VOLUME 27 · NUMBER 15 · MAY 20 2009

JOURNAL OF CLINICAL ONCOLOGY

ORIGINAL REPORT

Multicenter Validation of a 1,550-Gene Expression Profile for Identification of Tumor Tissue of Origin

Federico A. Monzon, Maureen Lyons-Weiler, Ljubomir J. Buturovic, C. Ted Rigl, W. David Henner, Christin Sciulli, Catherine I. Dumur, Fabiola Medeiros, and Glenda G. Anderson

From The Methodist Hospital and The Methodist Hospital Research Institute, Houston, TX; the Clinical Genomics Facility and Department of Pathology, University of Pittsburgh, Pittsburgh, PA; Pathwork Diagnostics, Sunnyvale, CA; Department of Pathology, VA Commonwealth University, Richmond, VA; and the Department of Laboratory Medicine and Pathology, Division of Laboratory Genetics, Mayo Clinic, Rochester, MN.

Submitted May 5, 2008; accepted December 2, 2008; published online ahead of print at www.jco.org on March 30, 2009.

Supported by a sponsored research agreement from Pathwork Diagnostics, Sunnyvale, CA (F.A.M., with a subcontract to C.I.D.).

Authors' disclosures of potential conflicts of interest and author contributions are found at the end of this article.

Presented in part in poster format at the Annual Meeting of the Association for Molecular Pathology, Los Angeles, CA, November 7-10, 2007; and the Annual Meeting of the United States and Canadian Academy of Pathology, San Diego, CA, March 24-30, 2007.

Corresponding author: Federico A. Monzon, MD, Medical Director of Molecular Diagnostics, The Methodist Hospital, 6565 Fannin St, MS 205, Houston, TX 77030; e-mail: famonzon@tmhs.org.

The Appendix is included in the full-text version of this article, available online at www.jco.org. It is not included in the PDF version (via Adobe® Reader®).

© 2009 by American Society of Clinical Oncology

0732-183X/09/2715-2503/$20.00

DOI: 10.1200/JCO.2008.17.9762

A B S T R A C T

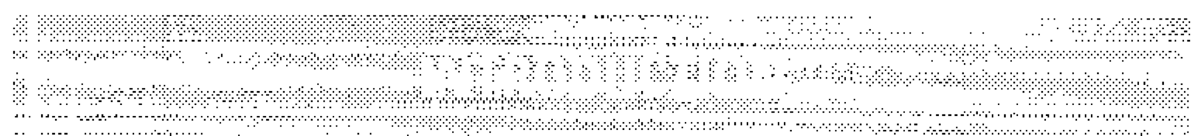
Purpose
Malignancies found in unexpected locations or with poorly differentiated morphologies can pose a significant challenge for tissue of origin determination. Current histologic and imaging techniques fail to yield definitive identification of the tissue of origin in a significant number of cases. The aim of this study was to validate a predefined 1,550-gene expression profile for this purpose.

Methods
Four institutions processed 547 frozen specimens representing 15 tissues of origin using oligonucleotide microarrays. Half of the specimens were metastatic tumors, with the remainder being poorly differentiated and undifferentiated primary cancers chosen to resemble those that present as a clinical challenge.

Results
In this blinded multicenter validation study the 1,550-gene expression profile was highly informative in tissue determination. The study found overall sensitivity (positive percent agreement with reference diagnosis) of 87.8% (95% CI, 84.7% to 90.4%) and overall specificity (negative percent agreement with reference diagnosis) of 99.4% (95% CI, 98.3% to 99.9%). Performance within the subgroup of metastatic tumors (n = 258) was found to be slightly lower than that of the poorly differentiated and undifferentiated primary tumor subgroup, 84.5% and 90.7%, respectively (P = .04). Differences between individual laboratories were not statistically significant.

Conclusion
This study represents the first adequately sized, multicenter validation of a gene-expression profile for tissue of origin determination restricted to poorly differentiated and undifferentiated primary cancers and metastatic tumors. These results indicate that this profile should be a valuable addition or alternative to currently available diagnostic methods for the evaluation of uncertain primary cancers.

J Clin Oncol 27:2503-2508. © 2009 by American Society of Clinical Oncology



Evidence-based management indicates that a thorough investigation of uncertain primary cancers should be performed to assist in therapeutic decisions.[1,2] This is typically carried out with immunohistochemistry (IHC) panels on the tumor specimen, and advanced whole body or site-directed imaging tests.[1,3-5] This work-up is associated with considerable resources, time, and expense[1,6,7]; however, the primary site remains unidentified in up to 30% of patients who present with an uncertain primary cancer.[1,8,9] Thus new approaches are needed to reduce diagnostic uncertainty in these patients. The use of gene expression–based signatures for classifying tumor tissue of origin (TOO) has been reported,[10-14] and these studies indicated that metastatic and poorly-differentiated specimens pose a significant challenge to gene expression–based classifiers.

To our knowledge, we present the first blinded, multicenter validation study conducted on a gene expression–based test to identify the tissue of origin, the Pathwork Tissue of Origin Test (Pathwork Diagnostics, Sunnyvale, CA). An interlaboratory reproducibility study of the 1,550-gene expression profile has been described previously.[15] Two important aspects of this study are: it is the first clinical validation of significant size (> 500 specimens) to be performed on a test for TOO; and it is the only reported study conducted entirely with metastatic

Information downloaded from jco.ascopubs.org and provided by HAM-TMC LIBRARY on May 18, 2009 from 206.83.53.249.
Copyright © 2009 by the American Society of Clinical Oncology. All rights reserved.

tumors and poorly differentiated or undifferentiated primary tumors chosen to resemble the expected population of difficult to diagnose cancers.

