

# Hybrid Model Integrating Immunohistochemistry and Expression Profiling for the Classification of Carcinomas of Unknown Primary Site

Barbara A. Centeno,<sup>\*†</sup> Gregory Bloom,<sup>‡</sup>  
Dung-Tsa Chen,<sup>§</sup> Zhihua Chen,<sup>‡</sup> Mike Gruidl,<sup>¶</sup>  
Aejaz Nasir,<sup>\*†</sup> and Timothy Y. Yeatman<sup>†</sup>

From the Departments of Anatomic Pathology,<sup>\*</sup> Biomedical Informatics,<sup>‡</sup> Biostatistics,<sup>§</sup> Experimental Therapeutics Program,<sup>¶</sup> and Gastrointestinal Tumor Program,<sup>†</sup> H. Lee Moffitt Cancer Center and Research Institute, Tampa, Florida

**Identification of the site of origin for ‘malignancy with unknown primary’ remains a challenge for modern pathology. Correct diagnosis is critical to defining the most beneficial treatment for the patient. Standard pathological approaches combine morphology and immunohistochemical (IHC) studies to first subclassify cytokeratin-positive carcinomas into adenocarcinoma, squamous cell carcinoma, neuroendocrine carcinoma, and urothelial carcinoma. Subsequently, organ-specific IHC-markers, if available, are used to assign the tumor’s primary site of origin. Previous gene expression classifiers have shown promise in tumor classification but cannot readily be integrated into standard practice because they ignore the algorithmic hierarchy used by pathologists. Here we present a novel hybrid approach integrating a hierarchy of gene expression classifiers into the algorithmic method used with IHC. In this method, a tumor is initially assigned to one of the carcinoma subclasses by the top tier classifier. Dependent on initial classification, one of three second-tier classifiers assign primary site resulting in both carcinoma subtype and primary site classification. First tier classifier accuracies were 89%, 88%, and 75% for cross-validation, independent, and institutional independent test sets, respectively. Second tier accuracies were 87%, 90%, and 87% for adenocarcinoma, squamous, and neuroendocrine carcinoma respectively. Therefore, we can successfully separate the four main subtypes of carcinoma and subsequently assign primary site by incorporation of gene expression-based classifiers into the standard algorithmic pathology approach. (*J Mol Diagn* 2010, 12:476–486; DOI: 10.2353/jmoldx.2010.090197)**

Identifying site of primary origin for carcinoma of unknown primary remains a challenge for the pathologist, even with modern pathological techniques. This carries serious implications for cancer therapy, as current onco-

logical therapeutic regimens are targeted to site of origin. Microarray-based gene expression studies are one potential technological solution to this problem, and the feasibility of this methodology for broad-based tumor classification has been established by a number of studies.<sup>1–7</sup> Approaches based entirely on gene expression data, however, limit these studies because they do not take into account well-understood differences in morphology and biological differentiation. Pathologists recognize and exploit these differences in their daily practice. Assessment of morphological features using routine hematoxylin and eosin stains is the first, and many times the last, step in pathological tumor classification, as many malignant neoplasms may be classified with morphology alone. Immunohistochemistry is often part of an algorithmic approach that first separates malignancies into general classes: hematolymphoid, carcinomas, mesothelioma, melanoma, CNS primaries, germ cell neoplasms, and sarcomas. Specific subtypes within each category, except for melanoma and mesothelioma, may be further refined with the use of specific markers. The first key breakpoint is the distinction of hematolymphoid or liquid malignancies from solid malignancies. The next breakpoint is distinguishing among the solid malignancies.

Identification of cytokeratin expression is a key component of this algorithm, as it will delineate carcinomas, the most frequent type of adult malignancy. Mesothelioma and some germ cell tumors also express cytokeratins. Further immunohistochemical studies will separate mesothelioma and germ cell tumor from carcinomas (Figure 1). Carcinomas are then further subtyped into squamous cell carcinoma, adenocarcinoma, neuroendocrine carcinoma; and urothelial carcinoma; these may then be refined by site of origin (Figure 2).

Although the current antibody panels are relatively effective at distinguishing among these various forms of carcinoma, there remain many instances in which the carcinoma type is not determined with objective certainty. Currently available antibody panels are used in a

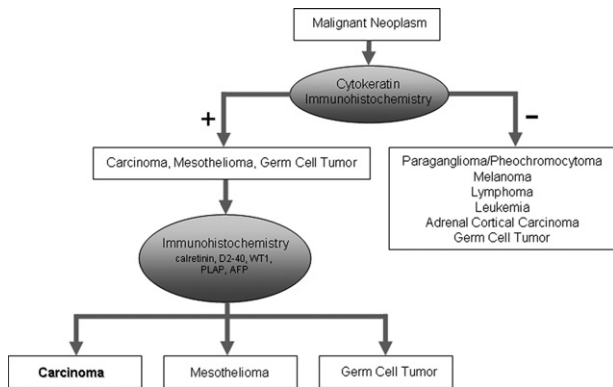
---

Supported by the National Institutes of Health, National Cancer Institute grant RO1 CA112215 (to T.J.Y.).

B.A.C. and G.B. contributed equally to this study.

Accepted for publication February 22, 2010.

Address reprint requests to Timothy J. Yeatman, M.D., Moffitt Cancer Center, 12902 Magnolia Drive, Tampa, FL 33612-9497. E-mail: timothy.yeatman@moffitt.org.

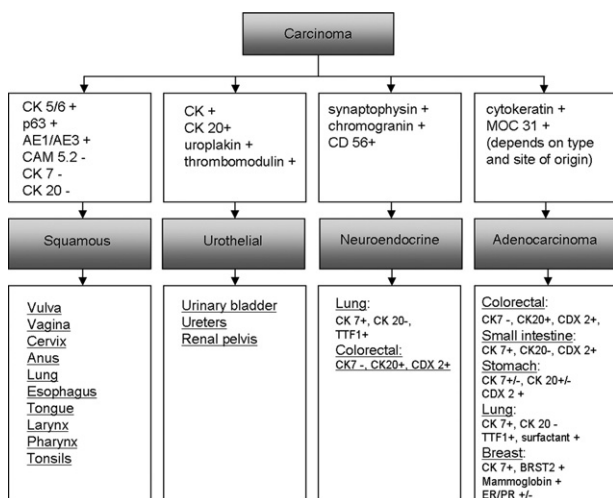


**Figure 1.** Diagram of IHC work flow for the identification of carcinoma from a malignant neoplasm. Initial Cytokeratin IHC separates the neoplasm into positive and negative for CK staining. A second panel of IHC delineates carcinoma from mesothelioma and germ cell tumors.

