Selection of Potential Markers for Epithelial Ovarian Cancer with Gene Expression Arrays and Recursive Descent Partition Analysis

Karen H. Lu,[2] Andrea P. Patterson,[1] Lin Wang,[1] Rebecca T. Marquez,[1] Edward N. Atkinson,[3] Keith A. Baggerly,[3] Lance R. Ramoth,[1] Daniel G. Rosen,[5] Jinsong Liu,[5] Ingegerd Hellstrom,[6] David Smith,[7] Lynn Hartmann,[7] David Fishman,[8] Andrew Berchuck,[9] Rosemarie Schmandt,[2] Regina Whitaker,[9] David M. Gershenson,[2] Gordon B. Mills,[4] and Robert C. Bast, Jr.[1]

[1]Ovarian Cancer Research Laboratory, Department of Experimental Therapeutics, and Departments of [2]Gynecologic Oncology, [3]Biostatistics, [4]Molecular Therapeutics, and [5]Pathology, University of Texas M. D. Anderson Cancer Center, Houston, Texas; [6]Pacific Northwest Research Institute, Seattle, Washington; [7]Mayo Clinic, Rochester, Minnesota; [8]Northwestern University Medical Center, Chicago, Illinois; and [9]Duke University Medical Center, Durham, North Carolina

ABSTRACT

Purpose: **Advanced-stage epithelial ovarian cancer has a poor prognosis with long-term survival in less than 30% of patients. When the disease is detected in stage I, more than 90% of patients can be cured by conventional therapy. Screening for early-stage disease with individual serum tumor markers, such as CA125, is limited by the fact that no single marker is up-regulated and shed in adequate amounts by all ovarian cancers. Consequently, use of multiple markers in combination might detect a larger fraction of early-stage ovarian cancers.**

Experimental Design: **To identify potential candidates for novel markers, we have used Affymetrix human genome arrays (U95 series) to analyze differences in gene expression of 41,441 known genes and expressed sequence tags between five pools of normal ovarian surface epithelial cells (OSE) and 42 epithelial ovarian cancers of different stages, grades, and histotypes. Recursive descent partition analysis (RDPA) was performed with 102 probe sets representing 86 genes that were up-regulated at least 3-fold in epithelial ovarian cancers when compared with normal OSE. In addition, a panel of 11 genes known to encode potential tumor markers [mucin 1, transmembrane (MUC1), mucin 16 (CA125), mesothelin, WAP four-disulfide core domain 2 (HE4), kallikrein 6, kallikrein 10, matrix metalloproteinase 2, prostasin, osteopontin, tetranectin, and inhibin] were similarly analyzed.**

Results: **The 3-fold up-regulated genes were examined and four genes [Notch homologue 3 (NOTCH3), E2F transcription factor 3 (E2F3), GTPase activating protein (RACGAP1), and hematological and neurological expressed 1 (HN1)] distinguished all tumor samples from normal OSE. The 3-fold up-regulated genes were analyzed using RDPA, and the combination of elevated claudin 3 (CLDN3) and elevated vascular endothelial growth factor (VEGF) distinguished the cancers from normal OSE. The 11 known markers were analyzed using RDPA, and a combination of HE4, CA125, and MUC1 expression could distinguish tumor from normal specimens. Expression at the mRNA level in the candidate markers was examined via semiquantitative reverse transcription-PCR and was found to correlate well with the array data. Immunohistochemistry was performed to identify expression of the genes at the protein level in 158 ovarian cancers of different histotypes. A combination of CLDN3, CA125, and MUC1 stained 157 (99.4%) of 158 cancers, and all of the tumors were detected with a combination of CLDN3, CA125, MUC1, and VEGF.**

Conclusions: **Our data are consistent with the possibility that a limited number of markers in combination might identify >99% of epithelial ovarian cancers despite the heterogeneity of the disease.**

INTRODUCTION

Ovarian epithelial carcinoma claims more lives than any other gynecological cancer in industrialized countries (1). It is the fifth most common cancer in American women and the fifth most common cause of cancer death (2). Whereas the 5-year survival for women presenting with early-stage disease is ~90%, the majority of women (75%) are diagnosed with late-stage disease (stage III or stage IV) and have a 5-year survival of less than 30% (3). Mortality might be reduced if the disease were detected in the early stages.

Screening for early-stage disease with individual serum tumor markers, such as CA125, is limited by the fact that no single marker is up-regulated and shed in adequate amounts by all ovarian cancers. Although CA125 represents the best available serum marker, achieving 50% sensitivity and 99% specificity for early-stage disease, it detects only ~80% of all ovarian cancers (4). In fact, epithelial ovarian cancer is a heterogeneous disease. Histological subtypes of epithelial ovarian cancer, including serous, endometrioid, mucinous, and clear cell carcino-

Received 10/15/03; revised 1/30/04; accepted 2/10/04.
Grant support: Supported by Ovarian SPORE Grant CA 83639, NIH, Department of Health and Human Services, and the CORE Grant CA 16772–28; K. Lu was supported by American Association of Obstetricians and Gynecologists Foundation.
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.
Note: K. Lu and A. Patterson contributed equally to this report.
Requests for reprints: Robert C. Bast, Jr., Department of Experimental Therapeutics, Box 355, University of Texas M. D. Anderson Cancer Center, 1550 Holcombe Boulevard, Houston, TX 77030. Phone: (713) 792-7743; Fax: (713) 742-7864; E-mail: rbast@mdanderson.org.

mas are known to have different clinical characteristics as well as different molecular features (4, 5). For example, many mucinous tumors do not secrete high levels of CA125 (6). Therefore, a panel of complementary tumor markers will be required to detect all cases of ovarian cancer at an early stage. If screening is to be performed with individual assays, a limited number of markers must encompass the heterogeneity of the disease.

