

Diagnostic Markers That Distinguish Colon and Ovarian Adenocarcinomas: Identification by Genomic, Proteomic, and Tissue Array Profiling

Satoshi Nishizuka, Sing-Tsung Chen,¹ Fuad G. Gwadry,² Jes Alexander,³ Sylvia M. Major, Uwe Scherf,² William C. Reinhold, Mark Waltham,⁴ Lu Charboneau, Lynn Young, Kimberly J. Bussey, Sohyoung Kim, Samir Lababidi, Jae K. Lee,⁵ Stefania Pittaluga, Dominic A. Scudiero, Edward A. Sausville, Peter J. Munson, Emmanuel F. Petricoin III, Lance A. Liotta, Stephen M. Hewitt, Mark Raffeld, and John N. Weinstein⁶

Genomics and Bioinformatics Group, Laboratory of Molecular Pharmacology [S. N., F. G. G., J. A., S. M. M., U. S., W. C. R., M. W., K. J. B., S. K., S. L., J. K. L., J. N. W.], Laboratory of Pathology [S-T. C., L. C., S. P., L. A. L., S. M. H., M. R.], Office of the Associate Director, Developmental Therapeutics Program, Division of Cancer Treatment and Diagnosis [E. A. S.], National Cancer Institute, Analytical Biostatistics Section, Mathematical and Statistical Computing Laboratory, Center for Information Technology [L. Y., P. J. M.], NIH, Bethesda, Maryland 20814; Science Applications International Corporation-Frederick-National Cancer Institute-Frederick Cancer Research and Development Center, Frederick, Maryland 21702 [D. A. S.]; and Tissue Proteomics Unit, Division of Therapeutic Proteins, Center for Biological Evaluation and Research, Food and Drug Administration, Bethesda, Maryland 20814 [E. F. P.]

ABSTRACT

Colon and ovarian cancers can be difficult to distinguish in the abdomen, and the distinction is important because it determines which drugs will be used for therapy. To identify molecular markers for that differential diagnosis, we developed a multistep protocol starting with the 60 human cancer cell lines used by the National Cancer Institute to screen for new anticancer agents. The steps included: (a) identification of candidate markers using cDNA microarrays; (b) verification of clone identities by resequencing; (c) corroboration of transcript levels using Affymetrix oligonucleotide chips; (d) quantitation of protein expression by “reverse-phase” protein microarray; and (e) prospective validation of candidate markers on clinical tumor sections in tissue microarrays. The two best candidates identified were villin for colon cancer cells and moesin for ovarian cancer cells. Because moesin stained stromal elements in both types of cancer, it would probably not have been identified as a marker if we had started with mRNA or protein profiling of bulk tumors. Villin appears at least as useful as the currently used colon cancer marker cytokeratin 20, and moesin also appears to have utility. The multistep process introduced here has the potential to produce additional markers for cancer diagnosis, prognosis, and therapy.

INTRODUCTION

Colon and ovarian carcinomas can be difficult to distinguish from each other histologically in ovarian masses, in peritoneal carcinomatosis, and in metastases to distant lymph nodes (1–6). It has been reported that 7% of ovarian masses are, in fact, metastases and that most of those are from primary lesions in the gastrointestinal tract (1, 2, 7). Peritoneal carcinomatosis, an unfortunate manifestation of colon and ovarian tumors, also often requires differential diagnosis at sites of dissemination. Overall, adenocarcinomas account for 60% of all cancers of unknown primary, as defined by the detection of one or more metastatic tumors after an initial standard evaluation fails to identify a primary tumor site (8). Additional pathological studies rarely identify a primary site and are indicated in only selected patient subgroups (8). Carcinoma of unknown primary has a generally bad

prognosis, with a median survival of 4–5 months, and the tumors tend to respond poorly to empirical therapy (8, 9). A number of markers have been used in an attempt to distinguish colon and ovarian malignancies by immunohistochemistry. Included have been CA125, CEA,⁷ CK7, and CK20 (3–6, 10). The combination of CK7 and CK20 has become the *de facto* standard against which other markers are judged, but there is still a clear need for additional markers (8, 10). The problem has practical importance for choice of therapy. Colorectal cancer is generally treated with 5-FU; ovarian cancer is generally treated with paclitaxel and a platinum agent (11, 12).

To identify additional markers for this differential diagnosis, we analyzed molecular profiling data on the 60 human cancer cell lines (the NCI-60) used by the Developmental Therapeutics Program of the NCI to screen potential anticancer agents (13–16). More than 100,000 chemical compounds plus a large number of natural product extracts have been screened since 1990, and their patterns of activity have been found to encode rich information about mechanisms of drug action and resistance (15–17). Screening the compounds for activity also profiles the cells for their patterns of drug sensitivity, thereby providing a unique perspective on the molecular pharmacology of cancer (15, 18–21). That perspective has been enhanced by profiling the cells at the molecular level. Many laboratories at the NCI and other institutions have characterized the cells one or a few molecules at a time. Our approach has been to profile the cells more broadly at the DNA, mRNA (19–21), and protein (18) levels. This profiling provided the starting point for the present study. Cell lines in culture are not fully representative of tumor cells *in vivo*, of course, but the RNA and protein profiling studies have indicated that they maintain molecular signatures of their tissues of origin (18–21). It would have been equally reasonable to start the identification process with clinical specimens, but we began with the cell lines because, unlike clinical tumors, they are available in large amounts for study and are homogeneous in cell lineage. Ultimately, the value of the approach would, we felt, depend on its results. For a reason to be discussed later, one of the histological markers identified would probably not have been found if we had begun by profiling bulk clinical tumors.

A schematic view of our overall approach is shown in Fig. 1. The steps were as follows: (a) select candidate markers based on differential colon/ovarian transcript expression levels obtained from 9706-clone cDNA arrays and based on the availability of suitable antibodies for detection as well as biological plausibility; (b) resequence the candidate clones to verify their identity; (c) cross-check the mRNA profiles obtained for candidate genes by cDNA microarray (19, 21) with those obtained for the same samples by Affymetrix oligonucleotide chip (20); (d) use protein lysate arrays (22, 23) in a newly

Received 11/27/02; revised 4/11/03; accepted 6/26/03.

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.

¹ Present address: Hematopathology, USlabs, 2601 Campus Dr., Irvine, CA 92612.

² Present address: Gene Logic Inc., 708 Quince Orchard Road, Gaithersburg, MD 20878.

³ Present address: Duke University School of Medicine, Duke University Medical Center, Durham, NC 27707.

⁴ Present address: St. Vincent's Institute of Medical Research, 9 Princes St., Fitzroy, Melbourne VIC 3065, Australia.

⁵ Present address: Department of Health Evaluation Sciences, P.O. Box 800717, University of Virginia School of Medicine, Charlottesville, VA 22908.

⁶ To whom requests for reprints should be addressed, at Genomics and Bioinformatics Group, Laboratory of Molecular Pharmacology, National Cancer Institute, NIH, Building 37, Room 5056, 9000 Rockville Pike, Bethesda, MD, 20892. Phone: (301) 496-9571; Fax: (301) 402-0752; E-mail: weinstein@dtapx2.ncifcrf.gov.

⁷ The abbreviations used are: CEA, carcinoembryonic antigen; CK, cytokeratin; NCI, National Cancer Institute; 5-FU, 5-fluorouracil; TARP, Tissue Array Research Program; HIER, heat-induced epitope retrieval.

