

An Expression-Based Site of Origin Diagnostic Method Designed for Clinical Application to Cancer of Unknown Origin

Richard W. Tothill,¹ Adam Kowalczyk,¹ Danny Rischin,² Alex Bousioutas,¹ Izhak Haviv,¹ Ryan K. van Laar,¹ Paul M. Waring,^{3,4} John Zalberg,² Robyn Ward,⁵ Andrew V. Biankin,⁶ Robert L. Sutherland,⁶ Susan M. Henshall,⁶ Kwun Fong,⁷ Jonathan R. Pollack,⁸ David D.L. Bowtell,¹ and Andrew J. Holloway¹

¹Ian Potter Centre for Cancer Genomics and Predictive Medicine, Departments of ²Haematology and Medical Oncology, and ³Pathology, Peter MacCallum Cancer Centre, St. Andrew's Place, East Melbourne, Victoria, Australia; ⁴Diagnostics and Pathology, Genentech Inc., South San Francisco, California; ⁵Department of Medical Oncology, St. Vincent's Hospital, Darlinghurst and School of Medicine, University of New South Wales, Sydney, Australia; ⁶Cancer Research Program, Garvan Institute of Medical Research, St. Vincent's Hospital, Darlinghurst, Australia; ⁷Department of Thoracic Medicine, Prince Charles Hospital, Chermide, Australia; and ⁸Department of Pathology, Stanford University, Stanford, California

Abstract

Gene expression profiling offers a promising new technique for the diagnosis and prognosis of cancer. We have applied this technology to build a clinically robust site of origin classifier with the ultimate aim of applying it to determine the origin of cancer of unknown primary (CUP). A single cDNA microarray platform was used to profile 229 primary and metastatic tumors representing 14 tumor types and multiple histologic subtypes. This data set was subsequently used for training and validation of a support vector machine (SVM) classifier, demonstrating 89% accuracy using a 13-class model. Further, we show the translation of a five-class classifier to a quantitative PCR-based platform. Selecting 79 optimal gene markers, we generated a quantitative-PCR low-density array, allowing the assay of both fresh-frozen and formalin-fixed paraffin-embedded (FFPE) tissue. Data generated using both quantitative PCR and microarray were subsequently used to train and validate a cross-platform SVM model with high prediction accuracy. Finally, we applied our SVM classifiers to 13 cases of CUP. We show that the microarray SVM classifier was capable of making high confidence predictions in 11 of 13 cases. These predictions were supported by comprehensive review of the patients' clinical histories. (Cancer Res 2005; 65(10): 4031-40)

Introduction

Gene expression profiling holds great potential as a new approach to cancer diagnosis and prognosis. A potential application of this technology lies in the development of molecular methods for the diagnosis of cancer site of origin. Whereas several groups have shown that tumors can be classified with respect to tissue of origin by expression profiling using either microarrays (1–3), publicly available SAGE data sets

(4, 5), or cell lines (6), it remains to be determined whether these tools can be effectively applied as a clinical diagnostic test.

In 3% to 5% of new cancer cases, the site of origin of a tumor cannot be readily identified, or a diagnosis of origin is equivocal (7). This disease manifestation is known as a cancer of unknown primary (CUP). These tumors represent a clinically diverse group, typically presenting with moderately to poorly differentiated tumors, often adenocarcinoma, involving multiple organs including liver, bone, lung, lymph nodes, pleura, and brain (8, 9). Patients with CUP represent a disproportionate fraction of cancer deaths due to their poor median survival, often measured in months (9). In many cases, patients receive a series of sequential treatments before a response, if any, is obtained. A large proportion of cases remain undiagnosed (10), with the result that therapy cannot be matched to their specific disease.

Gene expression profiling, applied as a site of origin diagnostic, seems likely to be one of the first areas of oncology where a microarray test influencing patient management could be applied, but important questions have yet to be resolved regarding design of such an assay. Recently, preexisting gene expression data sets have been combined to create very large databases that can be used to validate novel classification algorithms (11, 12). These studies are predicated on the belief that the use of an ever increasing number of samples will improve classification, although the influence of training set size has not been systematically evaluated. Likewise, previous studies have investigated major histopathologic types of tumors (1, 2) but have not investigated the importance of the inclusion of subtypes, such as histologic variants of ovarian cancer or estrogen receptor-positive and -negative breast tumors, in training sets. Finally, although metastatic tumors from known primary carcinoma have been correctly identified (1, 2, 12), no study has examined the efficacy of such tests for predicting site of origin of CUP, benchmarked against clinical diagnostic variables.

An additional consideration when developing any expression-based diagnostic is the clinical utility of the platform in terms of its ease of use and application to formalin-fixed paraffin-embedded (FFPE) tissue. Although microarray technology has matured considerably, it still has relatively lengthy protocols, with multiple enzymatic steps taking up to 3 days to complete. A number of studies have shown that accurate classification of multiple cancer types can be made using a reduced number of genes (2, 3, 11, 12). Hundreds rather than thousands of genes are therefore likely to be sufficient, indicating that a classification can be achieved using cheaper, faster, and more robust platforms for quantifying gene expression such as quantitative PCR. Quantitative PCR has several

Note: D.D.L. Bowtell and A.J. Holloway contributed equally to this work. P.M. Waring is currently at Diagnostics and Pathology, Genentech Inc., South San Francisco, California.

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

Requests for reprints: David Bowtell, Ian Potter Centre for Cancer Genomics and Predictive Medicine, Peter MacCallum Cancer Centre, Locked Bag 1, A'Beckett Street, Melbourne, Victoria 8006, Australia. Phone: 61-3-96561356; Fax: 61-3-96561414; E-mail: d.bowtell@petermac.org.

©2005 American Association for Cancer Research.

advantages over microarray, including its ability to use FFPE tissue. Several studies have reported robust expression analysis from fixed material using quantitative-PCR (13–15). Using multiplex reactions or generating low-density quantitative-PCR arrays therefore offers an attractive alternative to microarray for validation of classifier performance using existing FFPE tissue and eventual clinical application in a conventional pathology laboratory.

We describe here the development of a highly accurate multiclass classifier designed for clinical application to CUP. A large and comprehensive data set of gene expression was obtained from microarray analysis of 229 tumor samples, representing 14 commonly recognized sites of origin in the differential diagnosis for CUP. In collating the data set, we have purposely addressed the issue relating to molecular heterogeneity of specific tumor classes by including multiple histologic subtypes. The importance of sample coverage, particularly across the more heterogeneous classes, is shown by observing the expression of gene markers across subtypes and by systematically removing specific subtypes from training. We further used the microarray data set to choose an optimized series of tumor markers to create a quantitative PCR-based low-density array and generated a classifier that achieved similarly high prediction accuracies with both fresh and FFPE tissue to that obtained by microarray. Finally, we show the utility of the multiclass classifier for identifying the site of origin for CUP representing several clinical scenarios.

Materials and Methods

Tumor samples. Tumor specimens were collected through the Peter MacCallum Cancer Centre, Melbourne; The Garvan Institute of Medical Research, Sydney; St. Vincent's Hospital, Sydney; and The Prince Charles Hospital, Brisbane. Patient consent and Institutional Review Board approval were obtained according to National Health and Medical Research Council guidelines. Central pathology review was done on all samples. Histopathologic details of the tumors used for training the cDNA microarray classifier are provided in Supplementary Table S1: cDNA samples. Details of samples used for validating the quantitative-PCR classifier are provided in Supplementary Table S3: Quantitative PCR samples.

Unknown primary samples. Patients with CUP were referred to the study by treating oncologists at the Peter MacCallum Cancer Centre, Melbourne or St. Vincent's Hospital, Sydney. Thirteen patients with disseminated metastases and no clinically detectable sign of a primary tumor following a minimum investigation of histopathology and computed tomography imaging was identified for the study. Details of clinical evaluation for these patients are provided in Supplementary Table S2: CUP samples.

RNA extraction from fresh frozen tissue. Total RNA from fresh frozen tumor samples was isolated by phenol-chloroform extraction (Trizol; Invitrogen, Carlsbad, CA) and column chromatography (RNeasy, Qiagen, Valencia, CA) as previously described (16). Purified RNA was analyzed by agarose gel electrophoresis to assess the integrity of 28S and 18S rRNA bands.