METHODS

Patients and Tumor Specimens

Tumor specimens or tumor-derived microarray gene expression files from 622 patients were screened for inclusion. Three hundred fifty-one frozen tissue specimens were obtained from the Health Sciences Tissue Bank at the University of Pittsburgh (UPitt), the Mayo Clinic tissue bank, and commercial providers: Cytomyx (Lexington, MA), Proteogenix (Culver City, CA), and Asterand (Detroit, MI). In addition, electronic files of microarray data on 271 tumors were obtained from the International Genomics Consortium (IGC; Phoenix, AZ). Criteria for inclusion for frozen specimens were: ≥ 0.1 g of frozen tissue, histologic verification of minimal necrosis (≤ 20% of tumor tissue), and sufficient tumor representation (≥ 60% of tissue examined). Histologic verification was performed by a pathologist at the institution providing the tissue sample, who visually estimated the percent tumor cells. Inclusion criteria for all specimens (tissues and microarray files) were: characterization as a poorly differentiated or undifferentiated primary tumor (American Joint Committee on Cancer grade 3 or 4, or "high grade" in pathology report), or a metastatic tumor; and classification by the original pathology report as one of the 15 tissue types on the Pathwork TOO test panel (Data Supplement Table 1, online only). Sixteen specimens were excluded due to off-panel morphology: 45 due to less than 60% tumor content, 23 due to more than 20% necrosis, and six due to microarray quality control failures. A total of 547 specimens met all inclusion criteria for the validation analyses, with no fewer than 25 specimens for each of the 15 tissues on the panel. Characteristics of patients and tumor specimens are presented in Table 1. All specimens were collected and de-identified under institutional review board approved protocols.

Specimen Processing and Gene Expression Assays

Each specimen processing laboratory was trained to perform the test, and proficiency in performing the assay at each laboratory was verified using known total RNA samples and tissue from known specimens (n = 8 to 10). All laboratories obtained the correct tissue identification for these performance training samples (data not shown).

For the 547 specimens in the validation cohort, 276 frozen tumor tissues were processed at the Clinical Genomics Facility of UPitt, Cogenics (Morrisville, NC), and the Mayo Clinic as outlined in Figure 1. Tissue processing methods have been previously described and additional details are presented in Data Supplement Table 2 (online only).[15] Samples were hybridized to one of three microarrays: Pathwork Diagnostics Pathchip, Affymetrix GeneChip HG-U133A or HG-U133 Plus 2. The arrays were scanned using the Affymetrix GCS3000 scanner and intensity levels calculated using Affymetrix GCOS 1.1.3 or 1.4. The resulting raw intensity data files (.CEL), including the 271-gene expression data files from IGC, were processed at Pathwork Diagnostics for automated analysis and report generation. Probe-level intensity data were transformed into gene expression values and standardized using the 121-gene standardization method whose performance has been previously described,[15,16] before applying the 1,550-gene profile. Data from the 276 frozen tumor specimens have been deposited to the NCBI Gene Expression Omnibus (GEO)[17] under series accession number GSE12630. GEO accession numbers for the 271-gene expression data files from IGC are listed in Data Supplement Table 3 (online only).

1,550-Gene Profile for Tumor Tissue of Origin Identification

The 1,550-gene profile was trained using gene expression data files from a panel of 2,039 tumors comprising 15 tissue types and 60 different morphologies, as illustrated in Figure 2 and detailed in Data Supplement Table 1. The training set included both primary and metastatic tumors and well-differentiated to undifferentiated tumors. None of the validation specimens were used for algorithm training.

The 1,550-gene profile is a proprietary algorithm that uses the expression level of 1,550 transcripts to perform pair-wise comparisons between the test sample and each of the 15 tissues on the test panel. The results are presented as 15 similarity scores, one for each tissue included in the test panel.

Before analysis of the clinical validation study data, the 1,550-gene profile was locked based on its performance with the training data. Similarity score thresholds for determining absence and presence of tissue in the sample were also locked. The similarity scores were probability based, with a reported range from 0 to 100, and all 15 scores sum to 100. A similarity score of 30 or above indicates the presence of a given tissue in the specimen; a similarity score of 5 or less indicates the absence of a given tissue. Similarity scores between 5 and 30 are considered indeterminate. These criteria were used to make a tissue determination for each specimen.

The Pathwork System Software and 1,550-gene profile produced an automated report (Fig 1) for each specimen. An assessment of the biologic

Table 1. Patients and Tumors Characteristics Included in This Study

Characteristic	No.	%
Tumor		
Metastatic	258	47
Primary		
Grade 3	185	34
Grade 4	68	12
Not graded*	36	7
Patient		
Age, years†		
< 50	142	26
50-59	133	24
60-69	139	25
≥ 70	132	24
Sex‡		
Male	254	46
Female	290	53

*Melanoma, thyroid, and lymphoma tumors are not normally graded.
†Age data were available for 546 of 547 patients.
‡Sex data were available for 544 of 547 patients.