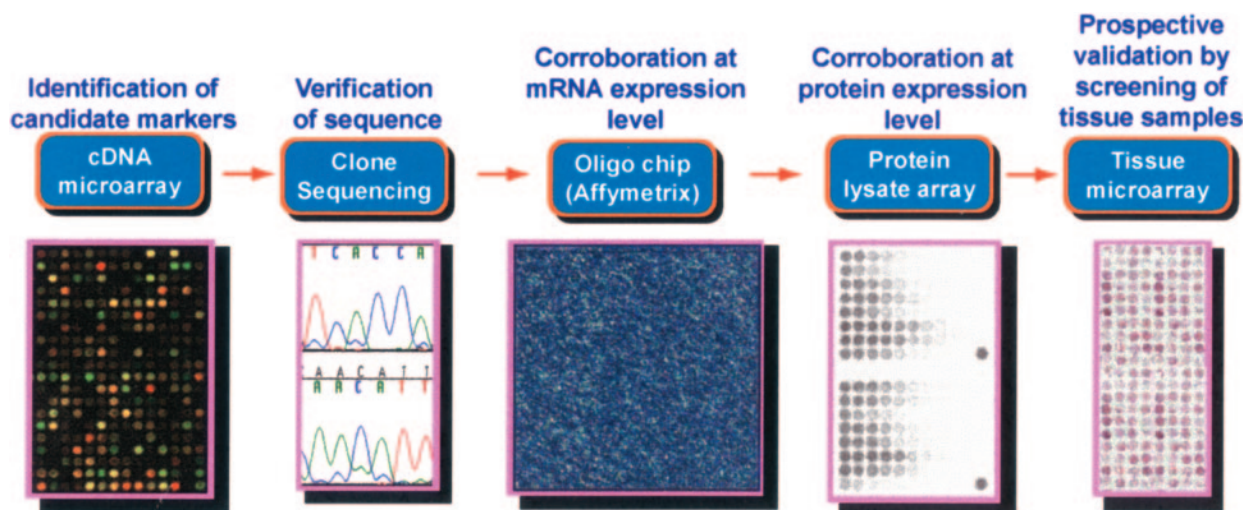


Fig. 1. A schematic overview of the information flow in genomic/proteomic identification and validation of candidate clinical markers for distinguishing ovarian and colon cancers. The process started with cDNA microarray profiling of the NCI-60 cancer cell line panel and continued through prospective validation using tissue microarrays.

developed high-density format to test whether the differential profiles seen at the transcript level would be reflected at the protein level as well; and (e) validate the markers prospectively on clinical materials using tissue microarrays with amplified immunohistochemical detection (24). Two candidate markers that survived this triage process have favorable characteristics for use in clinical pathology for distinguishing colon from ovarian cancer cells. Use of those markers, in combination with each other or with conventional markers, appears to improve the accuracy of discrimination.

MATERIALS AND METHODS

Cell Lines. The NCI-60 set includes seven colon carcinomas (COLO205, HCC-2998, HT29, KM12, SW-620, HCT-116, and HCT-15) and six ovarian carcinomas (SK-OV-3, IGROV1, OCVAR-4, OVCAR-3, OVCAR-5, and OVCAR-8). For this study, we focused on these 13 cell types out of the 60. Xenografts of these lines in nude mice differ widely in histology, from undifferentiated to well differentiated (25). The more differentiated ones, such as HT29 and OVCAR-5, tend to show tubular and acinar/glandular features (25). Immunohistochemical analysis showed CEA staining among the colon lines, particularly COLO205, HT29, and KM12. CA125 and CA19-1 stained primarily the ovarian lines, particularly OVCAR-3 and OVCAR-5 (25).

cDNA Microarrays. The microarrays and hybridization experiments have been described in detail previously (19, 21). Briefly, the arrays were produced by Synteni, Inc. (now Incyte, Inc., Palo Alto, CA). They consisted of robotically spotted, PCR-amplified cDNAs on coated glass slides. Approximately 8000 UniGene clusters were represented by the 9706 cDNA clones (Research Genetics, Huntsville, AL), including 3700 named genes, 1900 human genes homologous with those of other organisms, and 4104 expressed sequence tags of unknown function but defined chromosomal location. Sequence reverification suggested that 15–20% of the array elements contained misidentified cDNA clones or contaminants representing more than one gene (21). For each two-color comparative array hybridization, labeled cDNA was synthesized by reverse transcription from test cell mRNA in the presence of Cy5-dUTP and from reference pool mRNA in the presence of Cy3-dUTP. The reference pool (19, 21) consisted of 12 highly diverse cell lines out of the 60 cell lines. Included were leukemias HL-60 (TB) and K-562, non-small cell lung cancer NCI-H226, colon cancer COLO205, central nervous system cancer SNB-19, melanoma LOX-IMVI, ovarian cancers OVCAR-3 and OVCAR-4, renal cancer CAK-1, prostate cancer PC-3, and breast cancers MCF7 and HS 578T. The cells were harvested, and their RNA was stabilized within 1 min after removal from an incubator at 37°C. Polyadenylated mRNA was then prepared and

characterized as described previously (19, 21). The data used here are available online.⁸

cDNA Microarray Data Analysis. The experiments were those reported previously (19, 21), but the raw red and green channel data were normalized by a more refined algorithm based on fitting with a moving Gaussian kernel to remove the curvature generally observed in scatter plots of red and green channels against each other (26). After normalization, the expression level of each gene in the test cell relative to that in the reference pool was expressed as $E = \log_{10}(\text{CH2FL}/\text{CH1FL})$, where CH2FL and CH1FL are the signal amplitudes in the test cell and reference pool channels, respectively. To identify clones that might potentially distinguish colon and ovarian carcinomas, we then applied several statistical techniques to the expression profiles from the seven colon and six ovarian cell lines. Included as indices of selectivity were (a) the P from an unpaired t test comparing levels observed in the colon and ovarian lines, (b) the mean difference in expression ratio (E) between colon and ovarian lines, and (c) the median difference in amplitude of the single-channel test cell expression signal (CH2FL) between colon and ovarian lines. Single-channel values are not as accurately determined as ratios of red and green channels, but they suggest whether or not the amplitude of signal will be high enough for utility of the marker in *in situ* hybridization and, possibly, immunohistochemistry. On the basis of the three criteria, we identified the top 300 candidate genes in each direction and then clustered their expression patterns using a hierarchical average linkage algorithm with correlation metric (15, 18, 19) in our Clustered Image Map program package (cimminer).⁸

Sequence Verification of Candidate Clones. Once the candidate molecules had been identified (based on the three numerical criteria above plus the availability of antibody and biological plausibility), their IMAGE clones were obtained from Research Genetics or American Type Culture Collection (Manassas, VA) and resequenced. Sequencing was performed using a universal primer pair, GF200, with the pT3T7 or pBluescript vector. Sequences were aligned using either Entrez web tools (National Center for Biotechnology Information, Bethesda, MD) or the DoubleTwist software package (DoubleTwist, Inc., Oakland, CA) to confirm length and location of the clone insert in the full-length cDNA or genomic sequence. Only after full-length resequencing and verification that the clone was correctly identified did we proceed to further evaluation of the candidate marker.

Affymetrix Oligonucleotide Chips. After identification and sequence validation of the candidates, we cross-compared their mRNA expression patterns from the cDNA microarrays with expression patterns that we and our collaborators obtained by running 6800-gene Affymetrix oligonucleotide chips (HuFL 6.8K; Affymetrix, Santa Clara, CA) on the same samples (20). The combination of these two databases generated by different methods with

⁸ <http://discover.nci.nih.gov>.

different sources of error provided mutual corroboration and, therefore, more reliable gene expression data.⁹ The Affymetrix oligonucleotide chip data for the NCI-60 cell lines are available at our web site⁸ in the form used in this analysis.

Analysis of Protein Expression Using Reverse-phase Protein Lysate Arrays. To test whether the colon-ovarian selectivities seen at the transcript level would also be reflected at the protein expression level, we used reverse-phase protein lysate microarrays that included all 60 cell lines plus controls, each at 10 serial 2-fold dilutions. The arrays were designed and produced as described in detail elsewhere,¹⁰ using a modification of a previous protocol for making lower-density arrays (22). Briefly, the arrays of 648 spots were produced on nitrocellulose-coated glass slides (FAST Slides; Schleicher & Schuell, Dassel, Germany) using a pin-in-ring format Affymetrix GMS 417 Arrayer. Protein levels were assessed using the Catalyzed Signal Amplification System (DAKO, Carpinteria, CA), which is based on horseradish peroxidase and diaminobenzidine. Because of the 10 serial dilutions, the nominal dynamic range was ~1000-fold. Array images were processed using a modification of the P-SCAN program package (27). Before immunostaining of the arrays, each antibody was tested against a pool of all 60 cell lines by conventional Western blotting to check that it produced only a single major band at a loading of 20 μ g of protein.