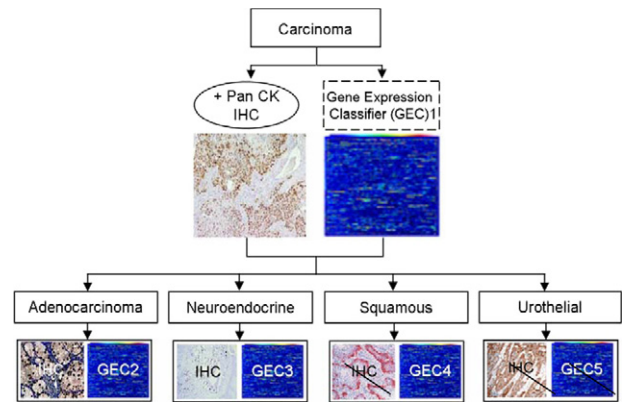
subjective and semiquantitative manner by pathologists, because of nonuniform criteria for determining what qualifies as positive expression.

The inability to distinguish carcinoma subtypes has therapeutic implications, because chemotherapeutic regimens for carcinomas are based not only on the site of primary origin but also on the subtype of carcinoma. As an example, the esophagus may develop both squamous cell carcinoma and adenocarcinoma, yet these different subtypes will each receive a different type of chemotherapy. A neuroendocrine carcinoma will be treated with a specific type of therapy, depending on its differentiation, regardless of site of origin. Thus, a classification of human tumors that skips these distinctions would be missing significant information needed for appropriate treatment decisions.

Here we construct a two-tiered classification scheme based on gene expression data that first delineate neoplasms at the first branch point of cytokeratin positive malignancies, that of carcinoma differentiation, then determine a tumor's site of origin using a group of second level clas-



**Figure 2.** Flow diagram of immunohistochemical staining used to delineate four categories of carcinoma; when available, antibodies used in IHC are shown for primary site of origin identification. Note the absence of primary site of origin antibodies for squamous and urothelial tissues.



**Figure 3.** Flow diagram showing parallel and complementary gene expression classifier and IHC staining used to separate carcinoma into the four major subtypes. **Strike thru lines** indicate no available IHC. No site of origin classifier was constructed for urothelial as origin site plays no role in treatment decision.

sifiers (Figure 3). This classification process is performed in a quantitative and objective manner. Unlike other gene expression-based classification schemes proposed to date, this approach follows the standard work flow used in everyday practice of pathology and can be directly compared with or integrated with the results of IHC staining for each of these critical decision points.<sup>1,4,6,7</sup>

## Materials and Methods

### Sources of Human Microarray Data

Data used to build all classifiers were exclusively derived from tissues arrayed on Affymetrix U133 Plus 2.0 GeneChips (Affymetrix, Santa Clara, CA). Two sets of data were used: microarray data published by the International Genomics Consortium Expression Project for Oncology (expO) and data derived from fresh frozen samples obtained from Moffitt Cancer Center tumor bank.

IGC expO has published more than 1900 tumor samples that have been analyzed on the identical U133 Plus 2.0 GeneChips, making the derived data comparable. The pathology information accompanying each sample was reviewed by a single pathologist (B.A.C.) to delineate the epithelial malignancies into one of the four carcinoma subtypes and into a primary site of origin. Only primary tumors were considered for the analysis. A total set of 561 carcinoma samples were used in this study. Table 1 lists all carcinoma types obtained from the expO dataset. Hepatocellular carcinomas and renal cell carcinomas were delineated as adenocarcinomas for the purpose of this study. In this cohort, adenocarcinomas represent the greatest number of carcinoma subtypes, followed by squamous cell carcinomas, urothelial carcinomas, and neuroendocrine carcinomas.

In addition to the expO data set, 413 tumor samples obtained from the Moffitt Cancer Center tumor bank were arrayed using the U133 Plus 2.0 GeneChip from Affymetrix. A summary of the tumor types and primary sites of origin profiled are listed in Table 2. As for the expO data, all tumor samples derived from the Moffitt Cancer

**Table 1.** IGC Data: Top Primary Sites per Histological Subtype

Histologic subtype	Primary site	Number
Adenocarcinoma	Vulva	1
Adenocarcinoma	Vagina	1
Adenocarcinoma	Appendix	1
Adenocarcinoma	Duodenum	1
Adenocarcinoma	GI tract, NOS	2
Adenocarcinoma	GE junction	2
Adenocarcinoma	Gallbladder	2
Adenocarcinoma	Breast	3
Adenocarcinoma	Small intestine	3
Adenocarcinoma	Unknown	5
Adenocarcinoma	Esophagus	5
Adenocarcinoma	Fallopian tube	6
Adenocarcinoma	Stomach	9
Adenocarcinoma	Cervix	9
Adenocarcinoma	Peritoneum	12
Adenocarcinoma	Pancreas	17
Adenocarcinoma	Colon (rectosigmoid)	35
Adenocarcinoma	Colon (rectum)	36
Adenocarcinoma	Lung	55
Adenocarcinoma	Prostate	59
Adenocarcinoma	Ovary	117
Adenocarcinoma	Uterus	157
Adenocarcinoma	Colon	254
Neuroendocrine	Small intestine	1
Neuroendocrine	Pancreas	2
Neuroendocrine	Lung	7
Squamous	Parotid	1
Squamous	Pharynx	1
Squamous	Penis	1
Squamous	Kidney	1
Squamous	Skin	2
Squamous	Tongue	2
Squamous	Bladder	3
Squamous	Vulva	7
Squamous	Cervix	24
Squamous	Lung	37
Urothelial	Ureterovesicle	1
Urothelial	Urinary tract	1
Urothelial	Ureter	3
Urothelial	Kidney	9
Urothelial	Bladder	27

Center tumor bank were reviewed by a single pathologist (B.A.C.). Cases were selected to include morphological variants when applicable and to include all grades of differentiation to develop classifiers that will be applicable to the widest range of histological variants of these malignancies. RNA extraction was performed using the RNeasy Mini Kit by Qiagen (Valencia, CA). RNase activity was minimized by using the RNase-Free DNase Set by Qiagen. Standard protocols for each of these products were followed. Specimen quality was assessed using the Agilent (Agilent, Santa Clara, CA) BioAnalyzer. The Bioanalyzer software calculates an RNA integrity number on a scale from 1 to 10 for each RNA sample run on the chip. An RNA integrity number >6.5 was taken as the cut-off for accepting the RNA as being of good quality. Specimens that were not of good quality were discarded.

### Expression Value Calculation (RMA)

Robust multi-array analysis (RMA) was used to normalize and calculate gene expression values for all samples

used. Each sample set was treated independently for the purposes of classifier training and testing to ensure that there was no unwanted bias.