Microarray analysis. Total RNA (3 µg) was amplified and labeled using a modified Eberwine method (17) and hybridized to cDNA microarrays containing ~10,500 elements as described previously (16). Reference RNA consisted of a pool of RNA isolated from eleven human tumor cell lines (18). Further details are provided in Supplementary Information Part 1: Methods. All MIAME compliant microarray data are available at ArrayExpress at EBI (www.ebi.ac.uk/arrayexpress; accession number: e-mexp-113).

Feature selection and support vector machine. Feature selection was done by ranking genes by absolute value of their signal-to-noise ratio (19) statistically comparing the gene expression observed within a single class against all other classes (one versus all). The top m ranked genes were selected for each class and combined for subsequent use in building a support vector machine (SVM) model. For supervised classification of

cancers, we used linear SVMs (20). Individual models were trained for discrimination of each cancer class from all other classes (one versus all), therefore an n class classifier was comprised of n class models. In testing a sample, the class model with the highest score was deemed the correct prediction. A decision margin was also given to the class prediction, calculated from the difference between the highest and the second highest score as follows: the decision margin was termed *high* if the difference in absolute SVM score between the first and second predictions was >50 , *medium* if the difference was between 26–50, or *low* if the difference was ≤ 25 . For more detailed description of SVM classification methods, refer to Supplementary Information Part 2: Generating a multiclass predictor using cDNA microarray.

Quantitative PCR using low-density arrays. Genes that effectively classified samples as gastric, colorectal, ovarian, pancreas, and breast were identified using a signal-to-noise metric (19) to analyze microarray expression data from 173 tumor samples. Twelve to fifteen of the most frequently selected genes for each class were chosen and the corresponding validated primer/probe sets were incorporated into a low-density array (Assay on Demand, Applied Biosystems, Foster City, CA). Six endogenous controls were added to the assay set, which were also used independently for quality assurance of cDNA using SYBR Green chemistry (Applied Biosystems). Twelve genes were added that represent tumor types outside of the five site differential and represented the class *other* to enable the identification tumors that did not belong to the tumor types present in the test. Genes selected for the quantitative-PCR low-density array are summarized in Supplementary Table S4: Low-density array gene list. See also Supplementary Table S5: Primer design for SYBR green endogenous controls.

For FFPE samples, five 10-µm sections were used for RNA extraction using a modification to the protocol previously described (13). Further information about RNA isolation, the generation of cDNA from fresh frozen and formalin-fixed material, gene selection for the quantitative-PCR low-density array, and subsequent quantitative-PCR analysis is described in Supplementary Information Part 1: Methods.

Quantitative-PCR data analysis. Normalization of quantitative-PCR assays was conducted using an average Ct value for all endogenous controls. Samples were then converted to a fold change ratio described using standard ΔCt formula:

$$X = 2^{\Delta\text{Ct}}, \text{ where } \Delta\text{Ct} = (\text{Ct target} - \text{Ct average endogenous controls})$$

Clustering of quantitative-PCR data was conducted by Pearson correlation using the program Cluster (21) and visualized using the program Mapletree (<http://mapletree.sourceforge.net/>). Analysis of quantitative-PCR data and the generation of a cross-platform model are described in detail in Supplementary Information Part 3: Translating a multiclass classifier to quantitative PCR.

Results

A comprehensive multiclass data set. To create a training set against which to compare cases of CUP, we profiled 229 tumors from 14 sites of origin on 10.5K spotted cDNA microarrays. Because ~90% of CUP tumors are thought to originate from epithelial cell types (10), we ensured a good representation of the major carcinoma types, defined by their anatomic tissue or organ of origin. Given also the histologic diversity of some carcinomas, we systematically represented histologic and molecular subtypes for some cancers [e.g., breast, ovarian, lung, gastric, and squamous cell carcinoma (SCC)]. Nonepithelial types, such as melanoma (22), seminoma (23), and mesothelioma (24), which can present with a cellular morphology and architecture indistinguishable from some poorly differentiated carcinomas, were also included. A summary of all tumors in the training set is detailed in Supplementary Table S6: Tumor classes.

Developing a site of origin classifier by machine learning.

We used an SVM algorithm to create a 13-class predictor based on anatomic site of origin (combining head and neck and skin SCC types; *SCCother*). In practice, the performance of an SVM depends critically on the subset of features (genes) selected for modeling and tuning relative to the regularization constant C (see Supplementary Information Part 2). The classification method was very robust, with marginal dependence on the tuning variables across the range of values. The leave-one-out cross-validation (LOOCV) accuracy was between 94% and 96.5% using at least 20 genes per class, with the best performance obtained using 50 genes per class (~600 nonredundant genes combined) and a C value of 10. Importantly, genes were reselected for every round of the LOOCV (25). The results of LOOCV, using optimal variables ($C = 10$, 50 genes/sample), are displayed for each class (Table 1), qualified by an associated decision margin (either high, medium, or low; see Materials and Methods). When the SVM prediction has a high or medium decision margin (combined into *strong*), there was a high likelihood of the prediction being correct: 203 of 207 cases (98.1%) called with strong decision margin were correctly classified (Table 1). In contrast, 4 of 18 (22%) low decision margin results were incorrect. Given that low decision margin predictions have a low but significant chance of being incorrect, we considered cases predicted with a low decision margin as unclassified. Considering only strong decision margin predictions, the adjusted accuracy for LOOCV was 89%. A similar high accuracy was achieved by splitting the entire data set into a two thirds training and a one third independent test set (Supplementary Information Part 2; see also Fig. 2B and C).

There were plausible explanations for the errors made by the SVM during LOOCV. For example, samples were confused based on close phenotypic similarities, such as the uterine endometrioid tumor misclassified as ovarian, due to similarity to an endometrioid ovarian gene expression signature, and among SCC-type tumors (Table 1). For several tumors, misclassification seems a result of representation by a single case example (chromophobe renal cell tumor; mixed cell type gastric tumor) or atypical morphology

(spindle cell-like melanoma sample; pancreatic tumor with an intestinal-like appearance). The pancreatic tumor that was predicted with high decision margin as colorectal most likely represents a recently described subtype of pancreatic adenocarcinoma that shares high molecular similarity to colorectal tumors (26).

Assessing the importance of training set coverage. Heterogeneous cancer subtypes, from organs such as the lung and ovary, may represent metaplastic or dedifferentiated variants that do not resemble a normal tissue counterpart or related subtypes morphologically or at the molecular level. Consistent with this, we found that despite the ability to measure expression across thousands of genes and a supervised approach to feature selection (top 20 ranked signal-to-noise ratio per class), there is a paucity of universally expressed site of origin markers for some cancer types (Fig. 1A, and see Supplementary Table S7 for gene list). For a higher-resolution view of gene expression, we selected several known markers from our refined list with previously validated tissue specificity (6, 27–37) and plotted the relative fold change for colorectal, breast, ovarian, and lung tumors. Markers can be identified which seem to be strongly and relatively uniformly expressed across the range of colorectal and breast tumors (Fig. 1B; *VILI* and *NOX1* for colorectal cancer, *PIP* and *GATA3* for breast). In contrast, the known histologic heterogeneity of ovarian and lung tumors correlates well with heterogeneity of expression of selected genes, including *RBPI*, *WT1*, and *STAR* (ovarian) and *TITF1*, *ADH7*, and *SFTPB* (lung). Our findings indicate the importance of inclusion of specific subtypes in training the classifier and the use of multiple markers for classification, particularly for heterogeneous tumor types such as lung and ovarian cancer.