Construction of Tissue Microarrays. TARP1 tumor tissue arrays were designed and produced by the NCI TARP using anonymous donor blocks obtained from the Cooperative Human Tissue Network. Donor blocks that met previously described inclusion criteria were selected and reviewed. Portions containing tumor were selected by one of us (S. M. H.). The arrays were produced as described previously (24), using a manual tissue arrayer (Beecher Instruments, Silver Spring, MD). The array design included specimens of 42 ovarian malignancies of surface epithelial origin, 45 lymphomas, 91 colon carcinomas, 90 non-small cell lung carcinomas, 97 prostate carcinomas, 89 breast carcinomas, 24 melanomas, 25 glioblastomas multiforme, and 62 normal human tissues of various types. Five- μ m sections were cut from the array blocks using tape sectioning materials from Instrumedics (Hackensack, NJ). Every fiftieth slide was stained with H&E for quality control review.

Immunohistochemical Staining of Tissue Microarrays. The arrays were stained using a method based on horseradish peroxidase-labeled polymer (Envision+; DAKO), preceded by an antigen retrieval procedure. The antigen retrieval step was carried out as follows for villin and moesin: the arrays were placed in a microwavable pressure cooker containing preheated buffer [0.1% Tween 20 and 0.01 M citrate buffer (pH 6.0)] and heated in a microwave oven at maximum power (800 W) for 8 min. For retrieval of CK7 and CK20 antigens, the arrays were pretreated with a solution of 0.1% trypsin for 20 min before the pressure cooker procedure. Primary antibodies were applied manually to the arrays, and incubation was carried out for 2 h or overnight at room temperature. The remainder of the procedure for all markers was performed using an automated immunostainer (DAKO) and the Envision+ detection kit (according to manufacturer's instructions). Murine monoclonal anti-villin antibody was obtained from Novocastra Laboratories (Newcastle upon Tyne, United Kingdom) and used at a dilution of 1:200. Murine monoclonal anti-moesin was obtained from Transduction Laboratories (Lexington, KY) and used at a dilution of 1:500. Murine antibodies directed against CK7 and CK20 were obtained from DAKO and used at a dilution of 1:200. The immunohistochemistry was scored as 0 (negative) or 1 (positive) according to the staining of epithelial components.

Statistical Evaluation of the Markers Identified. Statistical analysis was performed for the four single markers and all possible two-, three-, and four-marker combinations thereof. Parameters calculated (Table 3) included the accuracy, sensitivity, specificity, Pearson's χ^2 statistic, likelihood ratio χ^2 , RSquare, and Fisher's exact *Ps*. These calculations were done using the JMP

statistical software package (SAS Institute, Cary, NC). McNemar and bootstrap procedures were also performed, as noted in "Results."

RESULTS

Selection of Candidate Markers Based on cDNA Microarray Expression Studies. As described in "Materials and Methods," we performed unpaired *t* tests for each of the 9706 genes (*i.e.*, cDNA clones) to assess their colon-ovarian selectivities. The 300 genes with the lowest *Ps* for colon selectivity and the 300 genes with the lowest *Ps* for ovarian selectivity were selected for further consideration. Hierarchical clustering of these 600 genes in a Clustered Image Map (15) using only the colon and ovarian data (Fig. 2) showed a clear distinction between the two classes of markers. We next searched among the 600 genes for those with large expression ratio differences ($1/7\sum E_{co} - 1/6\sum E_{ov}$ or $1/6\sum E_{ov} - 1/7\sum E_{co}$, see "Materials and Methods" for definitions). In this way, we identified genes with high relative expression levels favoring either colon or ovarian carcinoma cell lines. Expressed sequence tags of unknown function and clones whose sequences had not been reverified were eliminated from consideration (21). Removing those clones reduced the numbers of candidate genes from 300 in each direction to 122 for colon and 135 for ovarian cancer. Among those genes, we selected candidates for immunohistochemical testing on the basis of the ranking criteria described above and also on the basis of antibody availability. As a result of these statistical and nonstatistical criteria, we focused on villin (IMAGE clone ID 469974, UniGene cluster Hs.166068) as a candidate marker for colorectal adenocarcinoma and moesin (IMAGE 486864, UniGene cluster Hs.170328) as a candidate for ovarian adenocarcinoma. The percentile ranks and indices of colon-ovarian selectivity of expression for these two genes are shown in Table 1 and Fig. 3A.

Sequence Verification. Because cDNA clones are notoriously subject to misidentification or contamination, the likelihood that a clone represents a pure sample of the correct clone on cDNA microarrays has been reported as 60–80% in various studies (21, 28–30). Hence, we resequenced the clones of candidate markers. When we did so, the insert sequence obtained for villin clone 469974 did not match any mRNA sequence available in dbEST or full-length cDNA entry in GenBank. However, when we examined available genomic sequence

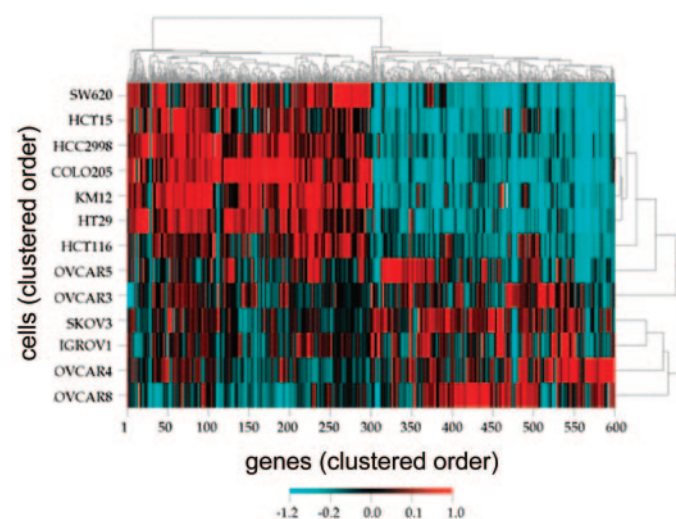


Fig. 2. Clustered Image Map of seven colon and six ovarian carcinoma cell lines for the top 600 candidate clones. Entries are color-coded expression ratio (*E*) values [$\log_{10}(\text{CH2FL}/\text{CH1FL})$]. Red indicates a high positive value; blue indicates a high negative value. There are two distinct expression patterns, corresponding to the two cell classes.

⁹ J. K. Lee, K. J. Bussey, F. G. Gwady, W. Reinhold, G. Riddick, S. L. Pelletier, S. Nishizuka, G. Szakacs, J.-P. Annereau, M. Gottesman, and J. N. Weinstein. Comparing cDNA and oligonucleotide array data: Concordance of gene expression across platforms for the NCI-60 cancer cells. *Genome Biology*, submitted.

¹⁰ Satoshi Nishizuka, Lu Charboneau, Lynn Young, Sylvia Major, William C. Reinhold, Mark Waltham, Hosein Kouros-Mehr, Kimberly J. Bussey, Jae K. Lee, Virginia Espina, Peter J. Munson, Emanuel Petricoin III, Lance A. Liotta, John N. Weinstein. Proteomic profiling of the NCI-60 cancer cell lines using new high-density "reverse-phase" lysate microarrays, submitted.

Table 1 Summary of statistical indices used to select candidate markers

	A. Villin					
	Mean difference in expression $\log_{10}(\text{ratio})^a$		Median difference in intensity levels ^b		<i>P</i> of <i>t</i> test ^c	
	Percentile ^d	Rank	Percentile	Rank	Percentile	Rank
Total genes (9706)	0.0083 (0.83%)	81	0.0093 (0.93%)	91	0.018	176
Sequence-verified genes (4047)	0.0074 (0.74%)	30	0.0091 (0.91%)	37	0.0170	70

	B. Moesin					
	Mean difference in expression $\log_{10}(\text{ratio})$		Median difference in intensity levels		<i>P</i> of <i>t</i> test	
	Percentile	Rank	Percentile	Rank	Percentile	Rank
Total genes (9706)	0.0073 (0.73%)	71	0.0044 (0.44%)	43	0.012	119
Sequence-verified genes (4047)	0.0099 (0.99%)	40	0.0064 (0.64%)	26	0.0130	52

^a Difference = [mean of $\log_{10}(\text{ratio})$ for colon lines] – (mean of $\log_{10}(\text{ratio})$ for ovarian lines).

^b Difference = [(median of single-channel intensity levels for colon lines) – (median of single-channel intensity levels for ovarian lines)]. Hence, the values are positive for villin and negative for moesin.