### Expression Value Calculation (Incremental RMA)

Incremental RMA (iRMA) is a technique wherein the quantile normalization means and probe-binding affinity parameters from one sample set are saved during the RMA procedure. These two value sets are then used directly by subsequent RMA procedures in lieu of recalculation of these model values for the new sample set.<sup>8</sup> This approach allows for the normalization of gene expression data from different sources to an initial data set without the need to perform RMA on the entire sample set due to the addition of a new sample or set of samples. In the same manner iRMA can be used to normalize an independent test set to the training set on which a classifier was built allowing independent testing of data without the introduction of chip set bias. Expression values for all Independent Test sets were calculated with iRMA using the quantile means and probe binding affinities derived from the previous RMA procedure on the corresponding Training-Test split set.

### Construction of the Carcinoma Subtype Classifier

#### Training-Test Split

The initial training set consisted of 30 randomly selected samples each of squamous cell carcinoma, adenocarcinoma, and urothelial carcinoma and 11 cases of neuroendocrine carcinoma, obtained from the expO data set. All available cases of neuroendocrine carcinoma were used ( $n = 11$ ).

#### Independent Test Set

The initial independent test set consisted of an additional randomly selected 10 samples of squamous cell carcinoma, adenocarcinoma, and urothelial carcinoma. No additional neuroendocrine samples were available from expO. All samples were obtained from the expO dataset. Samples used in the Training-Test split were not considered for selection here, as is the case for all independent test sets described.

#### Institutional Independent Test Set

The institutional independent test set consisted of randomly selected tissues from the Moffitt Cancer Center data set ( $n = 413$  tumor samples). Twenty samples each of squamous cell carcinoma, adenocarcinoma, urothelial carcinoma, and neuroendocrine carcinoma were used for testing.

**Table 2.** Histologic Type and Primary Site of Origin for Specimens Obtained from H. Lee Moffitt Cancer Center Tissue Procurement Facility

Site of origin	Carcinoma type	Total	Subtypes	Grades
Stomach	Adenocarcinoma	39	16 signet ring cell 20 intestinal 3 mucinous	Grade 1:7 Grade 2:13 Grade 3:19
Bladder, ureters, renal pelvis	Urothelial carcinoma	25	No specific subtypes	24 high-grade 1 low-grade
Liver	Hepatocellular carcinoma	31	1 fibrolamellar 1 clear cell 29 NOS	Grade 1:8 Grade 2:15 Grade 3:6 Grade 4:1 Grade X:1
Kidney	Renal cell	61	27 clear cell 10 sarcomatoid 16 papillary 8 chromophobe	Clear cell grades Grade 1:6 Grade 2:7 Grade 3:8 Grade 4:6
Breast	Adenocarcinoma	73	42 ductal 19 lobular 12 mixed	
Anus, rectum, colon	Squamous cell carcinoma	12		Grade 1:3 Grade 2:3 Grade 3:6
Esophagus	Squamous cell carcinoma	5		Grade 1:0 Grade 2:4 Grade 3:1
Larynx	Squamous cell carcinoma	42		Grade 1:7 Grade 2:20 Grade 3:15
Tongue	Squamous cell carcinoma	44		Grade 1:8 Grade 2:24 Grade 3:12
Penis	Squamous cell carcinoma	6		Grade 1:2 Grade 2:3 Grade 3:1
Pancreas	Adenocarcinoma	26	Oncocytic: 1 (no grade assigned) 1 adenosquamous 1 undifferentiated	Grade 1:8 Grade 2:11 Grade 3:5 Grade 4:1
Small intestine	Adenocarcinoma	19		
Various sites	Neuroendocrine carcinoma	25		Grade 1:15 Grade 2:2 Grade 3:8

*Construction of Adenocarcinoma Primary Site Classifier*

*Training–Test Split*

The initial training set consisted of 20 randomly selected samples each of kidney, ovary, uterus, colon, lung, prostate, breast ( $n = 140$ ) obtained from the combined expO dataset and Moffitt derived data. RMA was used for normalization and gene expression signal calculation.

*Independent Test Set*

The independent test set consisted of 10 randomly selected samples each of kidney, ovary, uterus, colon, lung, prostate, and breast obtained from the combined expO and Moffitt data sets. Incremental RMA was applied to this data set using the model values obtained during RMA of the initial training set.

*Construction of Squamous Primary Site Classifier*

*Training–Test Split*

The initial training set consisted of 25 randomly selected samples from a combined tongue and larynx group. Additionally, 3 vulva, 9 cervix, 5 penis, 18 lung, and 8 rectum were randomly selected from the combined expO and Moffitt data set. RMA was used for gene expression signal calculation.

*Independent Test Set*

The independent test set consisted of 11 randomly selected samples from a combined tongue and larynx group. Additionally, 2 vulvar, 4 cervical, 2 penile, 6 pulmonary, and 4 rectal squamous cell carcinomas were randomly selected from the combined expO and Moffitt



data set. Incremental RMA was applied to this data set using the model values obtained during RMA of the initial training set.

### *Construction of Neuroendocrine Primary Site Classifier*

#### *Training–Test Split*

The initial training set consisted of 11 randomly selected samples from a combined small bowel and duodenum group. Additionally, 7 pancreatic and 11 lung neuroendocrine carcinomas were randomly selected from the combined expO and Moffitt data set. RMA was used for gene expression signal calculation.

#### *Independent Test Set*

The independent set consisted of six randomly selected samples from a combined small bowel and duodenum group. Additionally, four pancreatic and four lung neuroendocrine neoplasms were randomly selected from the combined expO and Moffitt data set. Incremental RMA was applied to this data set using the model values obtained during RMA of the initial training set.

### *Identification of Discriminating Genes – Feature Selection*

Identification of a relatively small number of genes that have the ability to distinguish between different tumor categories is a great challenge that is inherent in all large-scale biological assays. To avoid the possibility of selecting a list of genes for the classifier where many or all of the highly significant genes distinguish a minimal number of tumor categories, the following approach was used. A series of four Kruskal–Wallis H tests were performed comparing a single tumor category versus the three remaining tumor categories. This one versus all approach results in four lists of probe sets that were subsequently sorted by *P* value. To construct a classifier with  $n = 50$  probe sets we simply chose the top probe sets from each of the four lists, and then continued to the second probe set from each list. This process is repeated until  $n = 50$  probe sets are chosen. Note that because a single gene is represented by more than one probe set on the Affymetrix U133 Plus chip, the list consists of 50 probe sets rather than 50 individual genes.

### *Classifier Construction*

An artificial neural network (ANN) was chosen for the classifier construction because of its ability to approximate any nonlinear function reasonably well and because no *a priori* assumptions need to be made about the relative importance of any single feature. Fifty input features (probe sets) and five hidden nodes were used to train the ANN for all classifiers constructed. We used a leave *k* out cross validation (LKOCV),  $k = 10\%$ , to assess the accu-

racy of all constructed classifiers. LKOCV in some cases can be slightly optimistic, and two independent training sets were used for further validation in the case of the carcinoma subtype classifier. It should be noted that we performed a “complete” analysis for each sample, meaning that both the gene selection procedure and subsequent ANN training steps were performed for each fold.

