

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 ($P = 0.16, 0.58, \text{ and } 0.16$). 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*; 18(14); 3952–60. ©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 evidence-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

Authors' Affiliations: ¹Department of Laboratory Medicine and Pathology, Mayo Clinic, Rochester, Minnesota; ²bioTheragnostics, Inc., San Diego, California; ³Department of Pathology and Laboratory Medicine, David Geffen School of Medicine, University of California Los Angeles, Los Angeles, California; and ⁴Department of Pathology, Massachusetts General Hospital, Boston, Massachusetts

Note: Supplementary data for this article are available at Clinical Cancer Research Online (<http://clincancerres.aacrjournals.org/>).

S.E. Kerr and C.A. Schnabel co-primary authors.

S.M. Dry and E.F. Brachtel co-senior authors.

Previous presentations: United States & Canadian Academy of Pathology, 2012 Annual Meeting.

Corresponding Author: Sarah M. Dry, Department of Pathology and Laboratory Medicine, David Geffen School of Medicine at UCLA, 13-145 CHS, Los Angeles, CA 90095. Phone: 310-825-8182; Fax: 310-267-2058; E-mail: sdry@mednet.ucla.edu

doi: 10.1158/1078-0432.CCR-12-0920

©2012 American Association for Cancer Research.

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

Characteristic	<i>N</i> (%)
<i>Gender</i>	
Male	385 (49)
Female	405 (51)
<i>Age, y (±SD)</i>	59 ± 16
<50	203 (26)
50–64	271 (34)
≥65	316 (40)
<i>Tumor</i>	
Primary	441 (56)
Grade I	35 (8)
Grade II	87 (20)
Grade III	189 (43)
Not graded ^a	130 (29)
Metastatic	349 (44)
<i>Specimen types</i>	
Limited tissue biopsy	109 (14)
Excision/resection	681 (86)
<i>Samples by site</i>	
Mayo	401 (51)
UCLA	191 (24)
MGH	198 (25)

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.

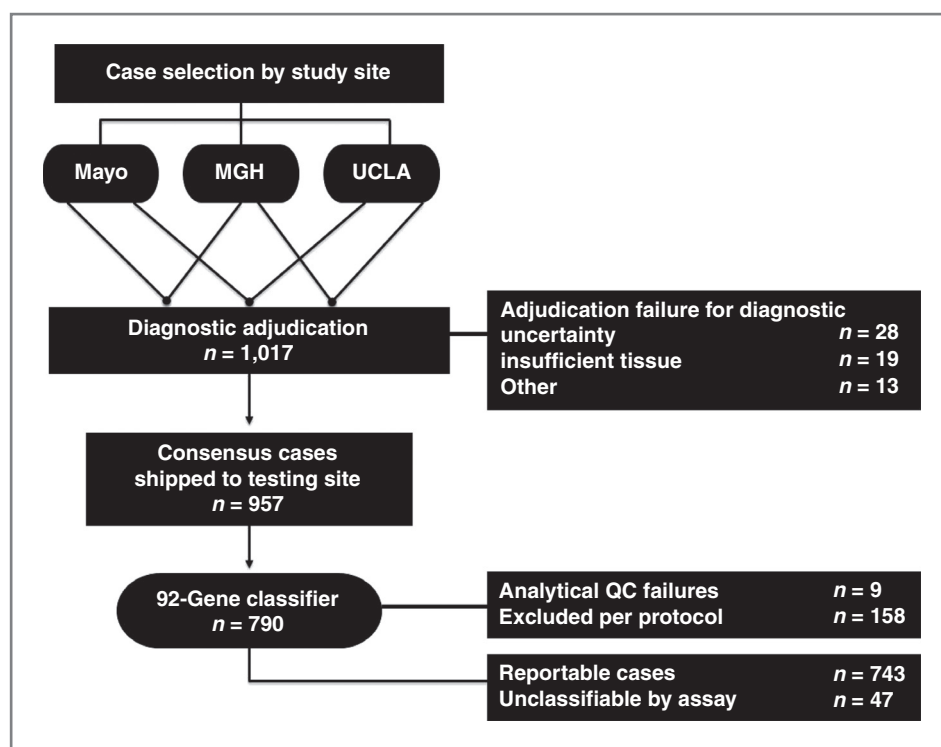


Figure 1. Case selection and flow diagram of the validation cohort. Cases were selected by a rolling enrollment process. Originating study sites identified cases that were submitted for adjudication at a second study site via whole slide imaging. Consensus cases were shipped for testing. A total of 790 cases comprised the blinded validation cohort; 743 (94.1%) with reported predictions and 47 (5.9%) unclassifiable cases.