^c Two-tailed *P* for the null hypothesis that there is no difference between the means of colon and ovarian expression ratios. The internal reference in each case is the cell pool, hence the expression ratios reflect relative colon-ovarian expression levels for villin and moesin.

^d Percentiles and ranks were calculated using only the sequence-verified genes in the set (*i.e.*, those not marked SID in the original database).

information using the DoubleTwist program (Oakland, CA), we found a genomic clone, RP11-398G12, that contains a sequence predicted to be in the last exon of villin. The last exon can be present or absent, depending on which of two polyadenylation signals is selected in the splicing process. Hence, there are villin transcripts of two different lengths (31, 32), both of which accumulate during differentiation of

colon epithelial cells (31–33). The size difference is entirely due to an 800-nucleotide extension of the 3' noncoding region in the longer mRNA. The coding sequences of the two mRNAs are identical (32). Clone 469974 (which is 791-bp long) represents the longer of the two villin transcripts. The two mRNAs for villin and the IMAGE clone alignment are shown schematically in Fig. 3B. Also shown in Fig. 3B is the IMAGE clone 486864 alignment for moesin mRNA based on resequencing.

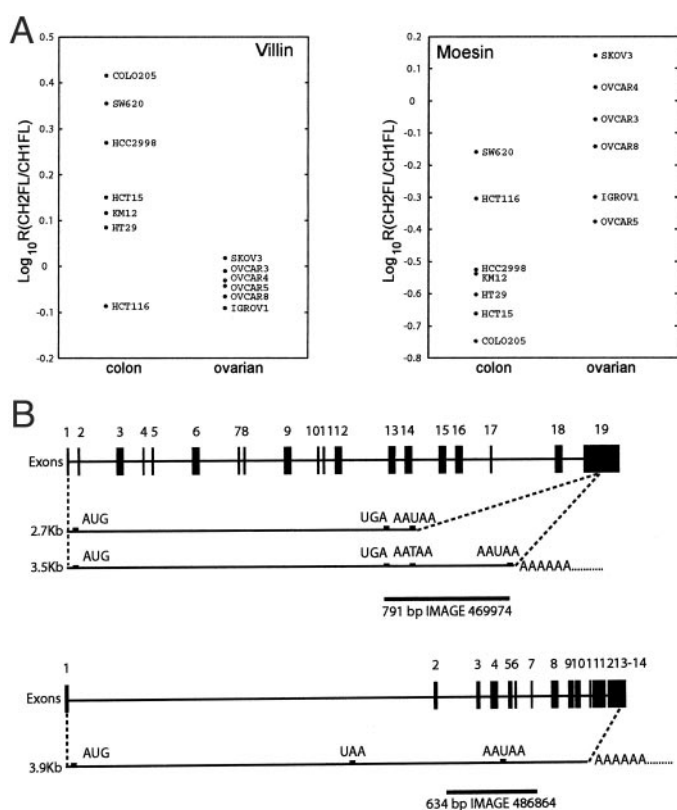


Fig. 3. A, \log_{10} mRNA expression ratio for seven colon and six ovarian carcinoma cell lines. B, top, the IMAGE clone sequence alignment for villin mRNA. Exons are indicated by solid boxes. The two mRNAs are represented below. Positions of translation initiation codons (AUG), stop codons (UGA), and polyadenylation signals (AAUAA or AAUAAA) are indicated. The alignment of the IMAGE clone (791 bp) is indicated by a solid bar. The difference between the two mRNAs in the 3' noncoding region is thought to be generated by alternative choice of polyadenylation signal rather than by alternative splicing of exons (32). Bottom, the IMAGE clone sequence alignment of moesin mRNA. Positions of initiation, stop, and polyadenylation codons are indicated as described for villin, except that the stop codon is UAA.

Corroboration of mRNA Expression Levels Using Affymetrix Oligonucleotide Chips. For transcripts indicated by microarray to be of interest, it is generally necessary to verify expression results using an independent method. In this instance, we had the advantage of having analyzed the same 60 cell samples using Affymetrix oligonucleotide GeneChip arrays (20), which are subject to a very different set of possible errors. Comparison of the seven colon and six ovarian lines revealed surprisingly high correlation between the two data sets for the candidate markers. Pearson's correlation coefficients between cDNA and oligonucleotide arrays for villin and moesin were 0.84 (two-tailed 95% bootstrap confidence interval of 0.54–0.95) and 0.85 (0.56–0.97), respectively. The corresponding Spearman correlation coefficients were 0.69 (0.10–0.95) and 0.78 (0.24–0.99), respectively. Therefore, we concluded that the candidates we had selected on the basis of cDNA microarray data were worthy of further investigation.

Verification of Selectivity at the Protein Level Using Reverse-phase Protein Lysate Microarrays. Because it is well understood that mRNA expression levels may not reflect protein expression levels, especially across disparate cell types, we asked whether the candidate markers selected at the mRNA level would show similar selectivity at the protein level. To address the question, we used reverse-phase protein lysate microarrays recently designed in this laboratory¹⁰ to study all 60 cell lines together on the same slide. First, the monoclonal antibodies to be used for detection of protein on the arrays were checked to see whether they would produce only a single band at the appropriate molecular weight on Western blots (with loading of $\sim 20 \mu\text{g}/\text{lane}$). The antibodies against villin and moesin met that criterion. Immunohistochemical staining of the lysate arrays with these antibodies then showed a qualitative difference between colon and ovarian cell lines. Quantitative analysis using a dose-interpolation algorithm¹⁰ showed a surprisingly high correlation between the cDNA and protein lysate microarray data for both molecules, suggesting transcriptional regulation of expression. The Pearson correlation co-

efficients between cDNA array and protein lysate array data were 0.92 (0.72–0.98) and 0.84 (0.46–0.97), respectively, for villin and moesin. The corresponding Spearman correlation coefficients were 0.88 (0.52–0.98) and 0.86 (0.45–0.99), respectively.

Immunohistochemistry on Tissue Microarrays. To test prospectively whether the results obtained for cell lines would carry over to human tumors, the 133 colon and ovarian tumor specimens on TARP tissue microarrays were examined with antibodies for villin, moesin, and “conventional” markers (CK7 and CK20) for distinguishing colon from ovarian carcinomas. CK7 is usually present in primary ovarian carcinomas but is infrequent in colon carcinomas. Conversely, CK20 is common in colon carcinomas but rare in ovarian carcinomas (4). On the TARP tissue microarrays, CK7 and CK20 showed clear staining patterns in ovarian and colon specimens, respectively. One of the candidate markers, villin, exhibited a strong, distinct pattern of staining for colon (Fig. 4, *a* and *c*). Most of the ovarian cases were villin-negative (Fig. 4*b*), but significant staining was seen in six cases. Stromal regions of both types of tumor were moesin-positive, but this staining was clearly distinguishable from the epithelial staining seen in the ovarian tumors (Fig. 4, *g* and *h*). The results for each specimen are recorded in Table 2.

Statistical Evaluation of the Candidate Markers. Table 2 lists the result obtained for each tissue array specimen when arrays were stained with antibody for villin, moesin, CK7, or CK20. Table 3 then summarizes statistical calculations from Table 2 for each marker and combination of markers. In doing the calculations, we treated villin and CK20 as markers for colon cancer and moesin and CK7 as markers for ovarian cancer. Because no single statistic can capture all aspects of a marker’s quality, a number of different statistics were tabulated. Nominally, the highest accuracy was seen for villin as a marker of colon (0.93), but the distinction from the next best, CK20 as a marker for colon (0.91), is not statistically significant ($P > 0.05$ by McNemar’s test). Note that sensitivity, specificity, positive predictive value, and negative predictive value are symmetrical with respect to colon and ovary in Table 3, given the reciprocal nature of the scoring. CK7 and CK20 showed good sensitivity (0.97 and 0.88) and specificity (0.84 and 0.97) as ovarian and colon markers, respectively. For villin as a colon marker, the corresponding values were 0.97 and 0.85. Moesin, showed reasonably good sensitivity (0.78) and specificity (0.92) as an ovarian marker, but the intensity of staining was not as strong as it was for villin (Fig. 4, *e* and *f*). The marker that gave the fewest incorrect predictions (eight cases) was the new candidate villin. A natural next step in the analysis was to consider the markers in all possible combinations to see whether results better than those for single markers could be obtained. When doing such calculations, decision rules of different types could have been invoked for marker combinations that were not unanimous in their prediction.