## **Results**

### *Classifier Accuracies*

The accuracies for the cross validation of the training set and each of the test sets is shown as confusion matrix tables (Table 3) for the carcinoma subtype classifier. The confusion matrix tables show class by class accuracy and cumulative accuracy. Note that the first independent test set did not include neuroendocrine carcinoma. The training set established an accuracy of 89%. The accuracy of the first independent test set was 88%. An institutional independent test set in which all samples originated from the Moffitt tissue bank resulted in 78% accuracy in separation of the four carcinoma subtypes. Note that the underlying primary site of origin for the IGC tumors, training set, was notably different from the Moffitt tumors, institutional training set contributing to the drop in accuracy. Accuracies for each of the three sites of origin classifiers are presented in Table 4. Again confusion matrices and accuracies for the training cross-validation and independent test sets are presented.

### *Gene Function Analysis*

A total of 32 discriminating genes were identified from the list of 50 Affymetrix probe sets, shown in Table 5. Most of the genes identified are not well characterized or studied in human tumors. However, the protein expression of four of these genes has been previously validated in human tissues. One such example is synaptic vesicle glycoprotein 2A (SV2A), a gene identified as a marker of neuroendocrine carcinomas. SV2 is an integral membrane protein, similar to synaptophysin, a well-established marker of neuroendocrine differentiation. SV2A is one of three well-characterized isoforms of SV2, which include SV2A, SV2B, and SV2C. SV2 immunoreactivity has been observed in neuroendocrine cells of normal stomach, intestines, parathyroid, thyroid, pancreas, and adrenal medulla, as well as nerve structures in all organs.<sup>9</sup> SV2 was found to be expressed in neuroendocrine carcinomas from a variety of organs.<sup>9,10</sup>

Cytokeratin 5, found as a marker of squamous cell carcinoma, is an established component of the antibody panel used to distinguish squamous cell carcinomas from the other carcinoma types. The antibody to cytotokeratin 5/6 combined with p63 is routinely used by pathologists to distinguish squamous cell carcinoma from adenocarcinoma and neuroendocrine carcinomas and a number of publications have confirmed its utility for specific problematic morphological differential diagnoses.<sup>11–14</sup> Dif-

**Table 3.** Confusion Matrices and Accuracies for Training, Independent, and Institutional Independent Test Sets

Training-Test Split				
Urothelial	Squamous	Adenocarcinoma	Neuroendocrine	
25	4	1	0	
3	26	0	0	
0	0	29	1	
0	1	1	9	
CM Accuracy = 89%				
Independent Test Set				
Urothelial	Squamous	Adenocarcinoma	Neuroendocrine	
25	4	1		
3	25	1		
0	0	29		
CM Accuracy = 88%				
Institution Independent Test Set				
Urothelial	Squamous	Adenocarcinoma	Neuroendocrine	
19	1	0	0	
9	11	0	0	
2	0	16	0	
0	5	0	14	
CM Accuracy = 78%				

Note no neuroendocrine samples were available from the expO dataset for the initial independent validation.

fuse expression for CK5/6 is a marker of squamous differentiation.<sup>15</sup>

Desmocollin-3 was shown to have differential expression between squamous cell carcinoma and the other carcinoma subtypes. Immunohistochemical analysis of lung carcinomas showed desmocollin-3 to be expressed in all squamous cell carcinomas but only 1 of 19 adenocarcinomas and 50% of large cell carcinomas. This study validates our findings.<sup>16</sup>

HNF4 is a marker of adenocarcinoma in our classifier. One study found HNF4  $\alpha$  to be a marker of ovarian mucinous carcinomas in fluids.<sup>17</sup> The exact specificity of HNF4 gamma remains to be studied. HNF4 gamma is expressed in the kidneys, gut, pancreas, and testes.<sup>18</sup>

### Discussion

We have successfully developed and tested a new hybrid approach that combines morphological and IHC assessment with a hierarchy of quantitative gene expression-based classifiers into the algorithmic method currently used by pathologists. This approach incorporates a hierarchy of gene expression-based classifiers into the algorithmic method currently used by pathologist to further refine and support their decision making process.

We begin at the point where the pathologist typically initiates the analysis: identifying a neoplasm as carcinoma-based on morphology and immunophenotypic expression for cytokeratins, and then determining whether it falls into one of four main subtypes of carcinoma. The first tier of our classifier similarly begins by assigning a neoplasm defined as carcinoma-based on morphology and cytokeratin expression into one of four carcinoma subtypes: squamous cell, neuroendocrine, adenocarcinoma,

and urothelial. First tier classifier accuracies were 89%, 88%, and 75% for cross-validation, independent, and institutional independent test sets, respectively showing an ability to separate these four subtypes of carcinoma. The identification of SV2, desmocollin-3, CK5, and HNF4 as discriminating genes suggests the selection of genes for the first tier of the classifier may be based on real biological differences because these proteins have been already shown to be differentially expressed in these human tissues. The other genes identified as being discriminatory will need to be validated in tissue samples. Although many carcinomas are easy to subclassify, many pose a challenge because they are poorly differentiated or appear to show combined features of differentiation, such as combined neuroendocrine carcinoma and adenocarcinoma or squamous cell carcinoma and combined adeno- and squamous cell carcinoma.

The next step in pathological assessment is to subclassify the carcinoma relative to the site of primary origin. Current immunohistochemical algorithms to define site of origin are only applicable to adenocarcinomas and well-differentiated neuroendocrine carcinomas. Although the antibody panels are effective at generally narrowing down possible primary sites, they are used in a subjective and qualitative or semiquantitative manner. Furthermore, squamous cell carcinomas are not classifiable as to site of primary origin using currently available antibody panels.