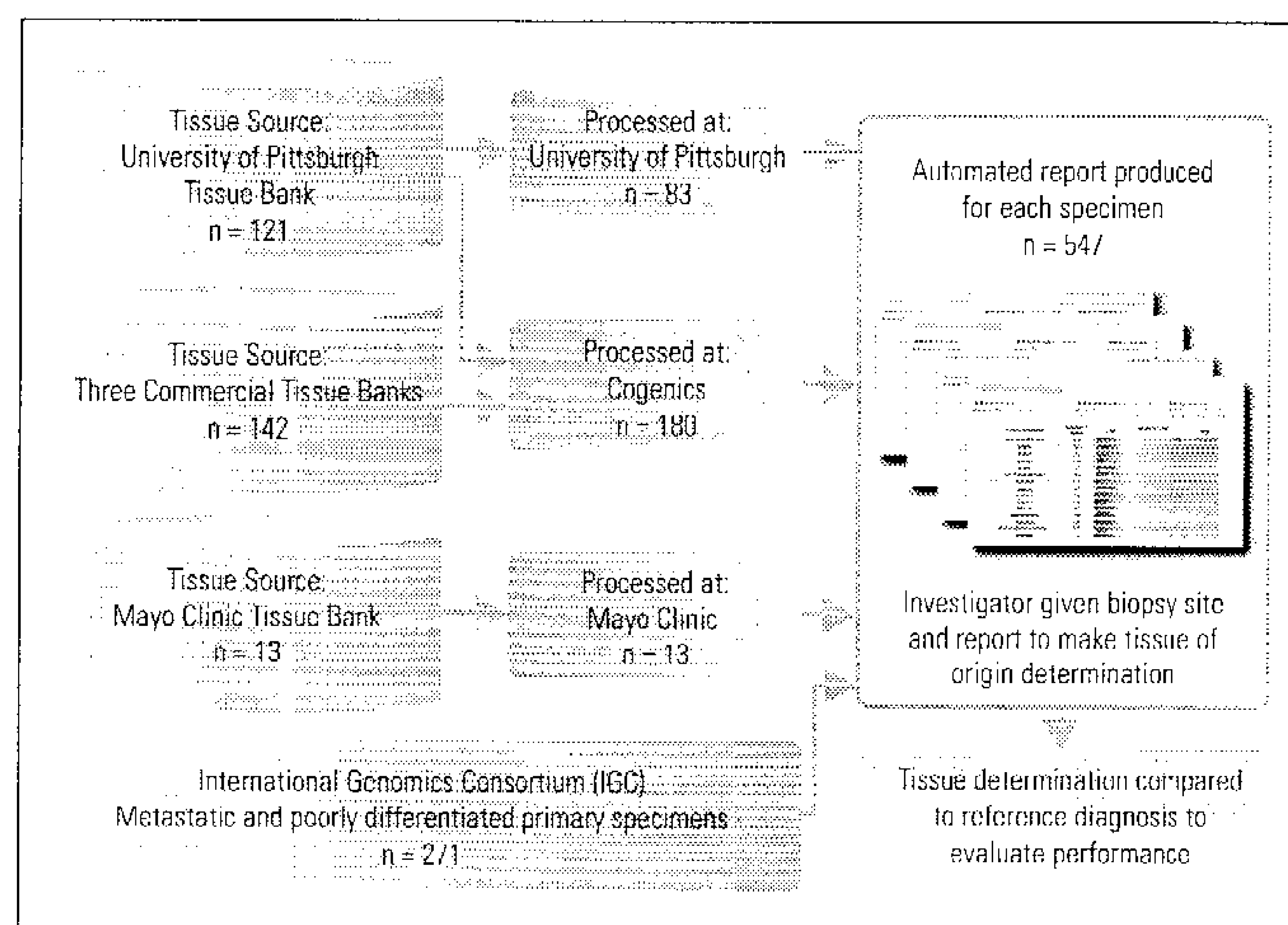


Fig 1. Validation study design. Gene expression data from 547 tumor samples generated by multiple laboratories were processed by the Pathwork Tissue of Origin test (Pathwork Diagnostics, Sunnyvale, CA) software. The test software transformed data into gene expression values, performed data verification and standardization, and generated reports that were evaluated in a blinded fashion by the investigators.

 © 2009 by American Society of Clinical Oncology

Information downloaded from jco.ascopubs.org and provided by HAM-TMC LIBRARY on May 18, 2009 from 206.83.53.249.
Copyright © 2009 by the American Society of Clinical Oncology. All rights reserved.