“Majority rules” would give the decision to the tumor type with the most markers in its favor. Ties (1 *versus* 1 or 2 *versus* 2) could be left as ambiguous or else resolved in favor of the “best” marker(s) by some criterion such as accuracy. For Table 3 we chose to use the most conservative rule, requiring unanimity for a prediction to be made. We did so in part to permit direct comparison among marker sets without invoking *ad hoc* tiebreakers (that would tend to favor villin over the others) for even numbers of markers. More importantly, we did so because the pathologist faced with conflicting predictions from different markers would be likely to consider the indications as ambiguous and seek other sources of information for making the differential diagnosis. The data in Table 2 can be used by those interested to introduce other decision rules for the calculations.

Pairing of the conventional markers, CK7 and CK20, provided more accurate prediction than did either marker alone. The CK7-CK20 pair yielded 87 correct, 3 incorrect, and 17 ambiguous among 107 informative cases. That is, 17 of the cases (16%) could not be diagnosed on the basis of the conventional markers. The immunostaining with villin and moesin yielded 85 correct, 3 incorrect, and 15 ambiguous among 103 informative cases for the same samples as those tested with antibodies to CK7 and CK20; 87 of 107 (81%) and 85 of 103 (83%) are indistinguishable by McNemar’s test ($P > 0.05$). Inclusion of a third marker decreased the number of incorrect predictions still further (to one or two), and use of all four markers left just one incorrect. The price of adding more markers and requiring unanimity was, of course, an increase in the number of cases left as ambiguous.

DISCUSSION

Colon and ovarian carcinomas can be difficult to distinguish histologically from each other in ovarian masses, in peritoneal carcinomatosis, and in metastases to distant lymph nodes (1–6). Misdiagnosis in this context may result in delayed identification of the primary lesion or misdirected clinical procedures. Misdiagnosis may also lead to inappropriate therapy because metastatic colon cancer is generally treated with 5-FU, whereas ovarian cancer is most often treated with paclitaxel and a platinum agent (11, 12). For patients with carcinoma of unknown primary, a number of palliative chemotherapy approaches using combinations of 5-FU, mitomycin-C, doxorubicin, and cyclophosphamide have been tried, but with limited success (8, 34). Colorectal adenocarcinomas can generally be differentiated from other carcinomas on the basis of ultrastructural features, such as the microfilamentous cores of microvilli, apical electron dense bodies, and glycocalyxal bodies (35). A prospective study by Hickey and Seiler (35) demonstrated high sensitivity (1.0) and specificity (0.97) by ultrastructural features alone. However, electron microscopic studies

Fig. 4. Staining with human anti-villin (*a–d*) and moesin (*e–h*) antibodies on TARP tissue microarrays. Overview of colon (*a* and *e*) and ovarian (*b* and *f*) array specimens. With anti-villin antibody, there is strong staining of colon epithelium (*c*) but no significant staining of anything in most of the ovarian sections (*d*). As shown at higher magnification, stromal cells in the colon sections are positively stained with moesin (*g*), but the tumor epithelium is largely unstained. In contrast, higher magnification of an ovarian specimen (*h*) shows strong epithelial staining. Because of the stromal staining by moesin, the low-power view of colon in (*e*) does not appear impressively negative.

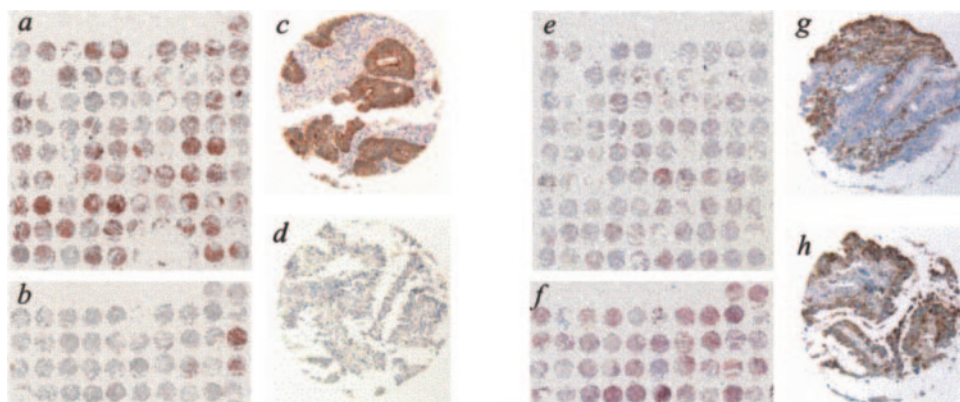


Table 2 *Prospective validation of markers on TARP tissue arrays*

Binary scoring of immunohistochemistry across 133 surgical specimens. Address indicates the position of each specimen on TARP tissue microarrays. Antibodies used are indicated.

Address	Ov/Co ^a	CK7	CK20	Villin	Moesin
1	Ov	NS ^b	NS	0	NS
2	Ov	1	0	0	1
3	Ov	1	0	0	1
5	Ov	1	0	0	TR
7	Ov	1	0	0	1
8	Ov	1	0	0	1
9	Ov	1	0	0	1
10	Ov	1	0	0	1
105	Ov	1	0	0	1
106	Ov	1	NS	0	1
107	Ov	1	0	1	1
109	Ov	1	0	1	1
110	Ov	1	0	1	NS
111	Ov	1	0	0	1
113	Ov	1	0	0	0
114	Ov	1	0	0	0
115	Ov	1	0	0	0
117	Ov	1	0	0	1
118	Ov	1	0	0	0
119	Ov	1	0	1	1
120	Ov	1	1	1	0
121	Ov	1	0	0	1
123	Ov	1	0	0	1
284	Ov	1	0	0	0
285	Ov	1	0	0	1
286	Ov	1	0	0	1
287	Ov	NS	NS	NS	NS
288	Ov	1	0	0	1
289	Ov	1	0	0	1
290	Ov	1	0	0	1
291	Ov	1	0	0	1
292	Ov	1	0	0	1
293	Ov	1	0	0	1
515	Ov	1	0	0	1
516	Ov	NS	NS	NS	NS
517	Ov	1	0	1	0
518	Ov	1	0	0	1
519	Ov	1	NS	0	1
521	Ov	0	0	0	0
522	Ov	1	0	0	1
523	Ov	1	0	0	1
524	Ov	1	NS	NS	NS
27	Co	0	1	1	0
28	Co	0	1	1	0
29	Co	NS	NS	NS	NS
30	Co	0	1	NS	0
31	Co	0	1	NS	0
32	Co	0	0	1	0
33	Co	0	1	1	0
34	Co	NS	NS	NS	NS
35	Co	0	0	1	0
36	Co	NS	NS	NS	NS
40	Co	NS	NS	NS	NS
41	Co	0	0	1	0
42	Co	0	1	1	0
43	Co	0	1	1	0
220	Co	0	1	1	0
221	Co	0	1	1	0
222	Co	1	1	1	0
223	Co	0	1	1	0
224	Co	1	1	1	0
225	Co	1	1	1	0
226	Co	0	0	1	0
227	Co	1	1	1	0
228	Co	0	1	1	0
229	Co	0	1	1	0
230	Co	0	1	1	0
231	Co	0	1	1	TR
232	Co	0	1	1	TR
233	Co	0	1	1	0
234	Co	0	1	1	0
235	Co	0	1	1	NS
236	Co	0	1	1	0
237	Co	0	1	1	0
238	Co	0	1	1	1
239	Co	0	1	1	0
240	Co	1	0	1	0
241	Co	1	1	1	1
242	Co	0	1	1	1
243	Co	NS	NS	1	0