To identify potential candidates for novel markers, we assembled a group of 42 ovarian cancers of different histological subtypes and compared their gene expression to that in five pools of normal ovarian epithelial tissue scrapings using Affymetrix arrays. Genes were sought whose level of expression in all cancers exceeded the level of expression in normal epithelial scrapings. We then used recursive descent partition analysis (RDPA) to seek up-regulated genes that would distinguish different histotypes of ovarian cancer from each other and from normal ovarian epithelial tissue. Similar analysis was undertaken with a panel of 11 genes encoding previously reported tumor markers: mucin 1, transmembrane (MUC1), mucin 16 (CA125), mesothelin, WAP four-disulfide core domain 2 (HE4), kallikrein 6, kallikrein 10, matrix metalloproteinase 2, prostasin, osteopontin, tetranectin, and inhibin. Each of these genes has been shown to be up-regulated in ovarian cancers. Individually, these genes have not been shown to have sufficient specificity to function as a tumor marker for ovarian cancer. We were interested in whether combinations of these genes would increase overall detection.

MATERIALS AND METHODS

Tumor Samples and RNA Preparation. Forty-two flash-frozen primary ovarian cancers were obtained from University of Texas M. D. Anderson Cancer Center (Houston, TX), Duke University (Durham, NC), and the Mayo Clinic (Rochester, MN; Table 1). All of the tumors were classified according to grade and stage using standard FIGO (Federation International of Gynecology and Obstetrics) criteria. Five pools of normal ovarian epithelial brushings from 42 different individuals were obtained from Northwestern University. Patients ranged in age from 43 to 79 years with a median age of 55; two-thirds were postmenopausal and one-third, premenopausal. Of the postmenopausal donors, ~60% were on hormone replacement therapy. The normal cells were collected using a cytobrush and were immediately suspended and frozen in RLT buffer (Qiagen, Valencia, CA). Total RNA was extracted from all of the ovarian cancers and normal ovarian epithelial scrapings using the RNeasy Kit (Qiagen) according to the manufacturer's protocol. Institutional Review Board approval had been obtained at each participating institution prior to the initiation of this study.

Gene Expression Analysis. The Affymetrix GeneChip Human Genome U95 set of oligoarrays (Affymetrix, Santa Clara, CA) was used to obtain gene expression data. This series tests the expression of more than 41,441 human genes and expressed sequence tags. The biotinylated cRNA preparation, hybridization, and scanning of the microarrays were performed according to the manufacturer's protocols. Data were collected using GeneChip software (Affymetrix). Data were analyzed using the software program dChip (7). We used Version 1.2 with the PM-only model to estimate differences. The genes listed had to pass two filters: the fold change for each gene between the normal ovarian epithelium scrapings and the ovarian cancers had to exceed 3, and the absolute difference in expression levels between the two means had to exceed 100 (to avoid distraction at the noise level, we found that on average 58% of the 41,441 genes and expressed sequence tags exhibited values greater than 100).

Statistics. RDPA was performed on the 102 most up-regulated probe sets, representing 86 genes, to find combinations of genes whose expression distinguished the samples based on histology. The analysis was performed using the JMP software (SAS Inc., Cary, NC). The values used for the test were the normalized hybridization intensities obtained from dChip. In RDPA (8, 9, 10), a classification tree is constructed that gives

Table 1 Clinical information for the 42 ovarian carcinoma tumor tissues[a]