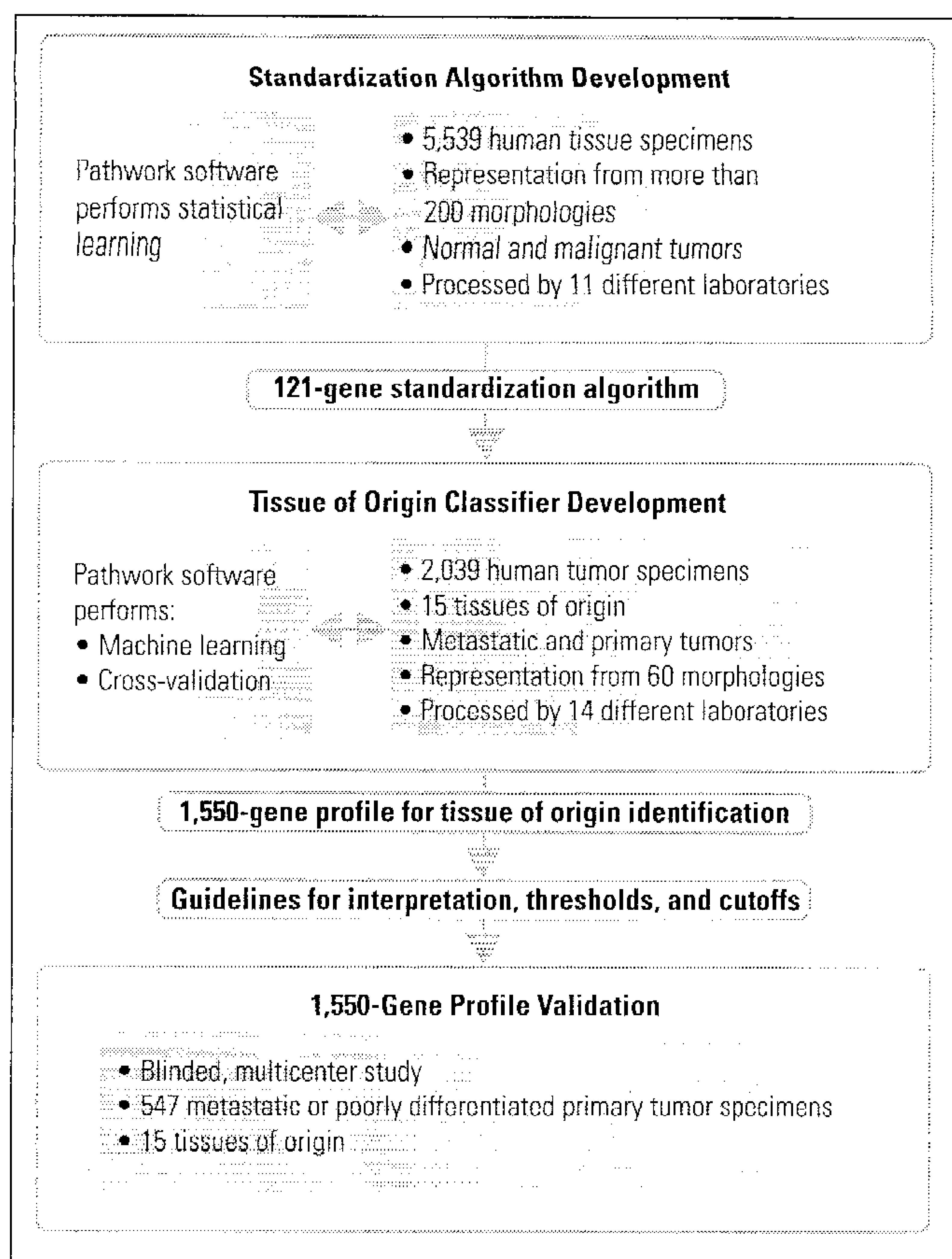


Fig 2. Development of the 1,550-gene profile to identify tissue of origin. A 121-gene standardization algorithm was used. The 1,550-gene profile for tissue of origin identification was trained using 2,039 primary and metastatic tumors. The algorithm was locked and thresholds for positive, negative, and indeterminate calls were predetermined before the multicenter validation.

plausibility for gene-tissue associations for the 60 genes with the strongest correlations with individual tissues is available in the online only Appendix.

Validation Study Design

The objective of this study was to determine the performance characteristics of the Pathwork Tissue of Origin Test in the identification of TOO for a series of metastatic and poorly differentiated or undifferentiated primary tumor specimens of known origin, which was considered the reference diagnosis. These specimens are representative of those that would likely be designated as uncertain primary cancer after initial histologic evaluation. The study evaluated agreement between the tissue determination made using the 1,550-gene profile and the reference diagnosis for each specimen. We also evaluated the nonagreement and indeterminate fractions.

Technical personnel performing the gene expression assays and investigators who interpreted the Pathwork Tissue of Origin Test results for making a tissue determination were blinded to patient sex, histology, or morphology information, and reference diagnosis. When making the tissue determination, investigators were provided only biopsy site and the 15 similarity scores for each specimen. Matching of reference diagnosis and the predicted site of origin was performed by an investigator not involved with any aspect of sample processing or tissue determination who was blinded to all the above information. Results were stratified by type of tissue (primary *v* metastatic), by processing site, and by site of origin, all of which are potential sources of variability.

Statistical Methods

Power calculations were based on the estimated 88% sensitivity found in cross-validation analyses of the training data set. Sample size was determined by calculating the minimum number of samples needed to detect a 5% reduction in performance (ie, a decrease from 88% to 83% sensitivity), determined to be clinically significant. One-tailed calculations indicated that 540 specimens would provide 95% power to detect this difference at a significance level of .05. We targeted no fewer than 25 samples per tissue type with a distribution reflecting the incidence of individual cancers, subject to specimen availability.

For each specimen, a tissue determination was made using the reported similarity scores and criteria described earlier, and compared to the reference diagnosis. A true-positive result was indicated when the tissue determination matched the reference diagnosis. When the tissue determination and the reference diagnosis did not match, the specimen was considered a false positive. For each tissue on the panel, sensitivity (or positive percent agreement) was defined as the ratio of true positive results to the total positive samples analyzed. Specificity (or negative percent agreement) was defined as the ratio: (1 − false positive)/(total tested including indeterminate − total positive). Diagnostic odds ratio was calculated as (sensitivity/(1 − specificity)/((1 − specificity)/sensitivity).[18]

RESULTS

Agreement With Reference Diagnosis

The 1,550-gene profile results showed 87.8% overall agreement with the reference diagnosis (480 of 547; 95% CI, 84.7% to 90.4%) for the 547 specimens. The overall sensitivity (positive percent agreement) and specificity (negative percent agreement) were 87.8% (95% CI, 84.7% to 90.4%) and 99.4% (95% CI, 98.3% to 99.9%), respectively (Table 2). Diagnostic odds ratios for all tissues are significantly greater than one, indicating that each of the individual tests is highly informative. Similarity scores reported for each of the 15 tissues on the panel for all samples are provided in Data Supplement Table 3. Overall rate of nonagreement for these specimens was 7.1% (39 of 547; 95% CI, 5.1% to 9.6%), and the rate of indeterminate calls was 5.1% (28 of 547; 95% CI, 3.4% to 7.3%; Table 3 and Data Supplement Table 4).

