</td>
<td>54</td>
</tr>
<tr>
<td>Brain</td>
<td>50</td>
<td>33</td>
</tr>
<tr>
<td>Ovary</td>
<td>36</td>
<td>7</td>
</tr>
<tr>
<td>Pleura/peritoneum</td>
<td>19</td>
<td>49</td>
</tr>
</tbody>
</table>
95% of solid tumors by incidence, with 26 to 228 specimens per tumor class that encompass a range of intratumor heterogeneity. In addition, gene expression is quantified using PCR methods, which allows broad clinical application, as routinely processed FFPE specimens can be used. In a recent analysis of 754 cases where the 92-gene assay was used as part of clinical care, reportable results were generated from 93% of the specimens; this provides further evidence that this platform is stable and highly compatible with standard clinical practice, including the inherent variations in tissue processing that routinely occurs between pathology laboratories (37). Finally, the data-driven approach used to develop the 92-gene panel directed the selection of genes with inherent discriminatory ability to classify a diverse spectrum of tumor types (22, 29). These key assay features contribute to a robust platform with high clinical compatibility.

Examination of assay misclassifications revealed the biologic and morphologic underpinnings of the 92-gene biomarker panel in lineage determination. The lowest performing tumor type was endometrial cancer, in which approximately half were predicted as ovarian (Fig. 2). Given current controversies over the ontogeny of female genital tract cancers (38–40), molecular profiling with the 92-gene assay may reflect this biologic intersection and provide additional insight into the origin of these tumors. Similarly, 4 breast cancers resulted in salivary gland predictions, and 1 salivary gland adenocarcinoma was predicted as breast cancer. Biologic proximity between these 2 tumor types is supported by HER2 amplification (41–43). Similar phylogenetic relationships are inherent to other types of tests using measures of similarity (44, 45). Precision may improve as reference libraries become more complete. The 92-gene assay reference database covers 95% of solid tumors based on incidence and is the most comprehensive to date.

The 92-gene assay conducted extremely well in distinguishing primary from metastatic tumors in common metastatic sites, an increasingly frequent scenario in cancer survivors with a new lesion on follow-up imaging. The 92-gene assay showed 100% PPV for identifying a primary tumor in lung, brain, and pleura/peritoneum. For distinguishing primary from metastatic carcinoma, no single immunohistochemical marker offers comparable performance. While immunohistochemical markers may improve performance, individual immunohistochemical markers are limited by lack of specificity, variable expression in poorly differentiated tumors, many technical sources of staining variance, and a subjective analytic approach and interpretation (46–52). Thus, the 92-gene assay may provide an important diagnostic adjunct, particularly in tumors with equivocal histopathology and immunohistochemistry, lack of specific clinical findings, and/or a range of differential diagnoses. Further support of the clinical use for the 92-gene assay is shown by a low false rule-out rate, wherein the 92-gene assay incorrectly excluded the adjudicated diagnosis in only 5% of the cases; these findings suggest that the 92-gene assay can be used to reduce the number of differential diagnoses with added certainty.

Indeed, recent clinical experience with the 92-gene assay in 2 analyses that included more than 1,000 patients shows the diverse clinical application of this assay in the management of patients with cancer (29, 37). Data abstracted from accompanying pathology reports in cases submitted for clinical testing show that the majority are not CUP but are presumed metastatic tumors that require either (i) molecular confirmation where one tumor type is favored (18%–20%) or (ii) additional molecular data to reduce the possibilities in a range of potential differential diagnoses (44%–52%; refs. 29, 37). Additional findings from cases submitted for 92-gene assay testing show remarkable variability in the number of immunohistochemical markers (0–35) used in routine pathologic work-up, as well as a wide range in the time to diagnosis (37). Collectively, these data highlight the clinical need for methodologies that permit more standardized diagnosis, and support a positive cost-benefit argument for the 92-gene assay toward improving health effectiveness and efficiency. Appropriate use of the 92-gene assay has the potential to prevent unnecessary additional testing, improve therapy, avoid toxicity and safety issues, decrease time to diagnosis, prevent excess tissue usage which may be preserved for downstream biomarker testing, and increase enrollment in clinical trials. Health economics studies to directly address these questions are currently ongoing.