Sample	Histology	Stage	Grade	Age
OV76	Serous	IC	III	70
OV134	Serous	IC	III	61
OV234	Serous	IA	III	73
OV632	Serous	IA	III	52
OV702	Serous	IC	III	55
OV1059	Serous	IA	III	44
OV1140	Serous	IA	III	43
OV746R	Serous	IC	III	68
OV979R	Serous	IC	III	55
MDA 1	Serous	III C	III	63
MDA 2	Serous	III C	III	69
MDA 4	Serous	III C	III	62
MDA 5	Serous	III C	III	69
MDA 6	Serous	III C	III	58
MDA 7	Serous	III C	III	82
MDA 8	Serous	III C	III	73
MDA 9	Serous	III C	III	52
ENDO 1389	Endometrioid	III C	III	72
ENDO 1462	Endometrioid	III C	III	52
ENDO 1720	Endometrioid	I B	II	31
ENDO 1931	Endometrioid	III	II	56
ENDO 1956	Endometrioid	II C	III	49
ENDO 1999	Endometrioid	I C	III	76
ENDO 2252	Endometrioid	II B	II	72
ENDO 2977	Endometrioid	III C	III	51
ENDO 3240	Endometrioid	II C	III	53
CC 2420	Clear Cell	IV	III	49
CC 2732	Clear Cell	II C	III	51
CC 2893	Clear Cell	III C	III	54
CC 2898	Clear Cell	I A	III	59
CC 2953	Clear Cell	III C	III	58
CC 2551	Clear Cell	I C	III	64
CC 2287	Clear Cell	III C	II	70
MUC 815	Mucinous	I	II	68
MUC 897	Mucinous	I A	I	48
MUC 1865	Mucinous	III C	II	63
MUC 2014	Mucinous	III C	I	88
MUC 1726	Mucinous	II C	II	75
MUC 1884	Mucinous	I A	I	36
MUC 2708	Mucinous	III C	III	39
MUC 2790	Mucinous	IV	III	38
MUC 3215	Mucinous	IV	II	47

[a] All OV samples were provided by the Mayo Clinic, all MDA samples were provided by M. D. Anderson Cancer Center, and all ENDO, CC and MUC samples were provided by Duke University Medical Center.

Table 2 102 probe sets (86 genes), up-regulated at least 3-fold compared with normal ovarian epithelium, used for recursive descent partition analysis

Accession no.	Probe set	Gene
U11863	37186_s_at	*amiloride binding protein 1*
AJ238979	32349_at	*annexin A10*
AF038451	38827_at	*anterior gradient 2 homolog*
AA632130	46588_f_at	*ATPase, H+ transporting, lysosomal, β polypeptide*
M18728	36105_at	*carcinoembryonic antigen-related cell adhesion molecule 6*
L33930	266_s_at	*CD24 antigen*[a]
M13699	39008_at	*ceruloplasmin*
D43950	40417_at	*chaperonin containing TCP1, subunit 5*
Y08374	36197_at	*chitinase 3-like 1*
X15998	38112_g_at	*chondroitin sulfate proteoglycan 2 (versican)*
X15998	38111_at	*chondroitin sulfate proteoglycan 2 (versican)*
AB000714	33904_at	*claudin 3*[a]
AI276313	87864_i_at	*claudin 4*[a]
AA044205	46205_at	*Cluster Including AA044205:zk50f08.r1*
AA046853	46274_at	*Cluster Including AA046853:zf14f11.r1*
AA164745	85613_at	*Cluster Including AA164745:zo93g03.s1*
AA464792	82705_at	*Cluster Including AA464792:zx83h05.r1*
AA523697	86402_f_at	*Cluster Including AA523697:ni50b10.s1*[a]
AA564760	80723_f_at	*Cluster Including AA564760:nk52f10.s1*[a]
AA584310	48774_at	*Cluster Including AA584310:nn79g01.s1*
AA703100	83553_f_at	*Cluster Including AA703100:zi78g02.s1*[a]
AA705851	72962_at	*Cluster Including AA705851:ah42f05.s1*
AA992380	56019_f_at	*Cluster Including AA992380:ot37g06.s1*[a]
AI141067	75658_f_at	*Cluster Including AI141067:oz43h05.x1*[a]
AI346914	48797_at	*Cluster Including AI346914:qp59c05.x1*
AI347618	82359_at	*Cluster Including AI347618:qo99f06.x1*[a]
AI636169	63904_at	*Cluster Including AI636169:tz06c10.x1*[a]
AI672389	78487_at	*Cluster Including AI672389:ty64f02.x1*
AI676059	77364_at	*Cluster Including AI676059:wc04g08.x1*
AI683338	63431_f_at	*Cluster Including AI683338:tw50h07.x1*
AI697470	45294_at	*Cluster Including AI697470:tq08h01.x1*
AI733409	77702_f_at	*Cluster Including AI733409:oq55c11.x5*
AI768465	78574_f_at	*Cluster Including AI768465:wh22c03.x1*
AI769970	67362_f_at	*Cluster Including AI769970:wj22h10.x1*
AI924096	82101_at	*Cluster Including AI924096:wn78g12.x1*[a]
AW020970	54904_at	*Cluster Including AW020970:df17a05.y1*
W55924	51263_at	*Cluster Including W55924:zc03g11.s1*
M26576	39333_at	*collagen, type IV, α1*
AA948682	73188_s_at	*collagen, type IV, α1*[a]
AI989792	73130_s_at	*cyclin-dependent kinase inhibitor 2A*
M22760	41223_at	*cytochrome c oxidase subunit Va*
AI961431	48684_at	*DKFZP434G032 protein*
D38550	41632_at	*E2F transcription factor 3*
AI765533	34335_at	*ephrin-B2*
U71207	35226_at	*eyes absent homolog 2 (Drosophila)*
AA128249	38430_at	*fatty acid binding protein 4, adipocyte*
X02761	31719_at	*fibronectin 1*
M10905	31720_s_at	*fibronectin 1*
M10905	311_s_at	*fibronectin 1*
U82984	50271_at	*GTPase activating protein*[a]
AI525822	56429_g_at	*hematological and neurological expressed 1*
AI525822	56428_at	*hematological and neurological expressed 1*[a]
AW005489	64363_f_at	*high-mobility group protein isoforms I and Y*[a]
L17131	39704_s_at	*high-mobility group protein isoforms I and Y*[a]
AI823992	59461_at	*hypothetical protein FLJ10461*
H67928	91728_at	*hypothetical protein PRO2605*
U97188	37558_at	*IGF-II mRNA-binding protein 3*
AI991845	48864_at	*interferon, α-inducible protein 27*[a]
J03909	39728_at	*interferon, γ-inducible protein 30*
X04430	38299_at	*interleukin 6*
M17017	35372_r_at	*interleukin 8*
U28386	40407_at	*karyopherin α2*
AI739473	74557_s_at	*KIAA0018 gene product*
D14657	38116_at	*KIAA0101 gene product*
AI762213	32821_at	*lipocalin 2*[a]
AF071219	41066_at	*mammaglobin 2*[a]
L22524	668_s_at	*matrix metalloproteinase 7*[a]