We followed the standard pathology work flow described above by developing a second tier of classifiers that assigned the primary site of origin to the tumor within adenocarcinoma, squamous cell carcinoma, or neuroendocrine carcinoma dependent on initial classification. Second tier classifier accuracies ranged from 83% to

**Table 4.** Confusion Matrices and Accuracies for the Training and Independent Test Sets for the Three Second Tier Primary Site of Origin Classifiers

Adenocarcinoma							
Train-Test Split							
Kidney	Ovary	Uterus	Colon	Lung	Prostate	Breast	
19	0	0	0	0	1	0	
1	9	9	0	0	0	1	
0	4	14	0	0	0	2	
0	0	0	20	0	0	0	
0	0	0	0	20	0	0	
0	0	1	0	0	19	0	
0	1	1	0	2	0	17	
CM Accuracy = 84%							
Independent Test Set							
Kidney	Ovary	Uterus	Colon	Lung	Prostate	Breast	
8	0	2	0	0	0	0	
0	10	0	0	0	0	0	
3	0	6	1	0	0	0	
0	0	1	8	1	0	0	
0	0	0	0	10	0	0	
0	0	0	1	0	9	0	
0	0	0	0	0	0	10	
CM Accuracy = 87%							
Squamous							
Train-Test Split							
Tongue-larynx	Vulva	Cervix	Penis	Lung	Anal-rectum		
24	0	1	0	0	0	0	
0	3	0	0	0	0	0	
2	0	6	0	1	0	0	
4	1	0	0	0	0	0	
1	0	1	0	16	0	0	
0	0	0	0	0	0	8	
CM Accuracy = 83%							
Independent Test Set							
Tongue-larynx	Vulva	Cervix	Penis	Lung	Anal-rectum		
11	0	0	0	0	0	0	
2	0	0	0	0	0	0	
0	0	6	0	0	0	0	
0	0	0	1	1	0	0	
0	0	0	0	4	0	0	
0	0	0	0	0	0	4	
CM Accuracy = 90%							
Neuroendocrine							
Train-Test Split							
Small-bowel	Pancreas	Lung					
9	0	2					
0	7	0					
0	0	11					
CM Accuracy = 93%							
Independent Test Set							
Small-bowel	Pancreas	Lung					
4	0	1					
0	4	0					
0	0	11					
CM Accuracy = 87%							

Samples from the Moffitt and expO data sets were first grouped and tumors randomly selected for both training and independent validation.

**Table 5.** Set of Genes Differentially Expressed among the Four Carcinoma Types

Tumor type	Gene title	Gene symbol
Adenocarcinoma	Hexokinase domain containing 1	<i>HKDC1</i>
	KIAA0152	<i>KIAA0152</i>
	Calmodulin-like 4	<i>CALML4</i>
	Amiloride binding protein 1 (amine oxidase [copper-containing])	<i>ABP1</i>
	Tripartite motif-containing 15	<i>TRIM15</i>
	Hepatocyte nuclear factor 4, gamma	<i>HNF4G</i>
Neuroendocrine	Crystallin, lambda 1	<i>CRYL1</i>
	Yes-associated protein 1, 65 kDa	<i>YAP1</i>
	Kinesin family member 1A	<i>KIF1A</i>
	Suppression of tumorigenicity 18 (breast carcinoma; zinc finger protein)	<i>ST18</i>
	Synaptic vesicle glycoprotein 2A	<i>SV2A</i>
	Cartilage associated protein	<i>CRTAP</i>
	Absent in melanoma 1	<i>AIM1</i>
	Tumor necrosis factor receptor superfamily, member 10b	<i>TNFRSF10B</i>
	Leucine zipper protein 1	<i>LUZP1</i>
	S100 calcium binding protein A16	<i>S100A16</i>
Squamous	Ribosomal protein L39-like	<i>RPL39L</i>
	Hypothetical protein MGC35402	<i>MGC35402</i>
	Lysosomal-associated membrane protein 3	<i>LAMP3</i>
	Keratin 5 (epidermolysis bullosa simplex, Dowling-Meara/Kobner/Weber-Cockayne types)	<i>KRT5</i>
	ATP-binding cassette, sub-family A (ABC1), member 13	<i>ABCA13</i>
	Pleckstrin homology domain containing, family A member 6	<i>PLEKHA6</i>
	Similar to OK/SW-CL.16	<i>LOC440552</i>
	Desmocollin 3	<i>DSC3</i>
	Interferon, gamma-inducible protein 16	<i>IFI16</i>
	Rho GTPase activating protein 23	<i>ARHGAP23</i>
	GATA binding protein 3	<i>GATA3</i>
	Dehydrogenase/reductase (SDR family) member 2	<i>DHRS2</i>
	Leucine-rich repeats and immunoglobulin-like domains 1/leucine-rich repeats and immunoglobulin-like domains 1	<i>LRIG1</i>
Sema domain, transmembrane domain (TM), and cytoplasmic domain, (semaphorin) 6D	<i>SEMA6D</i>	
Hypothetical protein LOC203274	<i>LOC203274</i>	
Ceramide kinase	<i>CERK</i>	
Urothelial		

93%, showing the ability of the gene expression-based classifiers to distinguish a large variety of primary sites.

A number of studies have demonstrated accurate prediction of tumor class by using gene expression-based tumor classification schemes. Most of these gene expression-based classifiers have started with an all-encompassing approach that did not incorporate differences in tumor cell morphology and biology. These studies included solid and liquid tumor types, and unrelated tumor types such as melanoma, carcinoma, and CNS malignancies,<sup>1,5,6</sup> all of which are usually easily distinguished with histomorphology and IHC. Additionally, other studies focused solely on subclassifying a limited spectrum of carcinomas as to site of origin,<sup>3,4,7,19</sup> without distinction as to carcinoma subtype. None of these studies incorporated an approach following the pathologist-based algorithm described in this article.

Two tests are commercially available in the United States, the Pathwork Tissue of Origin Test (Pathwork Diagnostics, Sunnyvale, CA) and the THEROS CancerTYPE ID\* by bioTheranostics (San Diego, CA). Both of these are mRNA-based products. The Pathwork Tissue of Origin Test issues a similarity score for 15 tumor types using a 1550-gene profile that uses the expression level of 1550 transcripts to perform pair-wise comparison between the test sample and each of the 15 tissues on the test panel. A validation study of this test was performed using 547

frozen specimens submitted from four institutions. The tissues were derived from either metastatic cancers or poorly or undifferentiated primary cancers. The test showed a sensitivity of 87.8% and a specificity of 99.4%.<sup>20</sup> A limitation of this validation study is that it was performed using frozen tissues. This validation study is significant because it focused on poorly differentiated or undifferentiated primary carcinomas, and metastatic carcinomas, which are the real challenges in tumor diagnosis. The Pathwork Tissue of Origin Test has now been developed for use in formalin-fixed, paraffin embedded (FFPE) tissues as the PathChip. In a study of 462 FFPE specimens, the test demonstrated 89% positive percent agreement with available diagnoses, and greater than 99% negative percent agreement in specimens that had previously been identified with existing methods as being among the 15 tumor types on the panel.<sup>21</sup> While identifying up to 15 tumor types, most may be distinguished with the application of simple ancillary studies, such as flow cytometry and gene rearrangement studies to diagnose non-Hodgkin lymphoma and immunohistochemistry to diagnose melanomas. Some of the recognized primaries, such as colorectal primaries and breast, have established immunohistochemical patterns. This test might be helpful for those tumor types that do not have a well-defined immunohistochemical pattern or are poorly differentiated or undifferentiated. This test, however, does



not report on differences in tumor morphology, such as squamous cell carcinoma, versus adenocarcinoma versus neuroendocrine carcinoma, features that are more important in predicting cancer therapy and prognosis.