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 $\geq 5\%$ 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

Table 2. 92-Gene assay performance characteristics for tumor classification based on concordance with the adjudicated diagnosis

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

Abbreviations: N, number of cases; Unc, unclassified by the 92-gene assay.

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 ($\geq 80\%$) 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).

	Adrenal	Brain	Breast	Cervix adenocarcinoma	Endometrium	Gastroesophageal	Germ cell	GIST	Head-neck-salivary	Intestine	Kidney	Liver	Lung-adeno/large cell	Lymphoma	Melanoma	Meningioma	Mesothelioma	Neuroendocrine	Ovary	Pancreaticobiliary	Prostate	Sarcoma	Sexcordstromaltumor	Skin basal cell	Squamous	Thymus	Thyroid	Urinary bladder	Unclassified	Total
Adrenal	24																					1								25
Brain		24																				1								25
Breast			16						4																				5	25
Cervix adenocarcinoma				13	1														2	2									7	25
Endometrium				1	10														10										4	25
Gastroesophageal						13	1			2									1	3									5	25
Germ cell							19																				4		2	25
GIST							1	23														1								25
Head-neck-salivary			1						21									1							1				1	25
Intestine					2					17									1										5	25
Kidney											29				1															30
Liver												24								1										25
Lung-adeno/large cell				1				2	1				15					1							1		2	2	25	
Lymphoma													1	21									3							25
Melanoma											1				22								2							25
Meningioma																25														25
Mesothelioma																	20			1		2							2	25
Neuroendocrine																		49												50
Ovary			1	1					2	1										31									4	40
Pancreaticobiliary						1				1											21							1	6	30
Prostate																						25								25
Sarcoma	1										1												57							60
Sexcordstromaltumor		4												1										20						25
Skin basal cell																									25					25
Squamous				1						2									1							25				30
Thymus											1															5	18	1		25
Thyroid																									1		24		25	
Urinary bladder										3																		14	3	25
Total	29	24	18	17	10	17	22	23	27	28	32	25	16	22	23	25	20	52	45	29	25	67	20	25	38	18	25	21	47	790

Figure 2. 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.

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, 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.

The study presented herein represents the most diagnostically comprehensive validation of a molecular classifier to

Table 3. 92-Gene assay performance in clinical subsets with reported predictions

Clinical variables	%	Sensitivity	<i>P</i>
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.

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

Biopsy site	Lesion type Primary (n)	Performance for prediction of primary Tumor				
		Metastatic (n)	Sensitivity (95% CI)	Specificity (95% CI)	PPV (%)	NPV (%)
Lung	34	62	79 (0.62–0.91)	100 (0.91–1.00)	100	90
Liver	8	54	100 (0.52–1.00)	96 (0.87–1.00)	80	100
Brain	50	33	98 (0.89–1.00)	100 (0.85–1.00)	100	97
Ovary	36	7	92 (0.78–0.98)	57 (0.18–0.90)	92	57
Pleura/peritoneum	19	49	95 (0.74–1.00)	100 (0.89–1.00)	100	98

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 panels 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, Y. 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

Development of methodology: S.E. Kerr, C.A. Schnabel, P.S. Sullivan, B. Carey, M. Erlander, 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, biostatistics, computational analysis): S.E. Kerr, C.A. Schnabel, Y. Zhang, V. Singh, B. Carey, S.M. Dry, E. Brachtel

Writing, review, and/or revision of the manuscript: S.E. Kerr, C.A. Schnabel, P.S. Sullivan, Y. Zhang, V. Singh, W.E. Highsmith, S.M. Dry, E. Brachtel

Administrative, technical, or material support (i.e., reporting or organizing data, constructing databases): S.E. Kerr, P.S. Sullivan, B. Carey
Study supervision: C.A. Schnabel, P.S. Sullivan, V. Singh, W.E. Highsmith, S.M. Dry, E. Brachtel

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.