Table 2 *Continued*

Address	Ov/Co ^a	CK7	CK20	Villin	Moesin
244	Co	0	1	1	0
245	Co	NS	NS	NS	0
246	Co	1	1	0	0
247	Co	0	1	1	0
248	Co	0	1	1	0
249	Co	NS	NS	NS	NS
250	Co	0	1	1	TR
251	Co	0	1	1	0
252	Co	0	1	1	0
259	Co	NS	NS	NS	NS
260	Co	0	1	1	0
261	Co	0	1	1	0
262	Co	0	1	1	0
263	Co	0	NS	1	0
265	Co	0	1	1	0
266	Co	0	1	1	0
267	Co	0	NS	NS	NS
268	Co	1	1	1	1
269	Co	0	1	1	0
270	Co	NS	NS	NS	NS
271	Co	NS	NS	1	NS
273	Co	0	1	1	TR
275	Co	1	0	1	0
276	Co	0	1	1	0
277	Co	1	1	1	0
278	Co	0	NS	1	1
279	Co	NS	1	NS	0
280	Co	1	1	NS	NS
281	Co	NS	NS	NS	NS
282	Co	NS	NS	NS	NS
283	Co	0	1	1	0
473	Co	0	1	1	0
474	Co	0	1	1	0
475	Co	0	1	1	0
476	Co	NS	NS	NS	NS
477	Co	0	1	1	0
478	Co	0	1	1	0
479	Co	1	0	0	1
480	Co	0	1	1	0
482	Co	0	1	1	0
483	Co	NS	NS	NS	0
484	Co	0	1	1	0
485	Co	0	1	1	0
486	Co	0	0	1	0
488	Co	0	1	1	0
489	Co	0	1	1	0
490	Co	NS	NS	NS	NS
492	Co	0	1	1	0
493	Co	0	0	1	0
494	Co	0	NS	1	0
495	Co	0	1	1	0
496	Co	0	1	1	0
497	Co	0	1	1	0

^a Ov, ovarian carcinomas; Co, colon carcinomas.

^b NS, non-scorable; TR, trace (counted as negative); 1, positive; 0, negative.

cannot be done routinely, so better markers for standard clinical pathology are required. Current candidates include the well-known marker molecules CEA, CA125, CA19-9, vimentin, CK7, and CK20 (2–6). There are reports that immunostaining for these markers can predict the primary site in metastatic adenocarcinoma correctly 60–80% of the time, but none of the markers is fully site specific (5, 10). Hence, additional markers would be advantageous, and our aim in this study was to provide them. To do so, we pursued the multistep protocol shown schematically in Fig. 1, starting with colon and ovarian cell lines in the NCI-60, using cDNA microarrays, clone sequencing, oligonucleotide arrays, protein lysate arrays, and tissue microarrays.

Clinical tumor specimens have provided the usual starting point for identification of diagnostic markers, sometimes using a high-throughput technology such as the cDNA array or Affymetrix oligonucleotide GeneChip (36, 37). Clinical materials have several drawbacks, however. First, it may be difficult to obtain the amount of RNA or protein necessary. Amplification methods for mRNA are problematical and

Table 3 Summary of statistical indices for single and multiple marker combination

Marker combination	CPC ^a	CPO	OPO	OPC	Ambiguous		Sensitivity		Specificity		PPV		NPV		Accuracy	Pearson χ^2	Likelihood ratio χ^2	RSquare	P	N
					Co	Ov	Co	Ov	Co	Ov	Co	Ov	Co	Ov						
1 Marker																				
Villin	71	2	33	6	NA	NA	0.97	0.85	0.85	0.97	0.92	0.94	0.94	0.92	0.93	79.3	87.3	0.60	<0.0001	112
Moesin	65	6	28	8	NA	NA	0.92	0.78	0.78	0.92	0.89	0.82	0.82	0.89	0.87	53.0	54.5	0.40	<0.0001	107
CK7	63	12	38	1	NA	NA	0.84	0.97	0.97	0.84	0.98	0.76	0.76	0.98	0.89	69.1	81.1	0.55	<0.0001	114
CK20	63	9	35	1	NA	NA	0.88	0.97	0.97	0.88	0.98	0.80	0.80	0.98	0.91	71.4	82.6	0.60	<0.0001	108
2 Markers																				
Villin/moesin	60	1	25	2	6	9	0.98	0.93	0.93	0.98	0.97	0.96	0.96	0.97	0.97	74.4	82.4	0.76	<0.0001	103
Villin/CK7	60	2	31	0	9	7	0.97	1.00	1.00	0.97	1.00	0.94	0.94	1.00	0.98	84.5	103.3	0.87	<0.0001	109
Villin/CK20	58	1	30	1	9	5	0.98	0.97	0.97	0.98	0.98	0.97	0.97	0.98	0.98	81.4	96.9	0.84	<0.0001	104
Moesin/CK7	54	3	28	1	11	7	0.95	0.97	0.97	0.95	0.98	0.90	0.90	0.98	0.95	69.5	80.2	0.73	<0.0001	104
Moesin/CK20	53	1	26	1	12	7	0.98	0.96	0.96	0.98	0.98	0.96	0.96	0.98	0.98	75.6	84.6	0.76	<0.0001	100
CK7/CK20	53	3	34	0	15	2	0.95	1.00	1.00	0.95	1.00	0.92	0.92	1.00	0.97	78.3	98.5	0.83	<0.0001	107
3 Markers																				
Villin/moesin/CK7	52	1	25	0	13	11	0.98	1.00	1.00	0.98	1.00	0.96	0.96	1.00	0.99	73.6	89.4	0.91	<0.0001	102
Villin/moesin/CK20	49	1	23	1	13	10	0.98	0.96	0.96	0.98	0.98	0.96	0.96	0.98	0.97	65.2	75.1	0.81	<0.0001	97
Villin/CK7/CK20	51	1	29	0	16	7	0.98	1.00	1.00	0.98	1.00	0.97	0.97	1.00	0.99	76.8	96.9	0.92	<0.0001	104
Moesin/CK7/CK20	46	1	26	0	18	8	0.98	1.00	1.00	0.98	1.00	0.96	0.96	1.00	0.99	68.8	86.5	0.91	<0.0001	99
4 Markers																				
CK7/CK20/villin/moesin	44	1	23	0	18	11	0.98	1.00	1.00	0.98	1.00	0.96	0.96	1.00	0.99	63.7	78.7	0.90	<0.0001	97

^a CPC, colon predicted as colon; CPO, colon predicted as ovary; OPO, ovary predicted as ovary; OPC, ovary predicted as colon; Ambiguous, cases that are not unanimous for all markers (*i.e.*, both colon and ovarian markers are positive); PPV, positive predictive value; NPV, negative predictive value. All calculations were done without ambiguous cases. Sensitivity (Co) = CPC/(CPC + CPO), Specificity (Co) = OPO/(OPC + OPO), PPV (Co) = CPC/(CPC + OPC), NPV = OPO/(CPO + OPO). Note that Specificity (Co) = Sensitivity (Ov) and PPV (Co) = NPV (Ov), Accuracy = (CPC + OPO)/(CPC + CPO + OPO + OPC), P, P of Fisher's exact test; N, total informative cases; NA, not applicable. Because of missing data, values of N differ.

do not exist for protein. Second, clinical tumors are grossly heterogeneous. They contain endothelial cells, infiltrating leukocytes, fibroblasts, and other stromal cells, as well as tumor cells. The latter may represent only a small fraction of the overall cellular material, effectively diluting out any inherent marker selectivity. Cell lines in culture have the obvious disadvantage that they are not fully representative of cells *in vivo*, but unlike clinical materials, they are available in unlimited numbers and are homogeneous with respect to cell lineage (15). Because they tend to retain gene expression signatures of their organs of origin (19, 21), we chose to use them as starting points for identification of differential markers.

The present study indicates that villin and perhaps moesin should be strongly considered for clinical use in distinguishing colon and ovarian carcinomas. Villin is a M_r 95,000 major microfilament-associated protein that binds actin filaments in a calcium-dependent manner (38, 39) and supports assembly of the actin core bundles of microvilli (33). Accumulation of villin has been observed at the transcript level in HT29 colon adenocarcinoma cells (31, 33) and at the protein level in colon adenocarcinomas but in no other tumor types tested (33). Because the most consistent and specific ultrastructural indicator of colorectal adenocarcinoma appears to be the set of long microfilamentous rootlets extending from the cell surface into the apical cytoplasm (35, 40), it makes sense that villin expression would provide a marker for colorectal carcinoma. Villin has previously been suggested as a marker for colon cancer (41), but, as far as we know, there has been no previous study that provides a statistical basis for its use in the differential diagnosis of colon and ovarian cancers.