Analysis by Relevant Subgroups

The rates of agreement between the test result and the reference diagnosis ranged from 94.1% for breast cancer specimens (n = 68) to 72.0% for gastric and pancreatic cancer specimens (n = 25 each; Table 3). Performance differences between tissue sites were statistically significant (χ^2 = 42.02; *P* = .04; *df* = 28; n = 547).

Performance of the test was found to be somewhat better with primary tumors (90.7% agreement; n = 289) than with metastatic specimens (84.5% agreement; n = 258) (Fisher's exact method two-sided *P* = .04). Rates of agreement between the test result and the reference diagnosis were 88.0%, 84.4%, 92.3%, and 89.7% at study sites 1 (Clinical Genomics Facility), 2 (Cogenics), 3 (Mayo Clinic), and 4 (IGC), respectively, and these differences were not statistically significant (χ^2 = 4.4, *P* = .62; *df* = 6; n = 547).

Nonagreements and Indeterminates

Of the 39 tissue determinations that were in nonagreement with the reference diagnosis, 11 matched the biopsy site for that sample. Of the 28 specimens with indeterminate results, 25 reported no similarity score above 30, and three reported two similarity scores greater than 30, neither of which could be excluded as the biopsy site. In 11 of these 28 indeterminate samples, the highest similarity score was that of the reference diagnosis tissue, and in only one result was the reference diagnosis ruled out due to a similarity score less than 5. When the 28

Information downloaded from jco.ascopubs.org and provided by HAM-TMC LIBRARY on May 18, 2009 from 206.83.53.249.
Copyright © 2009 by the American Society of Clinical Oncology. All rights reserved.

Table 2. Sensitivity and Specificity of the 1,550-Gene Profile for Tissue of Origin Identification

	Sample		Sensitivity			Specificity		
Reference Diagnosis	Algorithm Development	Multicenter Validation	Positive % Agreement	Ratio	95% CI	Negative % Agreement	Ratio	95% CI
Bladder	62	28	78.6	22/28	59.0 to 91.7	100.0	519/519	99.3 to 100.0
Breast	444	68	94.1	64/68	85.6 to 98.4	98.3	471/479	96.7 to 99.3
Colorectal	253	56	92.9	52/56	82.7 to 98.0	99.2	487/491	97.9 to 99.9
Gastric	52	25	72.0	18/25	50.6 to 87.9	99.4	519/522	98.3 to 99.9
Germ cell	121	30	73.3	22/30	54.1 to 87.7	100.0	517/517	99.3 to 100.0
Hepatocellular	151	25	92.0	23/25	74.0 to 99.0	99.8	521/522	98.8 to 100.0
Kidney	41	39	94.9	37/39	82.7 to 99.4	99.8	507/508	98.9 to 100.0
Melanoma	221	26	80.8	21/26	60.6 to 93.4	99.8	520/521	98.9 to 100.0
Non-Hodgkin's lymphoma	97	33	93.9	31/33	79.8 to 99.3	99.4	511/514	98.3 to 99.9
Non–small cell lung	69	31	87.1	27/31	70.2 to 96.4	98.6	509/516	97.2 to 99.5
Ovarian	189	69	92.8	64/69	83.9 to 97.6	99.0	473/478	97.6 to 99.7
Pancreas	43	25	72.0	18/25	50.6 to 87.9	99.8	521/522	98.9 to 100.0
Prostate	105	26	88.5	23/26	69.8 to 97.6	100.0	521/521	99.3 to 100.0
Soft to tissue sarcoma	122	31	83.9	26/31	66.3 to 94.5	99.4	513/516	98.3 to 99.9
Thyroid	69	35	91.4	32/35	76.9 to 98.2	99.6	510/512	98.6 to 100.0
Overall	2,039	547	87.8	480/547	84.7 to 90.4	99.4	NA	98.3 to 99.9

indeterminate results were excluded, the overall accuracy was 92.5% (480 of 519).

Gene expression–based classifiers for clinical applications should demonstrate strong reproducibility in sample processing, analytic performance, and clinical reported result. In this study, we show that the Pathwork Tissue of Origin Test can reliably identify the TOO in 87.8% of the 547specimens tested, and in 84.5% of the metastatic specimens. This compares favorably with current clinical practice standards, such as IHC, which has shown 66% to 88% agreement in blinded tests.[19-22] The performance of this test also compares favorably with other gene expression–based TOO classifiers with reported accuracies in the range of 76% to 89%.[10-14,23,24] Moreover, the results of this clinical validation study are consistent with the 86.8% agreement reported in our previous study.[15]