References

- American Cancer Society. Cancer facts & figures 2010. Atlanta, GA: American Cancer Society; 2010 [cited 2010 September 29]. Available from: <http://www.cancer.org/acs/groups/content/epidemiologysurveillance/documents/document/acspc-026238.pdf>.
- Altekruse SF, Kosary CL, Krapcho M, Neyman N, Aminou R, Waldron W, et al., editors. SEER Cancer Statistics Review, 1975–2007. Bethesda, MD: National Cancer Institute; 2010. Available from: http://seer.cancer.gov/csr/1975_2007/, based on November 2009 SEER data submission, posted to the SEER web site.
- Greco FA, Hainsworth JD. Cancer of unknown primary site. In: DeVita VTJ, Hellman S, Rosenberg SA, editors. Cancer: principles and practice of oncology. 9th ed. Philadelphia, PA: Lippincott Williams & Wilkins; 2011. p. 2033–51.
- Hillen HF. Unknown primary tumours. *Postgrad Med J* 2000;76:690–3.
- Martini M, Vecchione L, Siena S, Tejpar S, Bardelli A. Targeted therapies: how personal should we go? *Nat Rev Clin Oncol* 2012;9: 87–97.
- La Thangue NB, Kerr DJ. Predictive biomarkers: a paradigm shift towards personalized cancer medicine. *Nat Rev Clin Oncol* 2011;8: 587–96.
- Chapman PB, Hauschild A, Robert C, Haanen JB, Ascierto P, Larkin J, et al. Improved survival with vemurafenib in melanoma with BRAF V600E mutation. *N Engl J Med* 2011;364:2507–16.
- Kopetz S, Desai J, Chan E, Hecht JR, O'Dwyer PJ, Lee RJ, et al. PLX4032 in metastatic colorectal cancer patients with mutant BRAF tumors. *J Clin Oncol* 28:15s, 2010 (suppl; abstr 3534).
- Varadhachary GR. Carcinoma of unknown primary: focused evaluation. *J Natl Compr Canc Netw* 2011;9:1406–12.
- Anderson GG, Weiss LM. Determining tissue of origin for metastatic cancers: meta-analysis and literature review of immunohistochemistry performance. *Appl Immunohistochem Mol Morphol* 2010;18: 3–8.
- Levi F, Te VC, Erler G, Randimbison L, La Vecchia C. Epidemiology of unknown primary tumours. *Eur J Cancer* 2002;38:1810–2.
- DeYoung BR, Wick MR. Immunohistologic evaluation of metastatic carcinomas of unknown origin: an algorithmic approach. *Semin Diagn Pathol* 2000;17:184–93.
- Glover KY, Varadhachary GR, Lenzi R, Raber MN, Abbruzzese JL. Carcinoma of unknown primary. *Clinical oncology*. 4th ed. Philadelphia, PA: Churchill Livingstone; 2008. p. 2057–74.
- Tong KB, Murtagh KN, Hubert H, Santry B, Lenzi R, Hess KR, et al. Patient survival and health care utilization in Medicare beneficiaries diagnosed with cancer of unknown primary. *J Clin Oncol* 2006;24: 6064.
- Schapiro DV, Jarrett AR. The need to consider survival, outcome, and expense when evaluating and treating patients with unknown primary carcinoma. *Arch Intern Med* 1995;155:2050–4.