Villin offers some potential technical advantages over CK20 in that detection of CKs must generally be enhanced through antigen unmasking using a proteolytic enzyme such as trypsin. This step is difficult to standardize, and it is sensitive to variations in tissue fixation. In contrast, the method we introduced here for antigen unmasking of villin uses a HIER step, rather than an enzymatic degradation. It is our experience that HIER methodologies are easier to standardize and result in consistently good results, regardless of variations in tissue fixation.

The second new candidate marker, moesin, was originally isolated as a heparin-binding protein and identified as a member of the ezrin-radixin-moesin (ERM) family of homologous proteins that serve as general cross-linkers between plasma membranes and actin fila-

ments (42, 43). Interestingly, both villin and moesin are involved in the structural assembly of cytoskeletal elements. Perhaps coincidentally, they are also similar in that phosphorylations (of tyrosine in villin and threonine in moesin) play pivotal roles in the organization of actin-associated cytoskeleton (44, 45).

In our validation studies using TARP tissue microarrays, moesin proved less effective as a single marker than did villin, perhaps partly because ovarian cancers are more diverse than colon cancers. Antimoiesin antibody did not generally stain colon cancer cells but did stain surrounding stromal cells to a considerable extent. Hence, moesin might easily have been missed as a marker for distinguishing immunohistochemically between the two types of tumor cells if we had started with protein or transcript profiling of bulk tumors, rather than cultured cells. As was the case with villin, we were able to use a HIER protocol, rather than enzymatic degradation, for moesin antigen retrieval. However, this advantage was counterbalanced by relatively low expression levels in comparison with those of villin and the CKs. A better antimoiesin antibody and further methodological refinement may be beneficial. Overall, availability of high-quality antibodies was a limiting factor in the triage process described here. To alleviate that problem, we have now screened >400 antibodies of different specificities by Western blot and have incorporated the results (plus extensive information on the antibody reagents) into a user-friendly relational database¹¹ to aid ourselves and others in the search for additional markers.

In this study, we have presented a rather general set of steps that can be used to identify and validate candidate molecular markers. The protocol developed on a series of different types of "omic" (*i.e.*, genomic and proteomic) experiments and analyses (46, 47). Additional studies will, of course, be necessary to test the utility of villin and moesin for use in clinical pathology, but their characteristics as assessed to date using tissue microarrays are promising. Whether or not they might be similarly useful as serum markers remains to be explored.

ACKNOWLEDGMENTS

Multitumor tissue microarray slides were obtained from the TARP of the NCI, NIH (Bethesda, MD). Collaborations with D. Ross, P. O. Brown, M.

¹¹ Sylvia M. Major Satoshi Nishizuka, Rick Rowland, Frank L. Washburn, John N. Weinstein. Abminer: a relational database of antibodies screened against the NCI-60 cancer cell lines, manuscript in preparation.

Eisen, D. Botstein, *et al.* at Stanford and with J. Staunton, T. Golub, E. Lander, *et al.* at the Whitehead Institute produced the cDNA array and Affymetrix chip expression data sets, respectively. We also thank M. J. Miller for sequencing and M. Y. L. Song for immunohistochemical staining. We are grateful to staff of the NCI's Developmental Therapeutics Program for providing the NCI-60 cells and to the late K. D. Paull for his pioneering work on the informatics of the NCI-60 profiling system.