The THEROS CancerTYPE ID\* is designed to focus on those cases that are indeterminate and distinguishes among 39 tumor types. Included in these 39 tumor types are epithelial malignancies, lymphomas, mesotheliomas, meningiomas, stromal neoplasms, and pheochromocytoma. This test does provide information regarding tumor subtype, separating squamous cell carcinomas from adenocarcinomas for certain primary sites, but again uses an “all-encompassing” approach to tumor classification. Many of these separations are coarse distinctions that may be accomplished with the use of widely-available immunohistochemistry. For example, lymphomas may be distinguished from carcinomas with the use of immunohistochemical antibodies for cytokeratins and LCA and even finer distinctions may be routinely made with additional ancillary testing. For example, current practice is to use flow cytometry and gene rearrangement studies to subclassify non-Hodgkin’s lymphoma. Mutations in the *CKIT* gene or *PDGFR* gene are diagnostic for gastrointestinal stromal tumors. This approach is useful for the undifferentiated neoplasms, in which a primary line of differentiation cannot be determined.

The Veridex CUP assay (Raritan, NJ) uses 10 genes tested by RT-PCR to distinguish among six different primary sites of carcinoma: lung, breast, colon, ovary, pancreas, and prostate.<sup>22,23</sup> Although these studies demonstrate the feasibility of this assay, the assay itself left 48% of patients unassigned to an origin.

The CupPrint classifier, being developed by Agendia (Amsterdam, Netherlands), focuses on a finer distinction for adenocarcinoma of unknown primary.<sup>24,25</sup> The CupPrint classifier is developed by using the databases from another published classifier.<sup>5</sup> This is an RT-PCR-based test applicable to formalin-fixed paraffin-embedded tissue. It is a customized eight-pack microarray containing 495 genes that were selected as highly differentiated expressed between 48 tumor types. A weighted five-nearest neighbor algorithm was used to determine the five most molecularly similar tumors in the database. They achieved an accuracy of 83% for carcinomas with a known primary and 94% for a carcinoma of unknown primary. This study more closely approximated our approach, in that it focused mostly on adenocarcinomas, although urothelial carcinomas of the bladder and germ cell tumors, which are not strictly adenocarcinomas, were included in their scheme. Their classifier had a systematic problem in classifying lung and pancreatic carcinomas, misclassifying 63% and 100% of these, respectively. A satisfactory explanation for why this occurred is not provided. This is an important limitation, as these two primary sites most often give rise to adenocarcinoma of unknown primary.

One previous microarray-based gene expression study proposed a tumor classifier—based on a pathological tree-based framework<sup>26</sup> using a schema in which neoplasms were separated in a sequential coarse to fine approach, beginning with the separation of solid malignancies

from hematolymphoid malignancies. The authors further refined the epithelial malignancies into those of Mullerian (ovarian, endometrial) and non-Mullerian origin (breast, prostate, lung, colon, bladder, renal, pancreas). This approach more realistically organizes tumor classification to fit within a pathologist-based diagnostic algorithm. It does, however, leave out the first step typically performed by pathologists, the recognition of morphological subtypes of carcinomas, which include squamous cell carcinomas, urothelial carcinomas, adenocarcinoma, and neuroendocrine carcinomas.

In this study we began at a more practical breakpoint in the pathology-based approach to tumor classification, which starts with the subclassification of cytokeratin positive carcinomas into its four basic types, and follows up with the prediction of site of origin. Studies have focused solely on identifying site of primary origin for adenocarcinoma<sup>3,4,27</sup> proving the effectiveness of using gene expression to classify tumors within specific pathological carcinoma subtypes. Molecular classifiers for site of primary origin for squamous cell carcinoma and neuroendocrine carcinomas have not been developed. One study mentioned an attempt at classifying squamous cell carcinoma of unknown primary<sup>19</sup> and reported no success. Two studies have focused on a very specific differential diagnosis: distinguishing pulmonary from head and neck primary squamous cell carcinomas. One study developed a classifier using Affymetrix HG\_U95Av2 oligonucleotide microarray, which focused specifically on separating lung from tongue squamous cell carcinomas.<sup>28</sup> Another study developed a 10-gene classifier derived from Affymetrix U133 and HG\_U95Av2 data with a 96% accuracy.<sup>29</sup> None has presented a molecular classifier for neuroendocrine carcinoma of unknown primary.

An miRNA classifier has also been developed for carcinoma tissue origin by Rosetta Genomics (Rehovot, Israel).<sup>30</sup> This classifier uses a binary tree method of classification going from coarse to fine specifications. The decision at each node is a simple binary decision that can be performed using the expression levels of a few miRNAs. This classifier was tested on 400 paraffin-embedded and frozen samples from 22 different primary and metastatic tumor tissues. Overall accuracy was >90%. Accuracy for the test reached 89% in an independent data set. The approach described in this article is based on tumor cell differentiation, similar to the approach used by Shedden.<sup>26</sup> This study validates our approach, in that separate miRNAs distinguish among squamous cell and adenocarcinoma of the lung. Carcinoid of the lung is recognized as distinct from other malignancies of the lung. The distinction of neuroendocrine from squamous and glandular carcinomas is the one with which we begin.

Adenocarcinomas are known to have significant morphological variation; therefore, subtype is as important as the site of primary origin. Carcinoma subtype impacts the tumor classification. As an example, mucinous ovarian carcinomas classify with colonic or gastrointestinal primaries<sup>4,19</sup> rather than with ovarian serous type carcinomas. This demonstrates the necessity of including a variety of tumor subtypes and grades associated with a

particular tumor class. For this reason, our classifier is built on a variety of adenocarcinoma types per organ site and includes the various grades of differentiation per type.

Our classification system is the first to show successful classification of two other subtypes of carcinoma: squamous cell and neuroendocrine carcinoma. The squamous cell carcinoma classifier included vulva, cervix, penile, pulmonary, and ano-rectal carcinomas. In clinical practice, vulvar, cervical, and penile carcinomas would not be considered in the same patient, as these are gender-specific cancers. However, this classifier serves as a proof of principle for using gene expression-based classification for squamous cell carcinomas. Interestingly, the tongue and larynx primaries could not be separated, indicating the close embryological relationship of these organs. Future work will expand on the classifier, adding additional possible primary sites from the head and neck and also esophagus.