Published gene expression–based studies that show possible clinical application are criticized for one or more common flaws: reuse of the training samples in reported results, post hoc modification of the algorithm or thresholds, inadequate blinding, inadequate study size, and inappropriate handling of indeterminate results in reported performance.[25,26] Many groups have published multigene algorithms and results that demonstrate the promise of gene expression–based classifiers in TOO identification.[10-14,23,24] These studies have been restricted to smaller numbers of specimens (< 120), often dominated by well-differentiated primary cancers, and have often allowed post hoc modifications or enhancements to the algorithm design or thresholds. For example, in the study by Ma and coauthors where a panel of 92 genes was developed to identify 32 different tumor types, the same training set was repeatedly used to test different iterations of the classifier, and the final performance was evaluated in 119 tumors where representation from each tumor type ranged from 1 to 10 specimens.[23] Thus, in this test, correct identification of one single specimen was interpreted as 100% accuracy for that tissue type. Likewise, in a recent study by Rosenfeld et al, performance of a microRNA–based classifier was evaluated in 83 specimens, and representation of each of 22 tissue types ranged from 2 to 8 samples.[27] Clearly, these studies were inadequately sized to establish true diagnostic performance. In contrast, this validation study used 547 independent specimens with minimum tissue representation of 25 samples. Furthermore, Rosenfeld et al allowed post hoc enhancement of the test's performance by introducing a combination union classifier where sensitivity was calculated based on correct identification of TOO by either one of two algorithms (decision tree or k-nearest neighbor). Overall accuracy for the decision tree alone was 72% (60 of 83) for all samples and 59% (13 of 22) when only metastatic tumor samples were considered.

This is, to our knowledge, the largest clinical validation study of a gene expression assay for TOO determination to date. The study was designed and executed to avoid the common flaws mentioned earlier: all of the specimens used in the validation of the test were newly acquired; the algorithm was locked and thresholds predetermined based on the training set before the analysis of the validation specimens; indeterminate results are appropriately included in the reported performance; specimen identity was masked until the final analysis; and this study is the first to be adequately sized to provide performance data sufficient to support clinical use of a microarray-based test for TOO determination. Other strengths of this study are the wide range of tissues of origin evaluated, the characteristics of the challenging specimens, and the use of multiple laboratories for tissue processing and microarray analysis.

In a clinical scenario, the uncertainty of a tumor's origin usually arises in the context of metastatic and/or poorly differentiated to undifferentiated malignancies, and some of the previously published gene expression–based classifiers have shown decreased performance with less differentiated tumors.[12] Our results show that this test can identify the tissue of origin in poorly differentiated and undifferentiated tumor specimens, which is the clinically relevant population, since well-differentiated tumors rarely present a diagnostic challenge. Interestingly, we found a small but statistically significant reduction in the accuracy of the test when primary cancers and

 © 2009 by American Society of Clinical Oncology

Information downloaded from jco.ascopubs.org and provided by HAM-TMC LIBRARY on May 18, 2009 from 206.83.53.249.
Copyright © 2009 by the American Society of Clinical Oncology. All rights reserved.

Table 3. Effect of Possible Sources of Variability in Tumor Tissue of Origin Test Performance

Performance by	No. of Specimens	Agreement		Nonagreement		Indeterminate	
		No.	%	No.	%	No.	%
Reference diagnosis*							
Bladder	28	22*	78.6	4	14.3	2	7.1
Breast	68	64	94.1	4	5.9	0	< 0.1
Colorectal	56	52	92.9	4	7.1	0	< 0.1
Gastric	25	18	72.0	4	16.0	3	12.0
Germ cell	30	22	73.3	3	10.0	5	16.7
Hepatocellular	25	23	92.0	0	< 0.1	2	8.0
Kidney	39	37	94.9	1	2.6	1	2.6
Melanoma	26	21	80.8	2	7.7	3	11.5
Non-Hodgkin's lymphoma	33	31	93.9	1	3.0	1	3.0
Non–small-cell lung	31	27	87.1	2	6.5	2	6.5
Ovarian	69	64	92.8	3	4.3	2	2.9
Pancreas	25	18	72.0	5	20.0	2	8.0
Prostate	26	23	88.5	1	3.8	2	7.7
Soft tissue sarcoma	31	26	83.9	3	9.7	2	6.5
Thyroid	35	32	91.4	2	5.7	1	2.9
Overall	547	480	87.8	39	7.1	28	5.1
Overall 95% CI		84.7 to 90.4		5.1 to 9.6		3.4 to 7.3	
Metastatic *v* primary tumor samples†							
Metastatic	258	218†	84.5	23	8.9	17	6.6
Poorly and undifferentiated primary	289	262	90.7	16	5.5	11	3.8
At each processing laboratory‡							
IGC	271	243†	89.7	18	6.6	10	3.7
Cogenics	180	152	84.4	15	8.3	13	7.2
CGF-UPitt	83	73	88.0	5	6.0	5	6.0
Mayo clinic	13	12	92.3	1	7.7	0	< 0.1

Abbreviations: IGC, International Genomics Consortium; CGF, Clinical Genomics Facility; UPitt, University of Pittsburgh.
*χ^2 = 42.02; P = .04; df = 28; N = 547.
†Fisher's exact method two-sided P = .04.
‡χ^2 = 4.4; P = .62; df = 6; N = 547.

metastatic tumors were compared (90.7% and 84.5%, respectively). However, the performance in the metastatic samples still compares favorably with IHC, which is the current standard of care for tissue of origin identification. Importantly, similarly sized validation studies of IHC panels in clinical use today have not been performed, and in one of the largest blinded studies of IHC performance, Dennis and coauthors[20] reported 67% accuracy (20 of 30) in metastatic samples using a predetermined panel of ten antibodies.