Given our observation confirming that histologic subtypes can be diverse with respect to expression of tumor markers, we systematically quantitated the contribution of subtype representation in training the SVM and classifier performance (leave-subtype-out analysis). For this purpose, individual subtypes were sequentially removed from the training set, using the compromised classifier to predict the origin of these left out samples (Fig. 2A). Inclusion of certain subtypes in the training set was critical

Table 1. Results per tumor class for LOOCV on training set ($n = 229$) using the best variable SVM model

Class (n)	Correct (n) [H, M, L]	Errors (n) [H, M, L]	Histology of misclassified sample
Breast (34)	33 [30, 2, 1]	1 [0, 0, 1] Melanoma	Breast ductal adenocarcinoma, estrogen receptor negative
Colorectal (23)	23 [20, 1, 2]		
Gastric (15)	14 [12, 2, 0]	1 [0, 0, 1] Colorectal	Gastric mixed cell type
Melanoma (11)	10 [7, 3, 0]	1 [0, 0, 1] Lung	Melanoma spindle cell-like
Mesothelioma (8)	8 [7, 1, 0]		
Ovarian (50)	50 [38, 8, 4]		
Pancreas (9)	8 [4, 2, 2]	1 [1, 0, 0] Colorectal	Pancreatic adenocarcinoma, atypical intestinal-like morphology
Prostate (8)	8 [7, 0, 1]		
Renal (13)	12 [10, 1, 1]	1 [0, 0, 1] Breast	Renal cell carcinoma, chromophobe subtype
Testicular (3)	3 [3, 0, 0]		
SCCother (14)	13 [5, 4, 4]	1 [1, 0, 0] Lung	SCC of tongue spindle cell-like
Uterine (9)	8 [8, 0, 0]	1 [0, 1, 0] Ovarian	Uterine endometrioid subtype
Lung (32)	31 [24, 4, 3]	1 [1, 0, 0] SCCother	Lung SCC (moderately differentiated)

NOTE: Total number of samples predicted correctly or incorrectly is shown in bold whereas the distribution of predictions within the decision margin levels (high, medium, or low) is shown in brackets. The decision margin is determined by difference in absolute SVM score between the first and second highest predictions. A difference of greater than 50 defines high; 26–50, medium; and 0–25, low.

for certain tumor classes. For example, when mucinous ovarian tumors are left from training, test mucinous ovarian samples were commonly misclassified as tumors from the gastrointestinal tract. Similarly, classification of estrogen receptor–negative breast tumors was compromised when they are excluded from training, but interestingly estrogen receptor–positive breast tumors could be correctly predicted if the classifier was trained with estrogen receptor–negative samples. In clinical practice, these dilemmas

represent common and important diagnostic problems (38–40), underscoring the importance of including these subtypes in the training set. We observed that not only did the total number of incorrect classifications increase when specific subtypes were omitted, but the proportion of low confidence predictions (considered unclassified) also increased from 9.6% to 25% (Fig. 2B). Although an increased number of tumors could not be classified in the absence of particular subtypes, the accuracy of strong decision margins remained high (Fig. 2C).

A further consideration in building the classifier is the possibility of high decision margin predictions (false positive) when testing on tumor types not represented in training. To investigate the behavior of our classifier in this context, we systematically removed entire tumor types from training and then tested on these samples. This is similar to the leave-subtype-out analysis and we refer to it as a leave-type-out analysis. As expected, the number of high decision margin predictions decreased from 77% to 19.1%. Although the predictor decision margin is low for most samples when their representative class is not included in training, it seems that there may be exceptions if they share a strong molecular likeness to tumors of another class (Table 2). For example, when prostate tumors are absent from the training set, prostate test samples are exclusively predicted as breast (Table 2), revealing an association of these tumors based on the commonality of hormonal regulation (41, 42). Similarly, all mesothelioma samples were predicted as ovarian, perhaps reflecting their shared mesothelial cell lineage and expression of known markers such as mesothelin, retinol binding protein 1, and Wilms tumor 1 (43–45). Although such tumor types may rarely be involved in the same diagnostic differential, it further shows the importance of including a wide spectrum of tumor types in training the classifier for attaining high specificity.

Translation from cDNA microarray to quantitative PCR. Supervised learning using microarray data has shown that high-accuracy predictions can be achieved using a refined selection of gene markers, suggesting that such an expression-based test could be translated to a lower density platform such as quantitative PCR. To show this concept, we used the microarray data set to select gene markers for a refined differential of five sites: ovarian, breast, pancreas, colorectal, and gastric. Additionally, we selected a number of gene markers associated with tumors outside these five sites, which define a sixth class, *others*. The purpose of the *others* class was to assess if the specificity of a classifier could be increased by reducing the occurrence of strong decision margin predictions for unknown samples that do not originate from the five site differential. A total of 79 site-specific markers, in addition

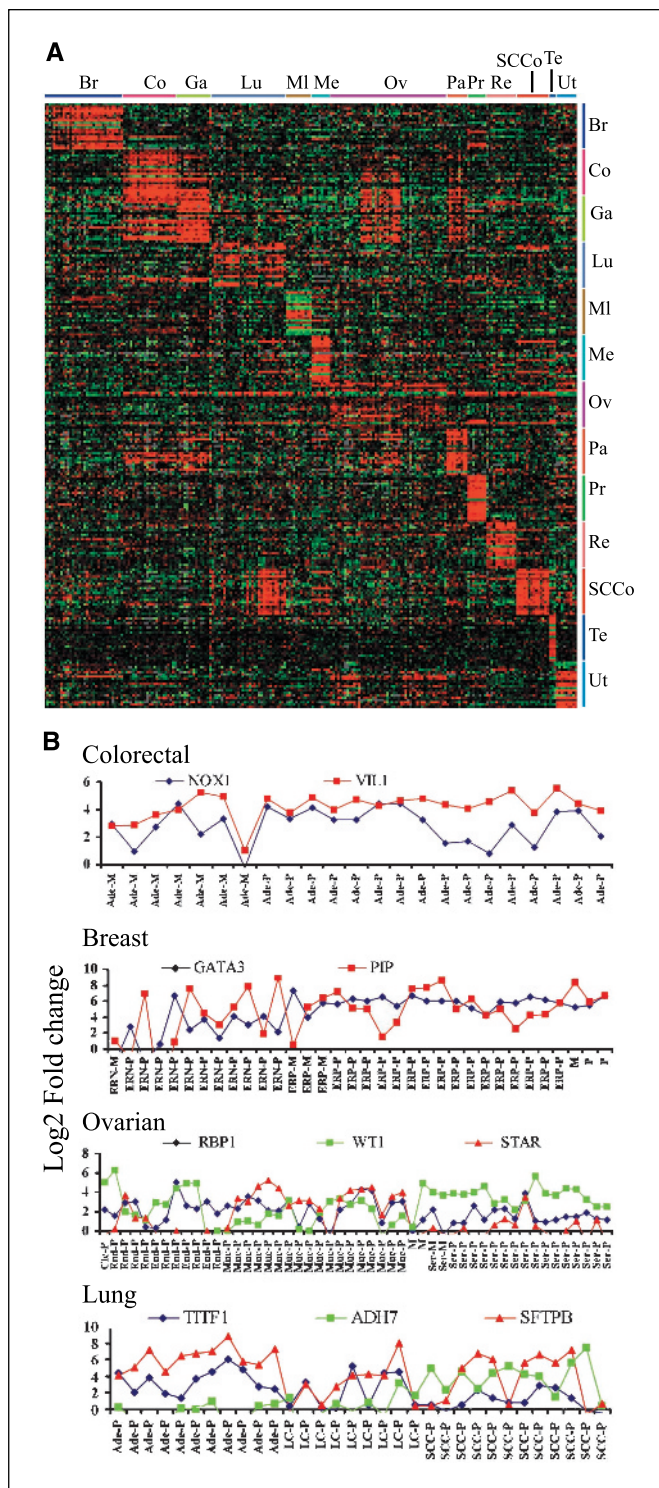


Figure 1. The expression of selected cancer type-specific genes across 13 classes and related subtypes. A, a heat map representation of median normalized array data for the top 20 genes per class (selected using the signal-to-noise metric using all 229 tumors) aligned respective to cancer class. Br, breast; Co, colorectal; Ga, gastric; Lu, lung; MI, melanoma; Me, mesothelioma; Ov, ovarian; Pa, pancreas; Pr, prostate; Re, renal; SCCo, SCC of skin or head and neck; Te, testicular; Ut, uterine. B, the median normalized log transformed expression data of several gene markers selected for comparison across histologic and molecular subtypes of four cancer classes. Expression values less than 0 were removed from display. Ade, adenocarcinoma; ERPI/ERN, estrogen receptor positive/negative; LC, large cell; P, primary; M, metastasis. VIL1, villin 1; NOX1, NADPH oxidase 1; PIP, prolactin-induced protein; GATA3, GATA binding protein 3; RBP1, retinol binding protein 1, cellular; WT1, Wilms tumor 1; STAR, steroidogenic acute regulator; TTF1, thyroid transcription factor; ADH7, alcohol dehydrogenase 7 (class IV), μ or σ polypeptide; SFTBP1, surfactant, pulmonary-associated protein B.