Neuroendocrine carcinoma, unknown primary continues to be a diagnostic problem in the current practice of oncology and pathology. Frequent sites of metastases include liver, lymph nodes, and bone. In a recent analysis of SEER data, up to 21% of low grade and 50% of high-grade neuroendocrine carcinomas were associated with metastases at the time of diagnosis.<sup>31</sup> In a proportion of these malignancies, the site of primary neuroendocrine carcinoma is not clinically evident. It is, therefore, important to develop diagnostic tools to accurately predict the origin of metastatic neuroendocrine carcinoma, so that the primary tumor may also be treated appropriately. The neuroendocrine carcinomas included in this analysis were from the three most frequent primary sites, pancreas, small bowel, and lung. Missing from these primaries is Merkel cell carcinoma, a primary high-grade neuroendocrine carcinoma of the skin. Merkel cell carcinoma may be distinguished from other neuroendocrine carcinomas by its characteristic CK 7-negative and CK 20-positive immunophenotypic pattern. IHC markers that can be used to determine the site of origin of 'metastatic low-grade neuroendocrine carcinomas' from unknown primary sites include TTF1, CDX2, cytokeratin 7 and 20, neuroendocrine secretory protein-55 (NESP-55), and pancreatic and duodenal homeobox factor-1 (PDX-1).<sup>32-35</sup> Despite site-specificity of these markers, a number of metastatic low-grade neuroendocrine carcinomas in the liver and other metastatic sites remain in the 'unknown primary' category. The molecular classifier proposed here will be a useful adjunct to the currently available IHC markers for more accurate prediction of primary site of origin in case of metastatic neuroendocrine carcinomas from unknown primary sites.

In conclusion, our study shows that distinction among the four basic subtypes of carcinoma and subsequent delineation of primary site of origin is feasible using a tumor classifier derived from standard practice based on morphology and immunohistochemistry, integrated with microarray-based gene expression profiling. This hybrid approach follows the standard pathological workflow for carcinoma classification. This success allows for both integration and direct comparison of microarray-based

classifiers to established pathological techniques for distinguishing carcinomas of unknown primary.

### Acknowledgment

We thank Magaly Mendez for manuscript preparation assistance.

### References

1. Bloom G, Yang IV, Boulware D, Kwong KY, Coppola D, Eschrich S, Quackenbush J, Yeatman TJ: Multi-platform, multi-site, microarray-based human tumor classification. *Am J Pathol* 2004, 164:9-16
2. Bridgewater J, van Laar R, Floore A, Van TVL: Gene expression profiling may improve diagnosis in patients with carcinoma of unknown primary. *Br J Cancer* 2008, 98:1425-1430
3. Buckhaults P, Zhang Z, Chen YC, Wang TL, St Croix B, Saha S, Bardelli A, Morin PJ, Polyak K, Hruban RH, Velculescu VE, Shih Ie M: Identifying tumor origin using a gene expression-based classification map. *Cancer Res* 2003, 63:4144-4149
4. Giordano TJ, Shedden KA, Schwartz DR, Kuick R, Taylor JM, Lee N, Misek DE, Greenson JK, Kardia SL, Beer DG, Rennert G, Cho KR, Gruber SB, Fearon ER, Hanash S: Organ-specific molecular classification of primary lung, colon, and ovarian adenocarcinomas using gene expression profiles. *Am J Pathol* 2001, 159:1231-1238
5. Ma XJ, Patel R, Wang X, Salunga R, Murage J, Desai R, Tuggle JT, Wang W, Chu S, Stecker K, Raja R, Robin H, Moore M, Baunoch D, Sgroi D, Erlander M: Molecular classification of human cancers using a 92-gene real-time quantitative polymerase chain reaction assay. *Arch Pathol Lab Med* 2006, 130:465-473
6. Ramaswamy S, Tamayo P, Rifkin R, Mukherjee S, Yeang CH, Angelo M, Ladd C, Reich M, Latulippe E, Mesirov JP, Poggio T, Gerald W, Loda M, Lander ES, Golub TR: Multiclass cancer diagnosis using tumor gene expression signatures. *Proc Natl Acad Sci USA* 2001, 98:15149-15154
7. Su AI, Welsh JB, Sapinoso LM, Kern SG, Dimitrov P, Lapp H, Schultz PG, Powell SM, Moskaluk CA, Frierson HF Jr, Hampton GM: Molecular classification of human carcinomas by use of gene expression signatures. *Cancer Res* 2001, 61:7388-7393
8. Eschrich SA, Hoerter AM, Bloom GC, Fenstermacher DA: Tissue-specific RMA models to incrementally normalize Affymetrix GeneChip data. *Conf Proc IEEE Eng Med Biol Soc* 2008, 1:2419-2422
9. Portela-Gomes GM, Lukinius A, Grimelius L: Synaptic vesicle protein 2. A new neuroendocrine cell marker. *Am J Pathol* 2000, 157:1299-1309
10. Jakobsen AM, Ahlman H, Wangberg B, Kolby L, Bengtsson M, Nilsson O: Expression of synaptic vesicle protein 2 (SV2) in neuroendocrine tumours of the gastrointestinal tract and pancreas. *J Pathol* 2002, 196:44-50
11. Kaufmann O, Fietze E, Mengs J, Dietel M: Value of p63 and cytokeratin 5/6 as immunohistochemical markers for the differential diagnosis of poorly differentiated and undifferentiated carcinomas. *Am J Clin Pathol* 2001, 116:823-830
12. Kargi A, Gurel D, Tuna B: The diagnostic value of TTF-1, CK 5/6, and p63 immunostaining in classification of lung carcinomas. *Appl Immunohistochem Mol Morphol* 2007, 15:415-420
13. Serrano MF, El-Mofty SK, Gnepp DR, Lewis JS Jr: Utility of high molecular weight cytokeratins, but not p63, in the differential diagnosis of neuroendocrine and basaloid carcinomas of the head and neck. *Hum Pathol* 2008, 39:591-598
14. Khayyata S, Yun S, Pasha T, Jian B, McGrath C, Yu G, Gupta P, Baloch Z: Value of P63 and CK5/6 in distinguishing squamous cell carcinoma from adenocarcinoma in lung fine-needle aspiration specimens. *Diagn Cytopathol* 2009, 37:178-183
15. Chu PG, Weiss LM: Expression of cytokeratin 5/6 in epithelial neoplasms: an immunohistochemical study of 509 cases. *Mod Pathol* 2002, 15:6-10
16. Monica V, Ceppi P, Righi L, Tavaglione V, Volante M, Pelosi G, Scagliotti GV, Papotti M: Desmocollin-3: a new marker of squamous