Table 2 Continued

Accession no.	Probe set	Gene
X16396	40074_at	*methylene tetrahydrofolate dehydrogenase*[a]
AI990317	75026_s_at	*methylene tetrahydrofolate dehydrogenase*
M97676	40199_at	*msh homeo box homolog 1 (Drosophila)*
J05581	38784_g_at	*mucin 1, transmembrane*[a]
X52229	700_s_at	*mucin 1, transmembrane*[a]
J05581	38783_at	*mucin 1, transmembrane*
J05582	927_s_at	*mucin 1, transmembrane*[a]
U97669	38750_at	*Notch homolog 3 (Drosophila)*
AL039458	34800_at	*ortholog of mouse integral membrane glycoprotein*
D13666	1451_s_at	*osteoblast specific factor 2 (fasciclin I-like)*[a]
AI971498	51016_s_at	*phosphoserine aminotransferase*[a]
U65011	157_at	*preferentially expressed antigen in melanoma*
AF027208	41470_at	*prominin-like 1 (mouse)*[a]
L33881	1602_at	*protein kinase C, iota*
AI763378	54147_at	*pyruvate dehydrogenase kinase, isoenzyme 4*
AA634799	77218_at	*pyruvate dehydrogenase kinase, isoenzyme 4*
AA131149	34319_at	*S100 calcium binding protein P*
AF052124	34342_s_at	*secreted phosphoprotein 1*[a]
J04765	2092_s_at	*secreted phosphoprotein 1*[a]
AI864016	74815_at	*secreted phosphoprotein 1*
X01683	36781_at	*serine (or cysteine) proteinase inhibitor, clade A*
AI985046	44982_s_at	*small inducible cytokine subfamily B, member 14*[a]
AB019987	34878_at	*SMC4*
U81800	33143_s_at	*solute carrier family 16, member 3*
AI587292	48048_at	*solute carrier family 34, member 2*
X02308	37899_at	*thymidylate synthetase*
L08044	31477_at	*trefoil factor 3 (intestinal)*
M93036	575_s_at	*tumor-associated calcium signal transducer 1*[a]
J04152	291_s_at	*tumor-associated calcium signal transducer 2*[a]
X77753	41286_at	*tumor-associated calcium signal transducer 2*
AW024572	73146_at	*vascular endothelial growth factor*[a]
AF022375	36100_at	*vascular endothelial growth factor*
AI565773	48563_at	*v-erb-b2 avian erythroblastic leukemia viral*
M34309	1585_at	*v-erb-b2 erythroblastic leukemia viral oncogene*
X63187	33933_at	*WAP four-disulfide core domain 2*[a]

[a] Probe sets that maintained a lower bound of at least 3-fold for a 90% confidence interval for the fold change differences between the normal ovarian epithelium and the ovarian carcinomas.

decision rules for assigning a sample to a category based on a series of sequential decisions. At each stage, a single predictor is used, and, depending on whether the value of the predictor is greater than or less than a selected cutoff value, the sample is assigned to a left or right node. Each of the resulting nodes is then analyzed using the same procedure, although different predictor variables and cutoff values may be used. The cutoff value is selected to maximize the likelihood-ratio χ^2 statistic for the test that the probability of a particular case belonging to a given group is independent of whether the predictor of that case is above or below the cutoff value. Thus, the cutoff value is chosen to make cases above and below the cutoff value as different as possible with respect to classification. This procedure continues until the data in each node are sufficiently well discriminated or until there are too little data in any node to support further analysis.