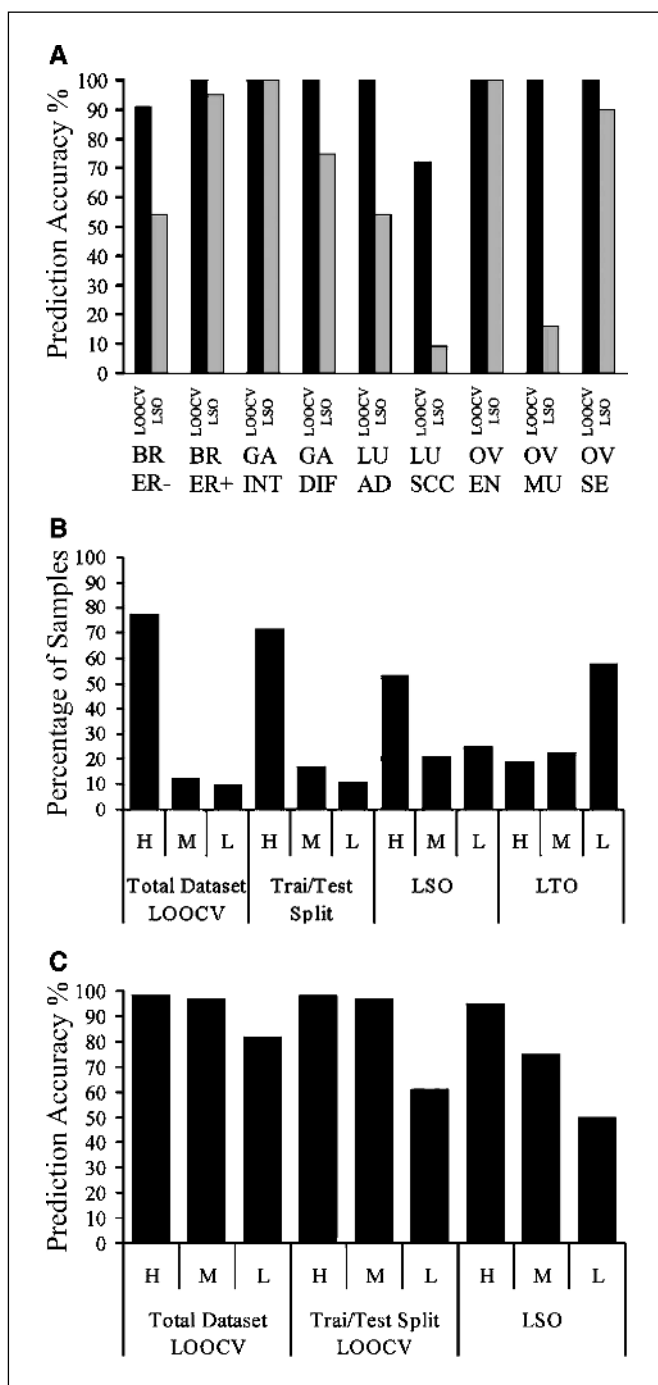


Figure 2. The effect of training set coverage on SVM prediction accuracies. **A**, prediction accuracies from leave-subtype-out analysis compared with the results for the same samples predicted during LOOCV. *BR ER-*, breast lobular or ductal adenocarcinoma, estrogen receptor negative; *BR ER+*, breast lobular or ductal adenocarcinoma, estrogen receptor positive. *GA INT*, gastric adenocarcinoma intestinal subtype; *GA DIF*, gastric adenocarcinoma, diffuse subtype; *LU AD*, lung adenocarcinoma; *LU SCC*, lung squamous cell carcinoma; *OV EN*, ovarian endometrioid carcinoma; *OV MU*, ovarian mucinous carcinoma; *OV SE*, ovarian serous papillary carcinoma. **B**, demonstrating the effect of data set size and complexity on distribution of predictions within three decision margin levels: high, medium, and low. *Total Data Set LOOCV*, results from a training set using all available samples ($n = 229$). *Trai/Test Split LOOCV*, results from training set only ($n = 167$). *LSO*, the accumulated results from iteratively leaving subtypes from training ($n = 105$). *LTO*, accumulated results from iteratively leaving entire tumor types from training ($n = 229$). **C**, accuracy of predictions within confidence levels high (H), medium (M), and low (L).

to six endogenous controls, were selected and were subsequently used for the design of a quantitative-PCR low-density array (Applied Biosystems).

First, to compare data generated from cDNA microarray and quantitative PCR, 42 fresh frozen tumor samples were profiled on both platforms. A Pearson correlation was calculated for each gene between platforms showing a high concordance between the data sets (median $r = 0.83$). Three genes from 79 were considered discordant ($r < 0.4$), despite sequence analysis confirming the identity of the clone used in generating the microarray. Differences between microarray and PCR data have been reported previously (46), and may be due to cross-hybridization of targets to microarray probes or expression of splice-specific isoforms not recognized in the PCR assay. These genes were removed from further analyses.

Classification using formalin-fixed tissue. Given the value of accessing fixed tissue samples, we next analyzed RNA obtained from FFPE cancer samples collected in a routine diagnostic pathology service. Twenty-five FFPE samples, collected 1 to 4 years ago and spanning all five classes of interest, were chosen. Of these, seven FFPE samples matched a fresh frozen sample already analyzed, whereas eight samples represent matched primary and metastatic tumors from the same patient. Hierarchical clustering of normalized quantitative-PCR expression profiles from fresh frozen and FFPE tissues showed that samples extracted from both fresh frozen and formalin-fixed material can be clustered accurately corresponding to the tissue of origin, with few exceptions (Fig. 3A).

Microarray and quantitative-PCR cross-platform predictor. Whereas it is possible to train and test a predictor solely using quantitative-PCR data, the relatively small number of samples analyzed by quantitative-PCR made construction of adequate independent training and test sets problematic. To circumvent this problem, we exploited cDNA microarray data to train the SVM, and then tested the predictor on an independent quantitative-PCR data set. Importantly, the microarray samples used to train the predictor were independent of those analyzed by quantitative-PCR.

Predictors were developed using either five or six class models. The five-class model is based on the five sites of origin: gastric, colorectal, pancreas, ovarian, and breast. The quantitative-PCR data set used for testing this model, composed of 55 samples, represented both fresh frozen and FFPE samples. The six-class model implements an additional class, others, that represents the combined signatures from cancers outside the original five-site differential. The test set used in this case was identical to that used for the five class model, except for the addition of four samples (melanoma, renal, prostate, and lung adenocarcinoma) representing the others class ($n = 59$). To develop cross-platform models, we normalized (rescaled) the data sets to cope with inherent differences in the data types. We developed a method of rank levels and compared this with the more rudimentary method of median normalization. Ranking consistently outperformed median normalization and enabled high-accuracy predictions of greater than 96% (Fig. 3B and Supplementary Information Part 3). The results from independently testing fresh and FFPE samples using five- and six-class models is shown in Fig. 3C, showing high accuracy is achieved for all specimen types.