- Hamilton CS, Langlands AO. ACUPS (adenocarcinoma of unknown primary site): a clinical and cost benefit analysis. *Int J Radiat Oncol Biol Phys* 1987;13:1497–503.
- Pavlidis N, Merrouche Y. The importance of identifying CUP subsets. In: Fizazi K, editor. Carcinoma of unknown primary site. New York, NY: Taylor & Francis Group; 2006. p. 37–48.
- Bugat R, Bataillard A, Lesimple T, Voigt JJ, Culine S, Lortholary A, et al. Summary of the standards, options and recommendations for the management of patients with carcinoma of unknown primary site (2002). *Br J Cancer* 2003;89 Suppl 1:S59–66.
- Varadhachary GR, Raber MN, Matamoros A, Abbruzzese JL. Carcinoma of unknown primary with a colon-cancer profile-changing paradigm and emerging definitions. *Lancet Oncol* 2008;9:596–9.
- Hainsworth JD, Schnabel CA, Erlander MG, Haines DW III, Greco FA. A retrospective study of treatment outcomes in patients with carcinoma of unknown primary site and a colorectal cancer molecular profile. *Clin Colorectal Cancer* 2012;11:112–8.
- Rosenfeld N, Aharonov R, Meiri E, Rosenwald S, Spector Y, Zepeniuk M, et al. MicroRNAs accurately identify cancer tissue origin. *Nat Biotechnol* 2008;26:462–9.
- Ma XJ, Patel R, Wang X, Salunga R, Murage J, Desai R, et al. Molecular classification of human cancers using a 92-gene real-time quantitative polymerase chain reaction assay. *Arch Pathol Lab Med* 2006;130: 465–73.
- Varadhachary GR, Talantov D, Raber MN, Meng C, Hess KR, Jatke T, et al. Molecular profiling of carcinoma of unknown primary and correlation with clinical evaluation. *J Clin Oncol* 2008;26:4442–8.
- Monzon FA, Lyons-Weiler M, Buturovic LJ, Rigl CT, Henner WD, Sciulli C, et al. Multicenter validation of a 1,550-gene expression profile for identification of tumor tissue of origin. *J Clin Oncol* 2009;27:2503–8.
- Ramaswamy S, Tamayo P, Rifkin R, Mukherjee S, Yeang CH, Angelo M, et al. Multiclass cancer diagnosis using tumor gene expression signatures. *Proc Natl Acad Sci U S A* 2001;98:15149–54.
- Su AI, Welsh JB, Sapinoso LM, Kern SG, Dimitrov P, Lapp H, et al. Molecular classification of human carcinomas by use of gene expression signatures. *Cancer Res* 2001;61:7388–93.
- Bloom G, Yang IV, Boulware D, Kwong KY, Coppola D, Eschrich S, et al. Multi-platform, multi-site, microarray-based human tumor classification. *Am J Pathol* 2004;164:9–16.
- Tothill RW, Kowalczyk A, Rischin D, Bousioutas A, Haviv I, van Laar RK, et al. An expression-based site of origin diagnostic method designed