One of the limitations of our study was the inability to independently verify the reference diagnosis used to assess the accuracy of the test. The diagnosis was extracted from the surgical pathology report that accompanied the specimen at the time it was banked. It is possible that some of these diagnoses are incorrect and this could result in an over- or underestimation of the test's accuracy. Another limitation is the requirement for frozen tissue. In many instances, the need to perform a tissue of origin determination is not known until after the specimen has been fixed. Although for this study we specified the need for > 0.1 g of tumor tissue, the assay requires 1 μg of total RNA; this quantity is obtainable from a needle core specimen if adequate tumor representation is present. However, validation of needle core biopsy material and/or formalin-fixed paraffin embedded tissues should be performed in separate studies.

This test is designed to be interpreted by a pathologist in conjunction with pathologic examination of the tissue and in consultation with the surgeon/oncologist. This is especially important in patients where the differential between a primary and a metastatic tumor is being considered, since the metastatic tumor specimens are expected to contain surrounding noncancerous tissue from the biopsy site. Due to the blinded nature of the study, the pathologists interpreting the TOO test results did not know the morphologic features of the specimen and/or the clinical features of the patient. It is expected that the clinical performance of the test will be favorably influenced by the availability of this information. In addition, it is important to note that although the test was trained and validated on a preselected panel of 15 tumor types which represent approximately 89% of the incident solid tumors[28] that are known to produce distant metastases, the possibility that an uncertain primary cancer might originate from a tissue site not covered by the panel must be considered. It is also important to acknowledge that in certain clinical situations, the need to test a sample that does not meet the quality control criteria for the test (≥ 60% tumor and ≤ 20% necrosis) could arise. As described previously, the best assay performance is achieved when these two criteria are met,[15] but there are insufficient data to adequately determine the impact of testing suboptimal specimens. Furthermore, the assay has been approved by the US Food and Drug Administration based on the stated sample quality thresholds.

In conclusion, this study represents the first adequately sized, multicenter validation of a prespecified diagnostic test for tissue of origin determination restricted to poorly differentiated and undifferentiated primary cancers and metastatic tumors. Our results confirm the diagnostic value of the 1,550 gene profile used in the Pathwork Tissue of Origin Test. This test should be a valuable addition to

© 2009 by American Society of Clinical Oncology

Information downloaded from jco.ascopubs.org and provided by HAM-TMC LIBRARY on May 18, 2009 from 206.83.53.249.
Copyright © 2009 by the American Society of Clinical Oncology. All rights reserved.

alternative to currently available diagnostic methods for the evaluation of uncertain primary cancers.

AUTHORS' DISCLOSURES OF POTENTIAL CONFLICTS OF INTEREST

Although all authors completed the disclosure declaration, the following author(s) indicated a financial or other interest that is relevant to the subject matter under consideration in this article. Certain relationships marked with a "U" are those for which no compensation was received; those relationships marked with a "C" were compensated. For a detailed description of the disclosure categories, or for more information about ASCO's conflict of interest policy, please refer to the Author Disclosure Declaration and the Disclosures of Potential Conflicts of Interest section in Information for Contributors.

Employment or Leadership Position: Ljubomir J. Buturovic, Pathwork Diagnostics (C); C. Ted Rigl, Pathwork Diagnostics (C); W. David Henner, Pathwork Diagnostics (C); Glenda G. Anderson, Pathwork Diagnostics (C) **Consultant or Advisory Role:** Federico A. Monzon, Pathwork Diagnostics (C) **Stock Ownership:** Ljubomir J. Buturovic, Pathwork Diagnostics; C. Ted Rigl, Pathwork Diagnostics; W. David Henner, Pathwork Diagnostics; Glenda G. Anderson, Pathwork Diagnostics **Honoraria:** Federico A. Monzon, Pathwork Diagnostics **Research Funding:** Federico A. Monzon, Pathwork Diagnostics **Expert Testimony:** None **Other Remuneration:** None

AUTHOR CONTRIBUTIONS

Conception and design: Federico A. Monzon, Ljubomir J. Buturovic, C. Ted Rigl, Glenda G. Anderson
Financial support: Glenda G. Anderson
Administrative support: Maureen Lyons-Weiler, C. Ted Rigl, Christin Sciulli
Provision of study materials or patients: Federico A. Monzon, Maureen Lyons-Weiler, C. Ted Rigl, Fabiola Medeiros
Collection and assembly of data: Federico A. Monzon, Maureen Lyons-Weiler, C. Ted Rigl, Christin Sciulli
Data analysis and interpretation: Federico A. Monzon, Ljubomir J. Buturovic, C. Ted Rigl, W. David Henner, Catherine I. Dumur, Fabiola Medeiros, Glenda G. Anderson
Manuscript writing: Federico A. Monzon, Ljubomir J. Buturovic, W. David Henner, Catherine I. Dumur, Fabiola Medeiros, Glenda G. Anderson
Final approval of manuscript: Federico A. Monzon, Ljubomir J. Buturovic, W. David Henner, Catherine I. Dumur, Glenda G. Anderson