- differentiation in undifferentiated large-cell carcinoma of the lung. *Mod Pathol* 2009, 22:709–717
17. Sugai M, Umezu H, Yamamoto T, Jiang S, Iwanari H, Tanaka T, Hamakubo T, Kodama T, Naito M: Expression of hepatocyte nuclear factor 4 alpha in primary ovarian mucinous tumors. *Pathol Int* 2008, 58:681–686
  18. Gerdin AK, Surve VV, Jonsson M, Bjursell M, Bjorkman M, Edenro A, Schuelke M, Saad A, Bjurstrom S, Lundgren EJ, Snaith M, Fransson-Steen R, Tornell J, Berg AL, Bohlooly YM: Phenotypic screening of hepatocyte nuclear factor (HNF) 4-gamma receptor knockout mice. *Biochem Biophys Res Commun* 2006, 349:825–832
  19. Tothill RW, Kowalczyk A, Rischin D, Bousioutas A, Haviv I, van Laar RK, Waring PM, Zalberg J, Ward R, Biankin AV, Sutherland RL, Henshall SM, Fong K, Pollack JR, Bowtell DD, Holloway AJ: An expression-based site of origin diagnostic method designed for clinical application to cancer of unknown origin. *Cancer Res* 2005, 65:4031–4040
  20. Monzon FA, Lyons-Weiler M, Buturovic LJ, Rigl CT, Henner WD, Sciulli C, Dumur CI, Medeiros F, Anderson GG: Multicenter validation of a 1,550-gene expression profile for identification of tumor tissue of origin. *J Clin Oncol* 2009, 27:2503–2508
  21. Pillai R, Deeter R, Rigl CT, Halks-Miller M, Buturovic L, Henner D: A microarray-based gene expression test as an aid to tumor diagnosis using formalin-fixed paraffin-embedded (FFPE) specimens. *Pathwork Diagnostics. Abstracts and Case Studies From the College of American Pathologists 2009 Annual Meeting. Arch Pathol Lab Med* 2009, 133:1608–1716
  22. Varadhachary GR, Talantov D, Raber MN, Meng C, Hess KR, Jatko T, Lenzi R, Spigel DR, Wang Y, Greco FA, Abbruzzese JL, Hainsworth JD: Molecular profiling of carcinoma of unknown primary and correlation with clinical evaluation. *J Clin Oncol* 2008, 26:4442–4448
  23. Talantov D, Baden J, Jatko T, Hahn K, Yu J, Rajpurohit Y, Jiang Y, Choi C, Ross JS, Atkins D, Wang Y, Mazumder A: A quantitative reverse transcriptase-polymerase chain reaction assay to identify metastatic carcinoma tissue of origin. *J Mol Diagn* 2006, 8:320–329
  24. Horlings HM, van Laar RK, Kerst JM, Helgason HH, Wesseling J, van der Hoeven JJ, Warmoes MO, Floore A, Witteveen A, Lahti-Domenici J, Glas AM, Van't Veer LJ, de Jong D: Gene expression profiling to identify the histogenetic origin of metastatic adenocarcinomas of unknown primary. *J Clin Oncol* 2008, 26:4435–4441
  25. van Laar RK, Ma XJ, de Jong D, Wehkamp D, Floore AN, Warmoes MO, Simon I, Wang W, Erlander M, van't Veer LJ, Glas AM: Implementation of a novel microarray-based diagnostic test for cancer of unknown primary. *Int J Cancer* 2009, 125:1390–1397
  26. Shedden KA, Taylor JM, Giordano TJ, Kuick R, Misek DE, Rennert G, Schwartz DR, Gruber SB, Logsdon C, Simeone D, Kardias SL, Greenston JK, Cho KR, Beer DG, Fearon ER, Hanash S: Accurate molecular classification of human cancers based on gene expression using a simple classifier with a pathological tree-based framework. *Am J Pathol* 2003, 163:1985–1995
  27. 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
  28. Talbot SG, Estilo C, Maghami E, Sarkaria IS, Pham DK, P Oc, Socci ND, Ngai I, Carlson D, Ghossein R, Viale A, Park BJ, Rusch VW, Singh B: Gene expression profiling allows distinction between primary and metastatic squamous cell carcinomas in the lung. *Cancer Res* 2005, 65:3063–3071
  29. Vachani A, Nebozhyn M, Singhal S, Alila L, Wakeam E, Muschel R, Powell CA, Gaffney P, Singh B, Brose MS, Litzky LA, Kucharczuk J, Kaiser LR, Marron JS, Showe MK, Albelda SM, Showe LC: A 10-gene classifier for distinguishing head and neck squamous cell carcinoma and lung squamous cell carcinoma. *Clin Cancer Res* 2007, 13:2905–2915
  30. Rosenfeld N, Aharonov R, Meiri E, Rosenwald S, Spector Y, Zepeniuk M, Benjamin H, Shabes N, Tabak S, Levy A, Lebanony D, Goren Y, Silberschein E, Targan N, Ben-Ari A, Gilad S, Sion-Vardy N, Tobar A, Feinmesser M, Kharenko O, Nativ O, Nass D, Perelman M, Yosepovich A, Shalmon B, Polak-Charcon S, Fridman E, Avniel A, Bentwich I, Bentwich Z, Cohen D, Chajut A, Barshack I: MicroRNAs accurately identify cancer tissue origin. *Nature Biotechnol* 2008, 26:462–469
  31. Yao JC, Hassan M, Phan A, Dagohoy C, Leary C, Mares JE, Abdalla EK, Fleming JB, Vauthey JN, Rashid A, Evans DB: One hundred years after "carcinoid": epidemiology of and prognostic factors for neuroendocrine tumors in 35,825 cases in the United States. *J Clin Oncol* 2008, 26:3063–3072
  32. Cai YC, Banner B, Glickman J, Odze RD: Cytokeratin 7 and 20 and thyroid transcription factor 1 can help distinguish pulmonary from gastrointestinal carcinoid and pancreatic endocrine tumors. *Hum Pathol* 2001, 32:1087–1093
  33. Jakobsen AM, Ahlman H, Kolby L, Abrahamsson J, Fischer-Colbrie R, Nilsson O: NESP55, a novel chromogranin-like peptide, is expressed in endocrine tumours of the pancreas and adrenal medulla but not in ileal carcinoids. *Br J Cancer* 2003, 88:1746–1754
  34. Srivastava A, Padilla O, Fischer-Colbrie R, Tischler AS, Dayal Y: Neuroendocrine secretory protein-55 (NESP-55) expression discriminates pancreatic endocrine tumors and pheochromocytomas from gastrointestinal and pulmonary carcinoids. *Am J Surg Pathol* 2004, 28:1371–1378
  35. Srivastava A, Hornick JL: Immunohistochemical staining for CDX-2, PDX-1, NESP-55, and TTF-1 can help distinguish gastrointestinal carcinoid tumors from pancreatic endocrine and pulmonary carcinoid tumors. *Am J Surg Pathol* 2009, 33:626–632