- for clinical application to cancer of unknown origin. *Cancer Res* 2005; 65:4031–40.
29. Erlander MG, Ma XJ, Kesty NC, Bao L, Salunga R, Schnabel CA. Performance and clinical evaluation of the 92-gene real-time PCR assay for tumor classification. *J Mol Diagn* 2011;13:493–503.
 30. Glas AS, Lijmer JG, Prins MH, Bonsel GJ, Bossuyt PM. The diagnostic odds ratio: a single indicator of test performance. *J Clin Epidemiol* 2003;56:1129–35.
 31. Pillai R, Deeter R, Rigl CT, Nystrom JS, Miller MH, Buturovic L, et al. Validation and reproducibility of a microarray-based gene expression test for tumor identification in formalin-fixed, paraffin-embedded specimens. *J Mol Diagn* 2011;13:48–56.
 32. Rosenwald S, Gilad S, Benjamin S, Lebanony D, Dromi N, Faerman A, et al. Validation of a microRNA-based qRT-PCR test for accurate identification of tumor tissue origin. *Mod Pathol* 2010;23:814–23.
 33. Raab SS, Grzybicki DM, Janosky JE, Zarbo RJ, Meier FA, Jensen C, et al. Clinical impact and frequency of anatomic pathology errors in cancer diagnoses. *Cancer* 2005;104:2205–13.
 34. Stang A, Pohlabein H, Muller KM, Jahn I, Giersiepen K, Jockel KH. Diagnostic agreement in the histopathological evaluation of lung cancer tissue in a population-based case-control study. *Lung Cancer* 2006;52:29–36.
 35. Stancel GA, Coffey D, Alvarez K, Halks-Miller M, Lal A, Mody D, et al. Identification of tissue of origin in body fluid specimens using a gene expression microarray assay. *Cancer Cytopathol* 2012;120: 62–70.
 36. Moraleda J, Grove N, Tran Q, Doan J, Hull J, Nguyen L, et al. Gene expression data analytics with interlaboratory validation for identifying anatomical sites of origin of metastatic carcinomas. *J Clin Oncol* 2004;22:9625.
 37. Schroeder BE, Laouri M, Chen E, Erlander MG, Schnabel CA. Pathological diagnoses in cases of indeterminate or unknown primary submitted for molecular tumor profiling [abstract]. In: Proceedings of the 101st Annual Meeting of the United States and Canadian Academy of Pathology; 2012 Mar 17–23. Vancouver, BC, Canada: Vancouver Convention Centre; 2012. Abstract nr. 429.
 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 carcinogenesis and its implications for screening for ovarian cancer. *Gynecol Oncol* 2011;120:470–3.
 39. Kurman RJ, Shih Ie M. Molecular pathogenesis and extraovarian origin of epithelial ovarian cancer—shifting the paradigm. *Hum Pathol* 2011;42:918–31.
 40. Wiegand KC, Shah SP, Al-Agha OM, Zhao Y, Tse K, Zeng T, et al. ARID1A mutations in endometriosis-associated ovarian carcinomas. *N Engl J Med* 2010;363:1532–43.
 41. Williams MD, Roberts DB, Kies MS, Mao L, Weber RS, El-Naggar AK. Genetic and expression analysis of HER-2 and EGFR genes in salivary duct carcinoma: empirical and therapeutic significance. *Clin Cancer Res* 2010;16:2266–74.
 42. Wick MR, Ockner DM, Mills SE, Ritter JH, Swanson PE. Homologous carcinomas of the breasts, skin, and salivary glands. A histologic and immunohistochemical comparison of ductal mammary carcinoma, ductal sweat gland carcinoma, and salivary duct carcinoma. *Am J Clin Pathol* 1998;109:75–84.
 43. Clauditz TS, Reiff M, Gravert L, Gnoss A, Tsourlakis MC, Munscher A, et al. Human epidermal growth factor receptor 2 (HER2) in salivary gland carcinomas. *Pathology* 2011;43:459–64.
 44. Sogawa K, Watanabe M, Sato K, Segawa S, Miyabe A, Murata S, et al. Rapid identification of microorganisms by mass spectrometry: improved performance by incorporation of in-house spectral data into a commercial database. *Anal Bioanal Chem* 2012;403:1811–22.
 45. Saffert RT, Cunningham SA, Ihde SM, Jobe KE, Mandrekar J, Patel R. Comparison of Bruker Biotyper matrix-assisted laser desorption ionization-time of flight mass spectrometer to BD Phoenix automated microbiology system for identification of gram-negative bacilli. *J Clin Microbiol* 2011;49:887–92.
 46. Guerriere-Kovach PM, Hunt EL, Patterson JW, Glembocki DJ, English JC III, Wick MR. Primary melanoma of the skin and cutaneous melanomatous metastases: comparative histologic features and immunophenotypes. *Am J Clin Pathol* 2004;122:70–7.
 47. Montone KT, van Belle P, Elenitsas R, Elder DE. Proto-oncogene c-kit expression in malignant melanoma: protein loss with tumor progression. *Mod Pathol* 1997;10:939–44.
 48. Bussolati G, Leonardo E. Technical pitfalls potentially affecting diagnoses in immunohistochemistry. *J Clin Pathol* 2008;61:1184–92.
 49. Leong TY, Cooper K, Leong AS. Immunohistochemistry—past, present, and future. *Adv Anat Pathol* 2010;17:404–18.
 50. Leong AS. Pitfalls in diagnostic immunohistochemistry. *Adv Anat Pathol* 2004;11:86–93.
 51. Agoff SN, Lamps LW, Philip AT, Amin MB, Schmidt RA, True LD, et al. Thyroid transcription factor-1 is expressed in extrapulmonary small cell carcinomas but not in other extrapulmonary neuroendocrine tumors. *Mod Pathol* 2000;13:238–42.
 52. Kubba LA, McCluggage WG, Liu J, Malpica A, Euscher ED, Silva EG, et al. Thyroid transcription factor-1 expression in ovarian epithelial neoplasms. *Mod Pathol* 2008;21:485–90.