Classification of cancers of unknown primary. To assess the clinical utility of our classifiers, we collected 13 cases of metastatic disease for which the primary tumor could not be unequivocally diagnosed at the time of presentation (Table 3, see also Supplementary Table S2 for detailed diagnostic workup and

Table 2. Confusion matrix for leave-type-out analysis based on cancer site of origin

Samples (<i>n</i>)	Leave-type-out SVM predictions										
	BR	CO	GA	ML	MS OV	PA	PR RE	TE	SCCo	UT	LU
BR (33)					13 [0, 2, 11]	8 [0, 0, 8]		3 [0, 0, 3]		9 [0, 0, 9]	
CO (23)			18 [4, 8, 6]			1 [0, 0, 1]					4 [0, 0, 4]
GA (14)		10 [1, 4, 5]			4 [0, 0, 4]						
ML (10)	2 [0, 0, 2]				1 [0, 0, 1]		1 [0, 0, 1]		5 [0, 0, 5]		1 [0, 0, 1]
MS (8)					8 [3, 3, 2]						
OV (50)	11 [0, 0, 11]	3 [0, 1, 2]	9 [2, 2, 5]	1 [0, 0, 1]		3 [0, 0, 3]	2 [0, 0, 2]			16 [1, 6, 9]	5 [0, 2, 3]
PA (7)		1 [0, 0, 1]	2 [2, 0, 0]		4 [0, 1, 3]						
PR (8)	8 [2, 5, 1]										
RE (12)					11 [0, 5, 6]						1 [0, 0, 1]
TE (3)											3 [0, 2, 1]
SCCo (13)											13 [9, 2, 2]
UT (8)					8 [8, 0, 0]						
LU (31)	5 [0, 4, 1]	3 [0, 0, 3]			3 [0, 0, 3]		1 [0, 0, 1]	1 [0, 0, 1]	18 [10, 3, 5]		

NOTE: Sample groups representing a cancer type were removed from SVM training and then tested using the resulting compromised predictor. The class left from training is represented by row and the test prediction by column. The total number of samples predicted is shown in bold, whereas confidence distribution is shown in brackets (high, medium, and low).

Abbreviations: BR, breast; CO, colorectal; GA, gastric; ML, melanoma; MS, mesothelioma; OV, ovarian; PA, pancreas; PR, prostate; RE, renal; TE, testicular; SCCo, SCCother; UT, uterine; LU, lung.

histology images). All cases fell into one of three categories: (a) metastatic disease without any prior history of cancer ($n = 7$); (b) metastatic disease with a prior history of cancer ($n = 5$) and; (c) presentation with two concurrent primary tumors ($n = 1$).

We tested the cDNA microarray data generated from CUP samples with the SVM classifier trained on all 229 samples in the known tumor data set. Based on our best variable model, realized from prior cross-validation, 11 of 13 patients were predicted with strong decision margin (i.e., decision margin > 25; Table 3). The two cases that were not predicted with strong confidence (P02864 and P02971) received equal SVM scores for SCCother and lung classes (i.e., decision margin = 0), suggesting the classifier has difficulty in discriminating between SCC arising from different primary sites.

We compared our SVM predictions with the most likely primary site, based on evaluation of the case by a medical oncologist after review of all subsequent investigations and additional pathologic evaluation. In all cases where a consistent and strong decision margin prediction was made (11 of 13, 85%), SVM classification was consistent with either the highest possibility of the tissue of origin or among the short list of likely sites. We also noted that where available, outcome information and further clinical evidence were supportive of the prediction. For example, case P1328, predicted to have breast cancer, presented initially with a differential diagnosis of ovary, gastric, or breast cancer and then later presented with metastatic deposits in axillary lymph nodes and supraclavicular fossa, consistent with a breast primary. Similarly, patient P01698, who presented with widespread metastatic adenocarcinoma, was strongly predicted as ovarian cancer, but was thought to have a gastrointestinal type tumor based on histopathologic review conducted by several pathologists. This, however, conflicted with other clinical evidence such as a raised plasma CA125 concentration and no identifiable gastrointestinal type primary from endoscopic investigation radiological imaging, including positron emission tomography scan. Eventually it was decided to treat the

patient with a regimen of taxol and carboplatin, as a broader-acting combination treatment was considered undesirable due to renal impairment. The patient had a good response to chemotherapy, consistent with the high decision margin prediction of ovarian cancer. This case represents a common diagnostic problem encountered with mucinous type tumors presenting in the ovary (47), and exemplifies the utility of the classifier for resolving such cases.

For a number of cases it can also be argued that the test would have significantly reduced the time taken to begin an appropriate chemotherapy regimen and may have given a survival benefit. For example, case P00563 had a 4-year history of ovarian cancer, and later presented with undifferentiated bone metastases. She was treated for recurrent ovarian cancer despite what would be an atypical presentation of a recurrence for this cancer type. Pathology and other diagnostic imaging did not suggest a likely alternative primary site until the patient presented 2 years later with an identifiable breast mass. Our microarray-based classifier, tested on the left neck mass, predicted breast cancer as the primary tumor, a diagnosis that would have resulted in an altered treatment plan 2 years earlier than otherwise applied. In such cases, where the patient has a previous history of cancer, it may be possible to access FFPE material from an earlier episode. Using our quantitative-PCR low-density array, we assayed FFPE tissue RNA extracts from both the 1994 ovarian tumor and the 2000 unidentified neck mass. Prediction using a cross-platform microarray/quantitative-PCR model (five- or six-class) resulted in strong decision margin predictions of ovarian and breast, respectively (see Supplementary Information Part 3).

Discussion

Several studies have shown that patterns of gene expression remain consistent with tissue of origin, both in cell lines (48) and tumor samples (1, 2, 49). Gene expression profiling may therefore

enable an accurate identification of the site of origin of a tumor, implying that such a technology could be developed into a clinically useful diagnostic test. We have shown for the first time translation of a genomics-based classifier to a more clinically amenable quantitative-PCR platform and obtained robust prediction of samples from both fresh and FFPE tissue. We tested our classifier on a spectrum of diagnostically challenging tumors. Our microarray-based predictor is capable of making confident predictions for 11 of 13 tumors, which in several cases can be strongly supported by their detailed clinical histories.

Utilizing our heuristic confidence measure, 89% of training samples were correct and predicted with a strong decision margin. Closer examination of the histology associated with misclassified samples revealed some systematic errors were made by the classifier. First, it seems that misclassification is possible for well-differentiated tumors sharing a common histologic appearance with other classes. The most obvious of these relates to the lung SCC and SCOther type tumors. Metastatic SCC of unknown origin represents a small but significant fraction of all CUP cases (7), with the differential often including head and neck cancer or a primary lung tumor (50). Other molecular based studies have used comparative genetic alterations (i.e., allelic loss) to assist in

matching the clonal origins of metachronous or synchronous SCC tumors (51); however, discordant results can occur due to genomic loss during tumor evolution (52), thus confounding the interpretation of results. Owing to the close molecular similarity of SCC type tumors, special attention is required to develop a more accurate gene expression-based classifier.

Errors were made for cases that represent a single sample of a particular subtype, consistent with the leave-subtype-out analysis, which showed that failure to represent specific histologic or molecular subtypes resulted in misclassification. This can be attributed to a paucity of gene markers that are truly universal across all subtypes, suggesting that some tumors do not retain expression of site of origin markers, but rather adopt an expression pattern underlying their new ectopic and differentiated form. This underscores the importance of representing class heterogeneity in respect to existing knowledge of cancer histopathology. As tumors may present in various states of differentiation, including mixed cell phenotypes, such examples must also be represented in training the classifier to cover the complete molecular heterogeneity that can arise from a specific site of origin. Unlike other studies (1), our classifier had no systematic difficulty in identifying poorly differentiated tumors. This may be attributed to the broad