A limitation of the study was exclusion of decalcified specimens and other challenging cytologic specimens including cell blocks from malignant effusions and slide scrapings of smeared aspirate material. Future studies of the 92-gene assay on these specimens would be helpful. Another limitation was that investigation of off-panel performance for tumors not covered by 92-gene assay was not conducted. Clinical samples submitted for testing may originate from cancers not represented in the assay reference database. Although beyond the scope of the current study, data on potential misclassification rates are important. Finally, a limit of this and of all molecular classifier studies lies in the diagnostic "gold standard" used. As noted, current knowledge of pathobiology for certain tumor types remains uncertain and subject to intense study and debate. Given this, despite rigorous diagnostic adjudication, it is possible that current practices are erroneous, which could affect assay performance as well as the integrity of the reference database. This limitation remains an ongoing challenge that awaits further resolution through continuing research efforts.

Concluding Summary

Results of this validation study support the clinical utility of the 92-gene assay in tumors of uncertain origin as a molecular adjunct to clinicopathologic evaluation for primary site diagnosis, discrimination between primary and metastatic tumor in common metastatic sites and for tumor subclassification. Prospective studies will help further define how
molecular data can be successfully integrated into the clinical decision making process and allow for increased diagnostic certainty.

Disclosure of Potential Conflicts of Interest
C.A. Schnabel, V. Zhang, V. Singh, M. Erlander are employees at bioTheranostics Inc and have ownership interest (including patents) for bioTheranostics Inc. P.S. Sullivan, W.E. Highsmith, S.M. Dry, and E. Brachtel have commercial research grants for bioTheranostics Inc. No potential conflicts of interest were disclosed by the other authors.

Authors’ Contributions
Conception and design: S.E. Kerr, C.A. Schnabel, W.E. Highsmith, S.M. Dry, E. Brachtel
Acquisition of data (provided animals, acquired and managed patients, provided facilities, etc.): S.E. Kerr, C.A. Schnabel, P.S. Sullivan, V. Singh, B. Carey, S.M. Dry, E. Brachtel
Analysis and interpretation of data (e.g., statistical analysis, bioinformatics, computational analysis): S.E. Kerr, C.A. Schnabel, Y. Zhang, V. Singh, B. Carey, S.M. Dry, E. Brachtel

References
Administrative, technical, or material support (i.e., reporting or organizing data, constructing databases): S.E. Kerr, P.S. Sullivan, B. Carey

Acknowledgments
The authors thank Mary Till and David Dvorak (Mayo Clinic) for coordinating case selection and data management; the Translational Pathology Core Laboratory (UCLA) for sample preparation and digital imaging; Tricia Della Pelle and Stephen Conley (MGH) for slide preparation and scanning and bioTheranostics Ranelle Salunga, Mariko Matsutani, Lavenia Correa, Yvette De la Torre, and Susan Hicks for expert technical assistance; Jose Galindo for data management; Jeff Anderson for manuscript support; and Brock Schroeder for critical review.

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 March 21, 2012; revised May 7, 2012; accepted May 22, 2012; published OnlineFirst May 30, 2012.
Multisite Validation of a 92-Gene Molecular Cancer Classifier


38. Seidman JD, Zhao P, Yemelyanova A. “Primary peritoneal” high-grade serous carcinoma is very likely metastatic from serous tubal intraepithelial carcinoma: assessing the new paradigm of ovarian and pelvic serous carcinoma genesis and its implications for screening for ovarian cancer. Gynecol Oncol 2011;120:470–3.


Multisite Validation Study to Determine Performance Characteristics of a 92-Gene Molecular Cancer Classifier

Sarah E. Kerr, Catherine A. Schnabel, Peggy S. Sullivan, et al.

Clin Cancer Res  Published OnlineFirst May 30, 2012.

Updated version  Access the most recent version of this article at: doi:10.1158/1078-0432.CCR-12-0920

Supplementary Material  Access the most recent supplemental material at: http://clincancerres.aacrjournals.org/content/suppl/2012/05/29/1078-0432.CCR-12-0920.DC1

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.