Semiquantitative RT-PCR. Semiquantitative reverse transcription (RT)-PCR was performed using the Bioanalyzer 2100 (Agilent Technologies, Germantown, MD). The DNA 500 LabChip kit (Agilent Technologies) was used to determine the expression levels of Notch homologue 3 (NOTCH3), E2F transcription factor 3 (E2F3), GTPase activating protein (RACGAP1), hematological and neurological expressed 1 (HN1), CA125, HE4, CLDN3, MUC1, and vascular endothelial growth factor (VEGF). All of the chips were prepared as instructed by the manufacturer. Data were normalized using glyceraldehyde-3-phosphate dehydrogenase (GAPDH) as an internal control. Negative control reactions without reverse transcriptase were used to identify genomic contamination. The linear range of each primer set was first determined. The final conditions for NOTCH3, E2F3, RACGAP1, and HN1 were 30 cycles and a 55°C annealing temperature, whereas the cycle number for the five remaining genes was 35. The expression level of each gene was determined for individual tumors and compared with the average expression level in the five normal pools.

Tissue Microarrays and Immunohistochemistry. Tissue microarrays were constructed using 158 epithelial ovarian cancer tissues. Included in the tissue microarrays were 113 serous, 23 endometrioid, 12 clear cell, and 9 mucinous ovarian carcinomas (OVCAs) with 5, 16, 7, and 7 early-stage samples, respectively. Immunohistochemistry was performed on the arrays as well as 5 normal ovarian epithelial tissues using antibodies against CA125, HE4, CLDN3, MUC1, and VEGF. Staining for CA125 was performed according to the manufacturer's protocol using 2 μg/ml of the OC125 antibody (Dako, Carpinteria, CA), an antigen retrieval of microwave for 10 min, and a

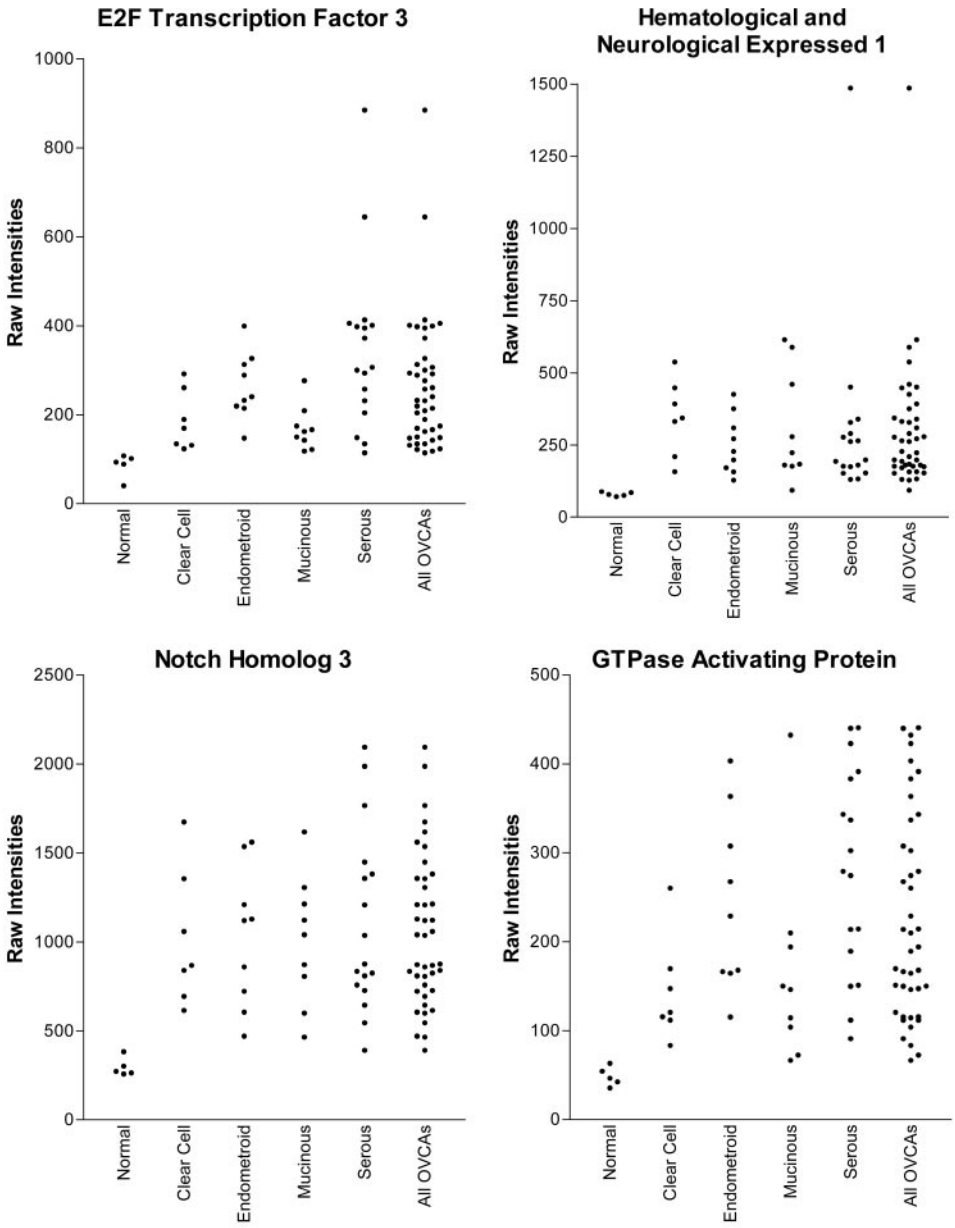

Fig. 1 Scatter plot of microarray data for E2F transcription factor 3, hematological and neurological expressed 1, notch homologue 3 and GTPase activating protein. With recursive descent partition analysis (RDPA), these genes independently separated tumor samples from normal samples. Data are shown separated by histotype and combined ovarian carcinomas (*OVCAs*) for all four genes.