REFERENCES

1. National Comprehensive Cancer Network: NCCN Clinical Practice Guidelines in Oncology, Occult Primary, v. 1.2007. http://www.nccn.org/professionals/physician_gls/PDF/occult.pdf
2. Briasoulis E, Tolis C, Bergh J, et al: ESMO Guidelines Task Force: ESMO minimum clinical recommendations for diagnosis, treatment and follow-up of cancers of unknown primary site (CUP). Ann Oncol 16:i75-i76, 2005 (suppl 1)
3. Pavlidis N, Briasoulis E, Hainsworth J, et al: Diagnostic and therapeutic management of cancer of an unknown primary. Eur J Cancer 39:1990-2005, 2003
4. Bugat R, Bataillard A, Lesimple T, et al: FNCLCC: Summary of the standards, options and recommendations for the management of patients with carcinoma of unknown primary site (2002). Br J Cancer 89:S59-S66, 2003 (suppl 1)
5. Varadhachary GR, Abbruzzese JL, Lenzi R: Diagnostic strategies for unknown primary cancer. Cancer 100:1776-1785, 2004
6. Chu PG, Weiss LM: Keratin expression in human tissues and neoplasms. Histopathology 40: 403-439, 2002
7. Schapira DV, Jarrett AR: The need to consider survival, outcome, and expense when evaluating and treating patients with unknown primary carcinoma. Arch Intern Med 155:2050-2054, 1995
8. Pavlidis N, Merrouche Y: The importance of identifying CUP subsets, in Fizazi K (ed): Carcinoma of Unknown Primary Site. New York, NY, Taylor & Francis Group, 2006, pp 37-48
9. Hillen HF: Unknown primary tumours. Postgrad Med J 76:690-693, 2000
10. Bloom G, Yang IV, Boulware D, et al: Multiplatform, multi-site, microarray-based human tumor classification. Am J Pathol 164:9-16, 2004
11. Buckhaults P, Zhang Z, Chen YC, et al: Identifying tumor origin using a gene expression-based classification map. Cancer Res 63:4144-4149, 2003
12. Ramaswamy S, Tamayo P, Rifkin R, et al: Multiclass cancer diagnosis using tumor gene expression signatures. Proc Natl Acad Sci U S A 98:15149-15154, 2001
13. Su AI, Welsh JB, Sapinoso LM, et al: Molecular classification of human carcinomas by use of gene expression signatures. Cancer Res 61:7388-7393, 2001
14. Tothill RW, Kowalczyk A, Rischin D, et al: An expression-based site of origin diagnostic method designed for clinical application to cancer of unknown origin. Cancer Res 65:4031-4040, 2005
15. Dumur CI, Lyons-Weiler M, Sciulli C, et al: Interlaboratory performance of a microarray-based gene expression test to determine tissue of origin in poorly differentiated and undifferentiated cancers. J Mol Diagn 10:67-77, 2008
16. Moraleda J, Grove N, Tran Q, et al: Gene expression data analytics with interlaboratory validation for identifying anatomical sites of origin of metastatic carcinomas. J Clin Oncol 22:862s, 2004 (suppl; abstr 9625)
17. Edgar R, Domrachev M, Lash AE: Gene Expression Omnibus: NCBI gene expression and hybridization array data repository. Nucleic Acids Res 30:207-210, 2002
18. Glas AS, Lijmer JG, Prins MH, et al: The diagnostic odds ratio: A single indicator of test performance. J Clin Epidemiol 56:1129-1135, 2003
19. Brown RW, Campagna LB, Dunn JK, et al: Immunohistochemical identification of tumor markers in metastatic adenocarcinoma: A diagnostic adjunct in the determination of primary site. Am J Clin Pathol 107:12-19, 1997
20. Dennis JL, Hvidsten TR, Wit EC, et al: Markers of adenocarcinoma characteristic of the site of origin: Development of a diagnostic algorithm. Clin Cancer Res 11:3766-3772, 2005
21. DeYoung BR, Wick MR: Immunohistologic evaluation of metastatic carcinomas of unknown origin: An algorithmic approach. Semin Diagn Pathol 17:184-193, 2000
22. Park S-Y, Kim B-H, Kim J-H, et al: Panels of immunohistochemical markers help determine primary sites of metastatic adenocarcinoma. Arch Pathol Lab Med 131:1561-1567, 2007
23. Ma XJ, Patel R, Wang X, et al: Molecular classification of human cancers using a 92-gene real-time quantitative polymerase chain reaction assay. Arch Pathol Lab Med 130:465-473, 2006
24. Talantov D, Baden J, Jatkoe T, et al: A quantitative reverse transcriptase-polymerase chain reaction assay to identify metastatic carcinoma tissue of origin. J Mol Diagn 8:320-329, 2006
25. Simon R, Radmacher MD, Dobbin K, et al: Pitfalls in the use of DNA microarray data for diagnostic and prognostic classification. J Natl Cancer Inst 95:14-18, 2003
26. Simon R: Roadmap for developing and validating therapeutically relevant genomic classifiers. J Clin Oncol 23:7332-7341, 2005
27. Rosenfeld N, Aharonov R, Meiri E, et al: MicroRNAs accurately identify cancer tissue origin. Nat Biotechnol 26:462-469, 2008
28. Estimated new cancer cases and deaths by sex for all sites, US, 2007. http://www.cancer.org/docroot/MED/content/downloads/MED_1_1x_CFF2007_Estimated_New_Cases_Deaths_by_Sex_US.asp

Acknowledgment

We thank Amelia Hensler and Aprel Delo for tissue procurement and preparation, Jill Hagenkord and Melissa Price for retrieval of pathology information, Rebecca Deeter for statistical analyses, Raji Pillai, PhD, for the biologic plausibility analysis (supplemental methods), Shawn Becker for helpful advice and Jane Seck for editorial support.

 © 2009 by American Society of Clinical Oncology

Information downloaded from jco.ascopubs.org and provided by HAM-TMC LIBRARY on May 18, 2009 from 206.83.53.249.
Copyright © 2009 by the American Society of Clinical Oncology. All rights reserved.