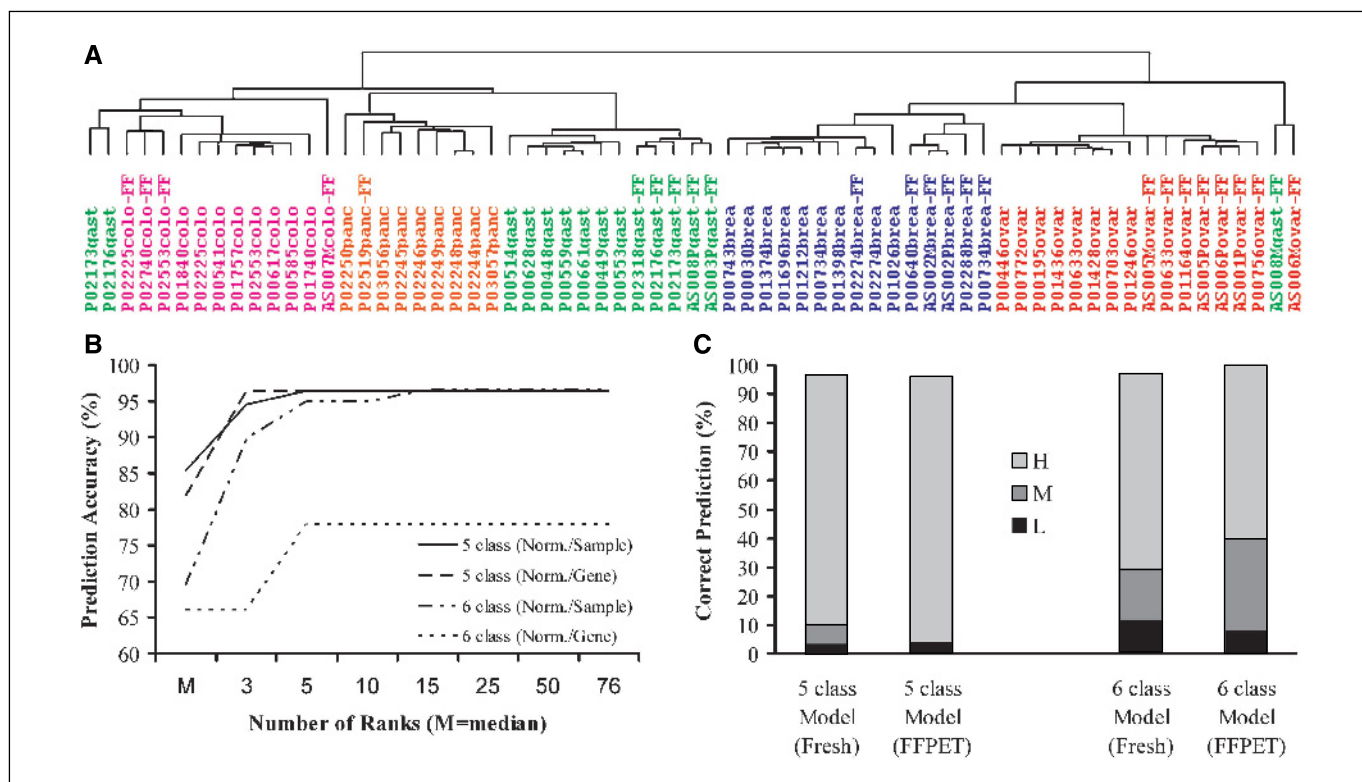


Figure 3. The translation of an expression-based classifier from microarray to quantitative PCR. **A**, hierarchical clustering of data generated from a quantitative-PCR low-density array platform. Median normalized quantitative-PCR data representing 67 tissue samples were subjected to average linkage clustering according to Pearson correlation using the program Cluster. Cluster data were visualized using the program Mapletree. Sample names refer to the patient unique identifier and primary site of origin. Samples isolated from formalin-fixed tissue are labeled FF. For unique identifiers with the prefix AS, primary and metastatic samples of the same cancer episode are labeled as P and M, respectively. **B**, comparison of methods for data transformation. Two methods for data transformation (median or rank normalization) were used independently in two directions (per gene across samples or per sample across genes) making the cDNA microarray and quantitative-PCR data sets compatible for building and testing SVM models. The method of ranking was tested using a titration of *n* rank levels (3-76) where all genes are firstly ranked from highest to lowest fold change, relative to directionality of normalization, and then binned into *n* ranks. The original measure of expression for a particular gene (i.e., a ratio representing fold difference to a reference for microarray or endogenous controls for quantitative PCR) is substituted with the value associated with its positional rank (see Supplementary Information Part 3). **C**, comparison of five- and six-class SVM models, using a per sample 15 rank normalization strategy, testing on samples processed from fresh frozen and FFPE tissue independently. The prediction accuracy is presented relative to the proportion of samples correctly predicted with a high (H), medium (M), or low (L) decision margin. The overall prediction accuracy is equivalent between fresh frozen and FFPE tissue. The five-class model outperforms the six-class model relative to the number of predictions with high decision margins.

Table 3. Summaries of clinical history and array predictions for unknown primary samples

Disease presentation and histology	Differential at initial presentation	Array prediction and outcome
P00459: 40-y-old male nonsmoker, no previous history. Supraclavicular and mediastinal lymphadenopathy, lymphangitis of lung, right upper lobe mass, and liver metastases. Poorly differentiated adenocarcinoma.	Clinical picture most consistent with lung but uncertain in a young nonsmoker.	Lung (70). Minor response to platinum/gemcitabine, stable disease for 3 mo on gefitinib and progressive disease with docetaxel.
P01328: 52-y-old female, no previous history. Extensive abdominal tumor. Adenocarcinoma.	Ovary, gastric, and breast	Breast (100). Left supraclavicular fossa and axillary nodes developed within 2 mo of chemotherapy.
P01405: 66-y-old male nonsmoker, no previous history. Paraaortic lymphadenopathy and bone metastases. Clear cell epithelioid tumor.	Pathology review favored sarcomatoid renal cell cancer; but renal CT and MRI normal.	Renal (88).
P01698: 37-y-old female, no previous history. Pelvic mass, ascites, and left pleural effusion. Moderately differentiated adenocarcinoma with occasional signet ring features.	Pathologist thought that morphology strongly suggested nonovarian origin (e.g., gastric, colorectal, pancreas, or lung). Clinical picture consistent with ovarian cancer.	Ovarian (92). Treated with taxol/carboplatin for presumed ovarian primary. Good clinical response with normalization of CA125
P01946: 49-y-old female smoker, no previous history. Liver, bone, adrenal, and mediastinal disease. Atypical infiltrating epithelial cells forming glandlike structures.	Lung, colorectal.	Lung (60)
P02971: 82-y-old female ex-smoker, no previous history. Bilateral cervical and mediastinal lymphadenopathy and bilateral lung metastases. Poorly differentiated adenocarcinoma	Pathologist suggested possible primaries included lung, endometrium, breast, and gastrointestinal. Clinical pattern of disease suggestive of lung or breast and colon needed to be excluded based on PET finding.	SCC \equiv Lung (0). Patient did not receive any active treatment.
P02989: 65-y-old male no previous history. Inguinal and mediastinal lymphadenopathy, bone, and lung metastases. Undifferentiated carcinoma.	Renal favored with differential of adrenal or hepatocellular carcinoma. However, no renal mass identified and histology atypical.	Renal (62). Treated as unknown primary with carboplatin and gemcitabine. Some improvement in symptoms with chemotherapy with best response of stable disease.
P00563: 67-y-old female diagnosed in 1994 with stage IIC poorly differentiated endometrioid ovarian cancer. Treated with TAH/BSO and chemotherapy. Presented in 1998 multiple sclerotic bone metastases. Undifferentiated carcinoma.	Pathologist in 1998 favored recurrent ovary or gastrointestinal primary. Clinical picture raised question of breast.	Breast (100). Patient deceased without formal identification of a breast primary although the clinical picture strongly supported breast as the primary site.
P00780: 81-y-old female smoker with a past history of unresectable papillary thyroid cancer diagnosed in 2000. Presented in 2001 with large mass of thyroid invading trachea. Poorly differentiated adenocarcinoma.	Lung or thyroid.	Lung (82). Good response to radiotherapy. In 2003 presented with left lower lobe collapse. Found to have tumor involving carina and left lower lobe bronchus.
P01169: 31-y-old female with history of stage I high-grade borderline mucinous ovarian tumor 6 y previously. Sclerotic metastases in pelvis and left femur. Adenocarcinoma.	Pathologist favored ovary but could not exclude breast, lung, or gastrointestinal primary.	Breast (48). Treated as unknown primary with ECF regimen and subsequently docetaxel without response

(Continued on the following page)

Table 3. Summaries of clinical history and array predictions for unknown primary samples (Cont'd)

Disease presentation and histology	Differential at initial presentation	Array prediction and outcome
P01382: 74-y-old female smoker with previous history of renal cell tumor. Presented with bone metastases right iliac crest. Poorly differentiated adenocarcinoma.	Renal or lung.	Lung (71).
P02864: 53-y-old male, ex smoker with past history of skin lesions removed from back, arm, and head. Presented with solitary axillary mass.	Skin, renal, and hepatocellular.	SCCo \equiv Lung (0). Presumed to be skin primary, received postoperative radiotherapy.
P01245: 60-y-old female presented with postmenopausal bleeding and underwent a hysterectomy. Found to have stage IC endometrial adenocarcinoma, colorectal tumor, and liver metastases.	Histology for liver metastasis favored colon but could not exclude endometrial origin.	Colorectal (100). Treated with oxaliplatin and 5-fluorouracil chemotherapy with partial response of liver metastases

NOTE: The decision margin between first and second SVM predictions is shown in parentheses. Abbreviations: CA125, tumor-associated antigen CA125; CT, computed tomography imaging; DVT, deep vein thrombosis; IVC, inferior vena cava; PET, positron emission tomography imaging; TAH/BSO, total abdominal hysterectomy/bilateral salpingo-oophorectomy.

selection of subtypes we have used, some of which by definition are poorly differentiated (i.e., large cell lung tumors). Although several studies have compiled multiple data sets in an effort to increase sample coverage (11, 12), to our knowledge no other study has attempted to compile a multiclass data set specifically addressing the issues of histologic diversity.