REFERENCES

- Gersell, D. J. Metastatic ovarian tumors: still a challenge. *Am. J. Clin. Pathol.*, **97**: 745–747, 1992.
- Daya, D., Nazerali, L., and Frank, G. L. Metastatic ovarian carcinoma of large intestinal origin simulating primary ovarian carcinoma. *Am. J. Clin. Pathol.*, **97**: 751–758, 1992.
- Wauters, C. C., Smedts, F., Gerrits, L. G., Bosman, F. T., and Ramaekers, F. C. Keratins 7 and 20 as diagnostic markers of carcinomas metastatic to the ovary. *Hum. Pathol.*, **26**: 852–855, 1995.
- Loy, T. S., Calaluce, R. D., and Keeney, G. L. Cytokeratin immunostaining in differentiating primary ovarian carcinoma from metastatic colonic adenocarcinoma. *Mod. Pathol.*, **9**: 1040–1044, 1996.
- Legendijk, J. H., Mullink, H., Van Diest, P. J., Meijer, G. A., and Meijer, C. J. L. M. Tracing the origin of adenocarcinomas with unknown primary using immunohistochemistry. *Hum. Pathol.*, **29**: 491–497, 1998.
- Multhaupt, H. A. B., Arenas-Elliott, C. P., and Warhol, M. J. Comparison of glycoprotein expression between ovarian and colon adenocarcinomas. *Arch. Pathol. Lab. Med.*, **123**: 909–916, 1999.
- Ulbricht, T. M., Lawrence, M. R., and Stehman, F. B. Secondary ovarian neoplasia. *Cancer (Phila.)*, **53**: 1164–1174, 1984.
- Hainsworth, J. D., and Greco, F. A. Management of patients with cancer of unknown primary site. *Oncology (Basel)*, **14**: 563–574, 2000.
- Greco, F. A., Burris, H. A., III, Erland, J. B., Gray, J. R., Kalman, L. A., Schreeder, M. T., and Hainsworth, J. D. Carcinoma of unknown primary site. *Cancer (Phila.)*, **89**: 2655–2660, 2000.
- Brown, R. W., Campagna, L. B., Dunn, J. K., and Cagle, P. T. Immunohistochemical identification of tumor markers in metastatic adenocarcinoma. *Am. J. Clin. Pathol.*, **107**: 12–19, 1997.
- Moertel, C. G. Chemotherapy for colorectal cancer. *N. Engl. J. Med.*, **330**: 1136–1142, 1994.
- Qazi, F., and McGuire, W. P. The treatment of epithelial ovarian cancer. *CA Cancer J. Clin.*, **45**: 88–101, 1995.
- Grever, M. R., Schepartz, S. A., and Chabner, B. A. The National Cancer Institute: cancer drug discovery and development program. *Semin. Oncol.*, **19**: 622–638, 1992.
- Boyd, M. R., and Paull, K. D. Some practical considerations and applications of the National Cancer Institute *in vitro* anticancer drug discovery screen. *Drug Dev. Res.*, **34**: 91–109, 1995.
- Weinstein, J. N., Myers, T. G., O'Connor, P. M., Friend, S. H., Fornace, A. J., Jr., Kohn, K. W., Fojo, T., Bates, S. E., Rubinstein, L. V., Anderson, N. L., Buolamwini, J. K., van Osdol, W. W., Monks, A. P., Scudiero, D. A., Sausville, E. A., Zaharevitz, D. W., Bunow, B., Viswanadhan, V. N., Johnson, G. S., Wittes, R. E., and Paull, K. D. An information-intensive approach to the molecular pharmacology of cancer. *Science (Wash. DC)*, **275**: 343–349, 1997.
- Paull, K. D., Shoemaker, R. H., Hodes, L., Monks, A., Scudiero, D. A., Rubinstein, L., Plowman, J., and Boyd, M. R. Display and analysis of patterns of differential activity of drugs against human tumor cell lines: development of mean graph and COMPARE algorithm. *J. Natl. Cancer Inst. (Bethesda)*, **81**: 1088–1092, 1989.
- Weinstein, J. N., Kohn, K. W., Grever, M. R., Viswanadhan, V. N., Rubinstein, L. V., Monks, A. P., Scudiero, D. A., Welch, L., Koutsoukos, A. D., Chiusa, A. J., and Paull, K. D. Neural computing in cancer drug development: predicting mechanism of action. *Science (Wash. DC)*, **258**: 447–451, 1992.
- Myers, T. G., Waltham, M., Li, G., Buolamwini, J. K., Scudiero, D. A., Rubinstein, L. V., Paull, K. D., Sausville, E. A., Anderson, N. L., and Weinstein, J. N. A protein expression database for the molecular pharmacology of cancer. *Electrophoresis*, **18**: 647–653, 1997.
- Scherf, U., Ross, D. T., Waltham, M., Smith, L. H., Lee, J. K., Tanabe, L., Kohn, K. W., Reinhold, W. C., Myers, T. G., Andrews, D. T., Scudiero, D. A., Eisen, M. B., Sausville, E. A., Pommier, Y., Botstein, D., Brown, P. O., and Weinstein, J. N. A gene expression database for the molecular pharmacology of cancer. *Nat. Genet.*, **24**: 236–244, 2000.
- Staunton, J. E., Slonim, D. K., Coller, H. A., Tamayo, P., Angelo, M. J., Park, J., Scherf, U., Lee, J. K., Weinstein, J. N., Mesirov, J. P., Lander, E. S., and Golub, T. R. Chemosensitivity prediction by transcriptional profiling. *Proc. Natl. Acad. Sci. USA*, **98**: 10787–10792, 2001.
- Ross, D. T., Scherf, U., Eisen, M. B., Perou, C. M., Rees, C., Spellman, P., Iyer, V., Jeffrey, S. S., Van de Rijn, M., Waltham, M., Pergamenschikov, A., Lee, J. C. F., Lashkari, D., Shalon, D., Myers, T. G., Weinstein, J. N., Botstein, D., and Brown, P. O. Systematic variation in gene expression patterns in human cancer cell lines. *Nat. Genet.*, **24**: 227–235, 2000.
- Paweletz, C. P., Charboneau, L., Bichsel, V. E., Simone, N. L., Chen, T., Gillespie, J. W., Emmert-Buck, M. R., Roth, M. J., Petricoin, E. F., III, and Liotta, L. A. Reverse phase protein microarrays which capture disease progression show activation of pro-survival pathways at the cancer invasion front. *Oncogene*, **20**: 1981–1989, 2001.
- Liotta, L., and Petricoin, E. Molecular profiling of human cancer. *Nat. Rev. Genet.*, **1**: 48–56, 2000.
- Kononen, J., Bubendorf, L., Kallioniemi, A., Barlund, M., Schraml, P., Leighton, S., Torhorst, J., Mihatsch, M. J., Sauter, G., and Kallioniemi, O. P. Tissue microarrays for high-throughput molecular profiling of tumor specimens. *Nat. Med.*, **4**: 844–847, 1998.
- Stinson, S. F., Alley, M. C., Kopp, W. C., Fiebig, H. H., Mullendore, L. A., Pittman, A. F., Kenney, S., Keller, J., and Boyd, M. R. Morphological and immunocytochemical characteristics of human tumor cell lines for use in a disease-oriented anticancer drug screen. *Anticancer Res.*, **12**: 1035–1053, 1992.
- Zhou, Y., Gwadry, F. G., Reinhold, W. C., Miller, L. D., Smith, L. H., Scherf, U., Liu, E. T., Kohn, K. W., Pommier, Y., and Weinstein, J. N. Transcriptional regulation of mitotic genes by camptothecin-induced DNA damage: microarray analysis of dose- and time-dependent effects. *Cancer Res*, **62**: 1688–1695, 2002.
- Carlisle, A. J., Prabhu, V. V., Elkahoul, A., Hudson, J., Trent, J. M., Linehan, W. M., Williams, E. D., Emmert-Buck, M. R., Liotta, L. A., Munson, P. J., and Krizman, D. B. Development of a prostate cDNA microarray and statistical gene expression analysis package. *Mol. Carcinog.*, **28**: 12–22, 2000.
- Knights, J. When the chips are down. *Nature (Lond.)*, **410**: 860–861, 2001.
- Halgren, R. G., Fielden, M. R., Fong, C. J., and Zacharewski, T. R. Assessment of clone identity and sequence fidelity for 1189 IMAGE cDNA clones. *Nucleic Acids Res.*, **29**: 582–588, 2001.
- Taylor, E., Cogdell, D., Coombes, K., Hu, L., Ramdas, L., Tabor, A., Hamilton, S., and Zhang, W. Sequence verification as quality-control step for production of cDNA microarrays. *Biotechniques*, **31**: 62–65, 2001.
- Pringault, E., Arpin, M., Garcia, A., Finidori, J., and Louvard, D. A human villin cDNA clone to investigate the differentiation of intestinal and kidney cells *in vivo* and in culture. *EMBO J.*, **5**: 3119–3124, 1986.
- Pringault, E., Robine, S., and Louvard, D. Structure of the human villin gene. *Proc. Natl. Acad. Sci. USA*, **88**: 10811–10815, 1991.
- Robine, S., Huet, C., Moll, R., Sahuquillo-Merino, C., Coudrier, E., Zweibaum, A., and Louvard, D. Can villin be used to identify malignant and undifferentiated normal digestive epithelial cells? *Proc. Natl. Acad. Sci. USA*, **82**: 8488–8492, 1985.
- Dowell, E. D., Garrett, A. M., Shyr, Y., Johnson, D. H., and Hande, K. R. A randomized Phase II trial in patients with carcinoma of an unknown primary site. *Cancer (Phila.)*, **91**: 592–597, 2001.
- Hickey, W. F., and Seiler, M. W. Ultrastructural markers of colonic adenocarcinoma. *Cancer (Phila.)*, **47**: 140–145, 1981.
- Beer, D. G., Kardia, S. L. R., Huang, C.-C., Giordano, T. J., Levin, A. M., Misek, D. E., Lin, L., Chen, G., Gharib, T. G., Thomas, D. G., Lizyness, M. L., Kuick, R., Hayasaka, S., Taylor, J. M. G., Iannettoni, M. D., Orringer, M. B., and Hanash, S. Gene-expression profiles predict survival of patients with lung adenocarcinoma. *Nat. Med.*, **8**: 816–824, 2002.
- Giordano, T. J., Shedden, K. A., Schwartz, D. R., Kuick, R., Yaylor, J. M. G., Lee, N., Misek, D. E., Greenson, J. K., Kardia, S. L. R., Beer, D. G., Rennert, G., Cho, K. R., Gruber, S. B., Fearon, E. R., and Hanash, S. Organ-specific molecular classification of primary lung, colon, and ovarian adenocarcinomas using gene expression profiles. *Am. J. Pathol.*, **159**: 1231–1238, 2001.
- Bretscher, A., and Weber, K. Villin: the major microfilament-associated protein of the intestinal microvillus. *Proc. Natl. Acad. Sci. USA*, **76**: 2321–2325, 1979.
- Bretscher, A., and Weber, K. Villin is a major protein of the microvillus cytoskeleton which binds both G and F actin in a calcium-dependent manner. *Cell*, **20**: 839–847, 1980.
- Sharma, S., Tan, J., Sidhu, G., Wiczek, R., Miller, D. C., and Cassi, N. D. Lung adenocarcinomas metastatic to the brain with and without ultrastructural evidence of rootlets: an electron microscopic and immunohistochemical study using cytokeratins 7 and 20 and villin. *Ultrastruct. Pathol.*, **22**: 385–391, 1998.
- Bacchi, C. E., and Gown, A. M. Distribution and pattern of expression of villin, a gastrointestinal-associated cytoskeletal protein, in human carcinomas: a study employing paraffin-embedded tissue. *Lab. Invest.*, **64**: 418–424, 1991.
- Lankes, W., Griesmacher, A., Grünwald, J., Schwartz-Albiez, R., and Keller, R. A heparin-binding protein involved in inhibition of smooth-muscle cell proliferation. *Biochem. J.*, **251**: 831–842, 1988.
- Lankes, W. T., and Furthmayr, H. Moesin: a member of the protein 4.1-talin-ezrin family of proteins. *Proc. Natl. Acad. Sci. USA*, **88**: 8297–8301, 1991.
- Hayashi, K., Yonemura, S., Matsui, T., and Tsukita, S. Immunofluorescence detection of ezrin/radixin/moesin (ERM) proteins with their carboxyl-terminal threonine phosphorylated in cultured cells and tissues. Application of a novel fixative protocol using trichloroacetic acid (TCA) as a fixative. *J. Cell. Sci.*, **112**: 1149–1158, 1999.
- Zhai, L., Zhao, P., Panebra, A., Guerrero, A. L., and Khurana, S. Tyrosine phosphorylation of villin regulates the organization of the actin cytoskeleton. *J. Biol. Chem.*, **276**: 36163–36167, 2001.
- Weinstein, J. N. Fishing expeditions. *Science (Wash. DC)*, **282**: 627, 1998.
- Weinstein, J. N. "Omic" and hypothesis-driven research in the molecular pharmacology of cancer. *Curr. Opin. Pharmacol.*, **2**: 361–365, 2002.