primary antibody incubation of 1 h at room temperature. HE4 antibody was prepared at the Pacific Northwest Research Institute, Seattle, WA (11). The staining was carried out with an antigen retrieval of microwave for 5 min, a primary antibody concentration of 10 μg/ml αHE4:2H5 antibody, and a primary antibody incubation of 4°C overnight. Commercially available antibodies for CLDN3 (Polyclonal Rabbit anti-Claudin 3, Zymed Laboratories Inc., South San Fransisco, CA), MUC1 (Muc-1 Core Glycoprotein monoclonal antibody, Novocastra Laboratories Ltd., Newcastle upon Tyne, United Kingdom), and VEGF [Vascular Endothelial Growth Factor Ab-3 (JH121), NeoMarkers, Inc., Fremont, CA] were used at a 1:50 dilution and stained according to manufacturers' protocols. With CLDN3, antigen retrieval was performed by boiling for 15 min, and the primary incubation was 1 h at room temperature. The antigens for MUC1 and VEGF were retrieved via pressure cooker and microwave for 10 min, respectively. The antibody incubations were performed at room temperature for 1 h (MUC1) or 2 h (VEGF). A 10% cutoff was used to determine staining in a given sample.

RESULTS

Expression Array Analysis and RDPA by Normal or Cancer Status. From the gene expression profiles of 42 ovarian cancer tumor tissue specimens and 5 pools of normal ovarian epithelium scrapings, we identified 86 genes that were up-regulated 3-fold or greater and reported absolute differences in

expression levels exceeding 100 (Table 2). To look for the most robust genes, we applied a third filter requiring the lower bound of a 90% confidence interval for the fold change to exceed 3 (7). Genes passing all three filters are annotated in Table 2. Using RDPA on the 86 genes, considering only tumor and normal as the classification criteria, we found four genes that perfectly separated the tumors from normal: E2F3, HN1, NOTCH3, and RACGAP1. The scatterplots of the microarray data in Fig. 1 show the perfect separation between normal and tumor for the four genes. A fairly wide range was noted in the amount of up-regulation between different cancers.

RDPA by Histotype. The 86 genes that were up-regulated 3-fold were evaluated for their ability to distinguish different histotypes from each other and from normal ovarian epithelial scrapings using RDPA. Two genes in combination, CLDN3 and VEGF, achieved a complete separation of the ovarian tumors from normal (Fig. 2*A*). Elevated CLDN3 identified all serous, endometrioid, and clear cell cancers. Elevated VEGF distinguished mucinous cancers from normal ovarian surface epithelium.

Subsequently, we examined the ability of genes for 11 previously known markers to distinguish different histotypes from each other and from normal ovarian epithelial scrapings. Among the 42 cancers, the number of tumors with a 2-fold elevation in each marker ranged from one for prostasin and inhibin to 37 for MUC1 (Table 3). When all 11 markers were elevated, RDPA was able to discriminate normal tissue from tumor using HE4, CA125, and MUC1 (Fig. 2*B*). Elevated HE4 separated 100% of the serous tumors, 89% of the endometrioid tumors, 43% of the clear cell tumors, and 22% of the mucinous cancers from the normals. An elevated CA125 separated the remaining clear cell tumors from the normals. MUC1 distinguished 100% of the clear cell tumors from the normals. Of note, 78% of the mucinous cancers tumors and 11% of the endometrioid tumors had a low CA125. Scatterplots of microarray data for the five genes identified by RDPA (CLDN3, VEGF, HE4, CA125, and MUC1) are shown in Fig. 3.

Semiquantiative RT-PCR. Overall, nine markers were identified by inspection of the initial scatterplots (Fig. 1) and by RDPA (Fig. 2). To confirm up-regulation of expression, semiquantitative RT-PCR was performed with 20 samples that included 5 cases from each of the four histotypes. In general, RT-PCR data correlated well with array data.

Immunohistochemical Reactivity of Five Markers. Antibodies could be obtained for five of the nine candidate markers. Immunohistochemical analysis of CLDN3, VEGF, HE4, CA125, and MUC1 in 158 ovarian cancers is shown in Fig. 4 and Table 4. A combination of CLDN3 and MUC1 detected 155 (98%) of 158 cases, a combination of CLDN3, CA125, and MUC1 included 157 (99.4%) of 158 cases, and all tumors were detected with a combination of CLDN3, CA125, MUC1, and VEGF.