The ability to accurately classify tumors using a refined number of gene markers suggests that translation of an expression-based classifier from microarray to quantitative PCR is possible. As a proof of principle, we have focused on a common differential diagnosis for CUP in women (7) covering the five sites of ovarian, gastric, colorectal, pancreas, and breast. A set of 79 site-specific markers translated to quantitative PCR allowed measurement of gene expression from either fresh or FFPE tissue. The robust quantification of mRNA from formalin-fixed tissue is consistent with several previous studies (13–15) but this is the first time to our knowledge where an accurate classifier has been generated using machine learning. Furthermore, the use of ranking to generate a classifier from microarray data, another novel feature of this work, negated the requirement to construct an entirely new training set using quantitative-PCR.

Our classifier was applied to a cohort of metastatic tumors in which the primary site could not be unequivocally identified at initial presentation, despite extensive clinical investigation. Although not all tumors in our series fit the classic definition of CUP, they represent a spectrum of real clinical scenarios where there was difficulty in diagnosing the origin of the tumor and determining clinical management. Predictions with strong decision margins could be made for all tumors, excluding the two cases of metastatic SCC. For several cases, compelling evidence became apparent during the course of the disease, which further validated the classifier predictions, albeit not until after the patient had endured extensive investigative procedures. It is difficult to obtain a definitive accuracy score for the classifier when testing on CUP tumors. This relates to the nature of these samples and that the origin of the majority of such tumors remains truly unknown (10).

Despite the bleak situation for patients with advanced stage cancer, treatments are becoming increasingly specific with approaches varying significantly depending on the cellular origin of the cancer. With recent advances in chemotherapy, specific regimens have led to improvements in survival and quality of life even in cancers that have traditionally been regarded as relatively chemoresistant (e.g., non-small-cell lung, colorectal, and pancreatic tumors; refs. 53–55). It is likely that there will be cost savings from more directed clinical evaluation of patients, enabled by a molecular genomics test. The average cost for diagnostic evaluation of CUP patients at a major US cancer center was ~\$18,000 when a large series was considered (56) whereas a test similar to that described here is likely to cost under \$1,000. Whereas a PCR or microarray-based classifier would not be expected to obviate the need for clinical investigation, it could allow much more focused testing, resulting in reduced cost, patient morbidity, and improved outcome.

Acknowledgments

Received 10/12/2004; revised 2/14/2005; accepted 2/25/2005.

Grant support: Australian National Health and Medical Research Council, The Cancer Council of New South Wales, the R.T. Hall Trust, St. Vincent's Clinical Foundation, the Prostate Cancer Foundation of Australia, and the Royal Australasian College of Surgeons.

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.

We are grateful to many people for their assistance with this work. In particular, we would like to thank Bianca Locandro, Kate Belfrage, and Dileepa Diyagama from the Peter Mac Microarray Core Facility for assistance with printing and hybridizing microarrays; Lisa Devereux, Justine Biggs, and Sam Coates from the Peter Mac Tissue Bank for collecting tissue samples and patient histories; Connie Mascarenhas, Patricia Goncalves, Melanie Trivett, and Neil O'Callaghan from the Peter Mac Department of Pathology and Peter Russell from the Royal Prince Alfred Hospital, Sydney, for assistance with review of patient specimens; and the members of the Centre for Cancer Genomics and Predictive Medicine, in particular Anna Tinker, David Thomas, Alan Christiansen, Ken Mitchelhill, Grant Macarthur, Sian Fereday, and Nadia Trafficante for helpful discussion. We are also grateful to Pat Brown, Matt van der Rijn, and David Botstein at Stanford University for sharing unpublished results. We would especially like to thank the many patients who donated tumor samples for this project.

We would like to dedicate this work to the memory of Professor Neil Della, a compassionate physician, talented researcher and valued friend, whose personal battle with cancer of unknown primary was the inspiration for this study.