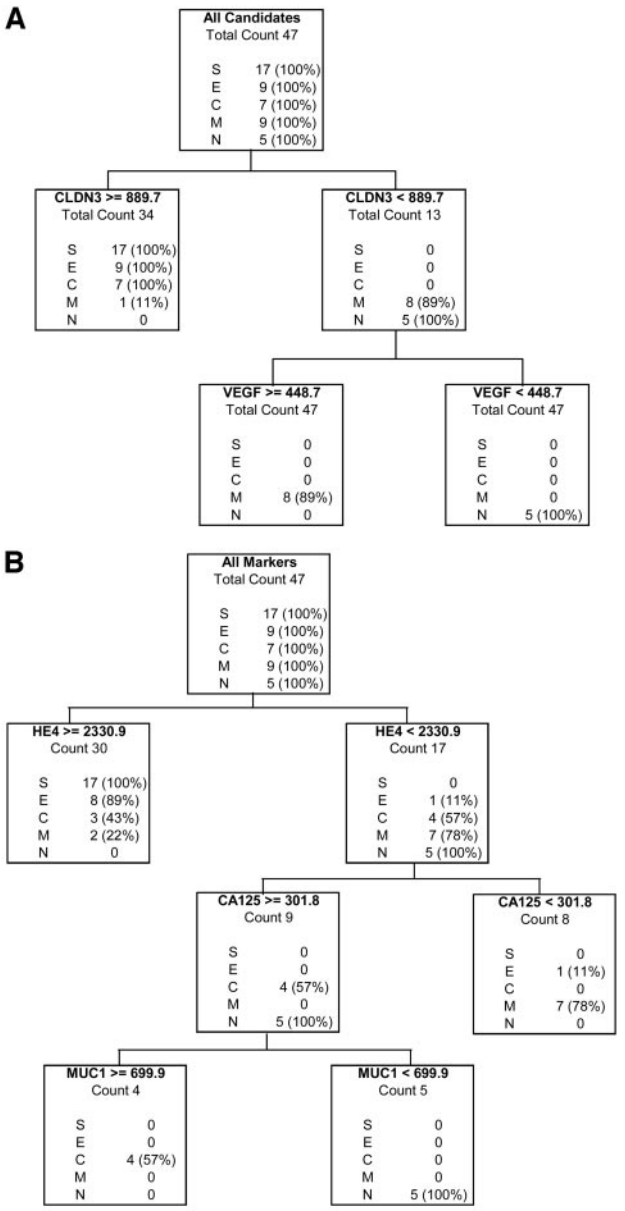

Fig. 2 Recursive descent partition analysis (RDPA). *A*, recursive descent partitioning tree for 102 candidate variables. Claudin 3 (*CLDN3*) initially separates all of the serous (*S*), endometrioid (*E*), and clear cell (*C*) tumors and one of the mucinous (*M*) tumors from all of the normal (*N*) samples and the rest of the mucinous tumors. CLDN3 is a robust and stable separator, because it would still separate the normal and mucinous from the other histotypes if the cutoff value were slightly adjusted. The next node demonstrates the ability of vascular endothelial growth factor (*VEGF*) to separate the remaining mucinous samples from the normal samples. *B*, recursive descent partitioning tree using the 11 selected known markers. The first node displays HE4 separating all of the serous (*S*), 89% of the endometrioid (*E*), 43% of the clear cell (*C*), and only 2 of the mucinous (*M*) tumors from the remainder of mucinous tumors and all of the normal (*N*) samples. CA125 separates the remaining endometrioid and mucinous tumors from the remaining clear cell tumors and normal samples. Finally, MUC1 separates the last clear cell tumors from all of the normal samples.

DISCUSSION

The use of gene expression microarrays permitted us to identify those genes that are highly overexpressed in ovarian cancers when compared with normal ovarian surface epithelium. Because the goal of our study was to identify potential markers

Table 3 Fold change for each of the 11 previously known markers in each tumor sample with fold changes that are 2 or greater in bold type