References

1. Ramaswamy S, Tamayo P, Rifkin R, et al. Multiclass cancer diagnosis using tumor gene expression signatures. *Proc Natl Acad Sci U S A* 2001;98:15149–54.
2. Su AL, Welsh JB, Sapinoso LM, et al. Molecular classification of human carcinomas by use of gene expression signatures. *Cancer Res* 2001;61:7388–93.
3. Giordano TJ, Shedden KA, Schwartz DR, et al. Organ-specific molecular classification of primary lung, colon, and ovarian adenocarcinomas using gene expression profiles. *Am J Pathol* 2001;159:1231–8.
4. Dennis JL, Vass JK, Wit EC, Keith WN, Oien KA. Identification from public data of molecular markers of adenocarcinoma characteristic of the site of origin. *Cancer Res* 2002;62:5999–6005.
5. Buckhaults P, Zhang Z, Chen YC, et al. Identifying tumor origin using a gene expression-based classification map. *Cancer Res* 2003;63:4144–9.
6. Nishizuka S, Chen ST, Gwadyr FG, et al. Diagnostic markers that distinguish colon and ovarian adenocarcinomas: identification by genomic, proteomic, and tissue array profiling. *Cancer Res* 2003;63:5243–50.
7. Briasoulis E, Pavlidis N. Cancer of unknown primary origin. *Oncologist* 1997;2:142–52.
8. Le Chevalier T, Cvitkovic E, Caille P, et al. Early metastatic cancer of unknown primary origin at presentation: a clinical study of 302 consecutive autopsied patients. *Arch Intern Med* 1988;148:2035–9.
9. Abbruzzese JL, Abbruzzese MC, Lenzi R, Hess KR, Raber MN. Analysis of a diagnostic strategy for patients with suspected tumors of unknown origin. *J Clin Oncol* 1995;13:2094–103.
10. Pavlidis N, Briasoulis E, Hainsworth J, Greco FA. Diagnostic and therapeutic management of cancer of an unknown primary. *Eur J Cancer* 2003;39:1990–2005.
11. Shedden KA, Taylor JM, Giordano TJ, et al. Accurate molecular classification of human cancers based on gene expression using a simple classifier with a pathological tree-based framework. *Am J Pathol* 2003;163:1985–95.
12. Bloom G, Yang IV, Boulware D, et al. Multi-platform, multi-site, microarray-based human tumor classification. *Am J Pathol* 2004;164:9–16.
13. Specht K, Richter T, Muller U, Walch A, Werner M, Hoffer H. Quantitative gene expression analysis in microdissected archival formalin-fixed and paraffin-embedded tumor tissue. *Am J Pathol* 2001;158:419–29.
14. Abrahamsen HN, Steiniche T, Nexø E, Hamilton-Dutoit SJ, Sorensen BS. Towards quantitative mRNA analysis in paraffin-embedded tissues using real-time reverse transcriptase-polymerase chain reaction: a methodological study on lymph nodes from melanoma patients. *J Mol Diagn* 2003;5:34–41.
15. Cronin M, Pho M, Dutta D, et al. Measurement of gene expression in archival paraffin-embedded tissues: development and performance of a 92-gene reverse transcriptase-polymerase chain reaction assay. *Am J Pathol* 2004;164:35–42.
16. Boussioutas A, Li H, Liu J, et al. Distinctive patterns of gene expression in premalignant gastric mucosa and gastric cancer. *Cancer Res* 2003;63:2569–77.
17. Van Gelder RN, von Zastrow ME, Yool A, Dement WC, Barchas JD, Eberwine JH. Amplified RNA synthesized from limited quantities of heterogeneous cDNA. *Proc Natl Acad Sci U S A* 1990;87:1663–7.
18. Bowtell DD, Sambrook JF. *DNA Microarrays: A Molecular Cloning Manual*. New York: Cold Spring Harbor Laboratory Press; 2002.
19. Golub TR, Slonim DK, Tamayo P, et al. Molecular classification of cancer: class discovery and class prediction by gene expression monitoring. *Science* 1999;286:531–7.
20. Vapnik V. *Statistical Learning Theory*. J. Wiley & Sons; 1998.
21. Eisen MB, Spellman PT, Brown PO, Botstein D. Cluster analysis and display of genome-wide expression patterns. *Proc Natl Acad Sci U S A* 1998;95:14863–8.
22. Bono A, Maurichi A, Moglia D, et al. Clinical and dermatoscopic diagnosis of early amelanotic melanoma. *Melanoma Res* 2001;11:491–4.
23. Wehrschütz M, Stoger H, Ploner F, et al. Seminoma metastases mimicking primary pancreatic cancer. *Onkologie* 2002;25:371–3.
24. Tang P, Vatsia SK, Teichberg S, Kahn E. Pulmonary adenocarcinoma simulating malignant mesothelioma. *Arch Pathol Lab Med* 2001;125:1598–600.
25. Ambroise C, McLachlan GJ. Selection bias in gene extraction on the basis of microarray gene-expression data. *Proc Natl Acad Sci U S A* 2002;99:6562–6.
26. Adsay NV, Merati K, Basturk O, et al. Pathologically and biologically distinct types of epithelium in intraductal papillary mucinous neoplasms: delineation of an “intestinal” pathway of carcinogenesis in the pancreas. *Am J Surg Pathol* 2004;28:839–48.
27. Geiszt M, Lekstrom K, Brenner S, et al. NAD(P)H oxidase 1, a product of differentiated colon epithelial cells, can partially replace glycoprotein 91phox in the regulated production of superoxide by phagocytes. *J Immunol* 2003;171:299–306.
28. Perou CM, Jeffrey SS, van de Rijn M, et al. Distinctive gene expression patterns in human mammary epithelial cells and breast cancers. *Proc Natl Acad Sci U S A* 1999;96:9212–7.
29. Brown RW, Campagna LB, Dunn JK, Cagle PT. Immunohistochemical identification of tumor markers in metastatic adenocarcinoma. A diagnostic adjunct in the determination of primary site. *Am J Clin Pathol* 1997;107:12–9.
30. Acs G, Pasha T, Zhang PJ. WT1 is differentially expressed in serous, endometrioid, clear cell, and mucinous carcinomas of the peritoneum, fallopian tube, ovary, and endometrium. *Int J Gynecol Pathol* 2004;23:110–8.
31. Hwang H, Quenneville L, Yaziji H, Gown AM. Wilms tumor gene product: sensitive and contextually specific marker of serous carcinomas of ovarian surface epithelial origin. *Appl Immunohistochem Mol Morphol* 2004;12:122–6.
32. Cvitkovic D, Williams SJ, Hamilton TC. Loss of cellular retinol-binding protein 1 gene expression in microdissected human ovarian cancer. *Clin Cancer Res* 2003;9:1013–20.
33. Ishikura H, Sasano H. Histopathologic and immunohistochemical study of steroidogenic cells in the stroma of ovarian tumors. *Int J Gynecol Pathol* 1998;17:261–5.
34. Zamecnik J, Kodet R. Value of thyroid transcription factor-1 and surfactant apoprotein A in the differential diagnosis of pulmonary carcinomas: a study of 109 cases. *Virchows Arch* 2002;440:353–61.
35. Chhieng DC, Cangiarrella JF, Zakowski MF, Goswami S, Cohen JM, Yee HT. Use of thyroid transcription factor 1, PE-10, and cytokeratins 7 and 20 in discriminating between primary lung carcinomas and metastatic lesions in fine-needle aspiration biopsy specimens. *Cancer* 2001;93:330–6.
36. Khoor A, Whitsett JA, Stahlman MT, Olson SJ, Cagle PT. Utility of surfactant protein B precursor and thyroid transcription factor 1 in differentiating adenocarcinoma of the lung from malignant mesothelioma. *Hum Pathol* 1999;30:695–700.
37. Wang T, Hopkins D, Schmidt C, et al. Identification of genes differentially over-expressed in lung squamous cell carcinoma using combination of cDNA subtraction and microarray analysis. *Oncogene* 2000;19:1519–28.
38. Seidman JD, Kurman RJ, Ronnett BM. Primary and metastatic mucinous adenocarcinomas in the ovaries: incidence in routine practice with a new approach to improve intraoperative diagnosis. *Am J Surg Pathol* 2003;27:985–93.
39. Lee KR, Young RH. The distinction between primary and metastatic mucinous carcinomas of the ovary: gross and histologic findings in 50 cases. *Am J Surg Pathol* 2003;27:281–92.
40. Varadhachary GR, Abbruzzese JL, Lenzi R. Diagnostic strategies for unknown primary cancer. *Cancer* 2004;100:1776–85.
41. Gustafsson JA. Estrogen receptor β —a new dimension in estrogen mechanism of action. *J Endocrinol* 1999;163:379–83.
42. Chang WY, Prins GS. Estrogen receptor- β : implications for the prostate gland. *Prostate* 1999;40:115–24.
43. Ordonez NG. Application of mesothelin immunostaining in tumor diagnosis. *Am J Surg Pathol* 2003;27:1418–28.
44. Hecht JL, Lee BH, Pinkus JL, Pinkus GS. The value of Wilms tumor susceptibility gene 1 in cytologic preparations as a marker for malignant mesothelioma. *Cancer* 2002;96:105–9.
45. Mohr S, Rihn B. Gene expression profiling in human mesothelioma cells using DNA microarray and high-density filter array technologies. *Bull Cancer* 2001;88:305–13.
46. Rajeevan MS, Vernon SD, Taysavang N, Unger ER. Validation of array-based gene expression profiles by real-time (kinetic) RT-PCR. *J Mol Diagn* 2001;3:26–31.
47. Gersell DJ. Metastatic ovarian tumors: still a challenge. *Am J Clin Pathol* 1992;97:745–7.
48. Ross DT, Scherf U, Eisen MB, et al. Systematic variation in gene expression patterns in human cancer cell lines. *Nat Genet* 2000;24:227–35.
49. Khan J, Wei JS, Ringner M, et al. Classification and diagnostic prediction of cancers using gene expression profiling and artificial neural networks. *Nat Med* 2001;7:673–9.
50. Califano J, Westra WH, Koch W, et al. Unknown primary head and neck squamous cell carcinoma: molecular identification of the site of origin. *J Natl Cancer Inst* 1999;91:599–604.
51. Leong PP, Reza B, Koch WM, et al. Distinguishing second primary tumors from lung metastases in patients with head and neck squamous cell carcinoma. *J Natl Cancer Inst* 1998;90:972–7.
52. Sasatomi E, Finkelstein SD, Woods JD, et al. Comparison of accumulated allele loss between primary tumor and lymph node metastasis in stage II non-small cell lung carcinoma: implications for the timing of lymph node metastasis and prognostic value. *Cancer Res* 2002;62:2681–9.
53. The non-small cell lung cancer group. Chemotherapy in non-small cell lung cancer: a meta-analysis using updated data on individual patients from 52 randomized clinical trials. *BMJ* 1995;311:899–909.
54. Burris HA III, Moore MJ, Andersen J, et al. Improvements in survival and clinical benefit with gemcitabine as first-line therapy for patients with advanced pancreas cancer: a randomized trial. *J Clin Oncol* 1997;15:2403–13.
55. Cunningham D, Pyrhonen S, James RD, et al. Randomised trial of irinotecan plus supportive care versus supportive care alone after fluorouracil failure for patients with metastatic colorectal cancer. *Lancet* 1998;352:1413–8.
56. Schapira 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.