Sample	CA 125 AI923224	MUC 1 X52229	HE4 X63187	Osteopontin J04765	Kallikrein 6 U62801	Kallikrein 10 AF055481	Mesothelin U40434	MMP2 AA528255	Tetranectin X64559	Prostasin L41351	Inhibin M13981
OV1059	0.9	**2.4**	**4.6**	**2.6**	**3.0**	1.8	1.1	0.5	**2.4**	1.4	1.2
OV1140	1.4	**4.4**	**3.7**	**6.8**	**2.3**	**2.6**	1.0	0.7	1.5	1.1	1.0
OV134	1.4	**6.3**	**4.2**	**3.6**	1.0	0.7	0.8	0.5	0.9	1.0	1.0
OV234	0.7	**7.7**	**5.4**	**11.1**	**3.1**	0.6	0.5	0.4	**3.1**	1.0	1.3
OV632	**12.8**	**7.8**	**5.7**	**5.8**	1.4	1.9	1.5	0.6	1.1	1.2	1.2
OV702	1.7	**2.3**	**5.7**	1.7	1.2	0.6	1.2	0.5	1.2	1.4	1.0
OV746	0.7	**5.6**	**6.5**	**8.4**	0.9	0.4	0.5	0.7	0.8	1.1	0.8
OV76	1.7	**2.8**	**3.8**	**4.0**	**2.7**	**2.2**	**2.2**	0.5	1.2	1.3	1.0
OV979	1.5	**6.8**	**5.5**	**5.3**	1.5	1.1	**2.0**	0.5	0.9	1.0	0.9
MDA1	**2.7**	**6.0**	**9.8**	**4.5**	1.2	1.0	**3.0**	0.7	0.9	1.1	1.2
MDA2	**3.9**	**3.9**	**7.0**	**3.5**	1.1	0.8	**2.2**	0.9	1.1	1.2	0.8
MDA4	**7.3**	**11.3**	**4.9**	**6.0**	**5.4**	**2.6**	1.3	0.6	0.9	1.0	0.9
MDA5	**3.2**	**6.4**	**5.6**	**4.9**	1.6	0.8	1.5	0.6	0.8	1.0	0.9
MDA6	**8.7**	**3.8**	**2.1**	**8.1**	1.5	1.8	1.5	0.8	1.0	1.0	0.9
MDA7	**2.3**	**4.4**	**3.0**	1.9	**2.8**	0.7	1.6	0.5	0.9	0.9	1.0
MDA8	1.1	**5.1**	**5.5**	**2.7**	1.0	0.5	1.1	0.6	1.0	1.0	0.9
MDA9	**3.3**	**9.0**	**6.7**	1.4	**2.5**	**4.3**	1.4	0.4	1.0	0.9	1.2
ENDO1389	1.1	**9.2**	**6.6**	**10.9**	1.2	1.1	**2.3**	0.6	0.9	0.9	0.9
ENDO1462	**2.5**	**10.6**	**4.6**	**6.6**	**2.1**	**4.3**	0.9	0.5	1.0	1.3	0.9
ENDO1720	1.2	**4.9**	**5.5**	**3.6**	1.3	0.7	0.5	0.8	1.3	1.2	1.2
ENDO1931	1.7	**2.9**	**3.0**	**4.2**	0.9	0.8	0.5	**2.1**	1.0	1.3	0.9
ENDO1956	**2.4**	**2.5**	**4.8**	**4.4**	**2.5**	0.9	1.4	0.4	1.1	1.0	1.1
ENDO1999	1.6	**4.7**	**6.3**	**2.6**	1.3	0.6	0.9	0.4	1.1	1.1	1.1
ENDO2522	1.2	**8.2**	**4.8**	**13.1**	1.2	0.8	0.6	**2.5**	0.8	1.2	0.9
ENDO2977	0.7	0.9	0.7	**2.5**	0.9	0.5	0.5	0.8	1.0	1.1	1.0
ENDO3240	**2.3**	**5.7**	**2.8**	**12.8**	1.3	**2.8**	0.9	0.6	1.0	0.8	0.9
CC2251	1.0	**2.6**	1.5	**3.6**	**2.3**	0.5	1.6	0.4	1.3	1.4	1.0
CC2287	**2.9**	1.8	**4.5**	**7.6**	**7.9**	**3.7**	**7.3**	**2.4**	1.3	**8.6**	**4.6**
CC2420	1.6	1.9	0.9	**34.7**	**2.4**	0.6	0.8	1.0	1.1	1.2	1.0
CC2732	1.5	**2.5**	1.7	**26.8**	1.4	0.7	0.5	0.4	1.4	1.3	1.4
CC2893	0.9	**8.9**	**4.3**	**9.8**	0.8	0.4	0.7	0.8	1.0	1.3	1.0
CC2898	0.7	**3.5**	1.9	**3.2**	1.1	0.4	0.4	0.7	1.0	1.2	0.9
CC2953	**2.8**	**10.1**	**4.5**	**3.5**	**3.4**	**3.2**	1.5	1.1	1.0	1.7	1.0
MUC1726	0.6	**7.0**	0.8	0.3	0.9	0.5	0.4	0.5	0.9	1.0	0.9
MUC1865	0.7	1.3	0.9	**9.7**	0.9	0.5	0.8	**2.9**	1.5	1.3	0.9
MUC1884	0.5	**9.1**	1.5	1.4	0.8	0.8	0.8	0.5	1.0	1.0	0.8
MUC2014	0.7	0.9	0.8	**4.6**	0.9	0.4	0.9	**2.4**	**2.0**	1.1	0.8
MUC2708	0.6	**2.5**	1.0	**7.4**	1.2	0.9	0.7	0.9	1.1	1.0	0.9
MUC2790	0.6	**6.7**	0.9	**7.5**	1.1	0.8	0.5	0.8	1.5	1.1	0.8
MUC3215	**4.0**	**3.1**	**4.3**	**3.5**	0.9	0.5	1.0	0.6	0.8	1.3	0.8
MUC815	0.6	**2.1**	1.1	**5.3**	1.1	0.8	0.6	1.0	1.1	1.2	0.9
MUC897	**5.0**	**3.4**	**3.5**	**2.0**	**2.1**	**2.0**	0.8	0.6	1.3	1.2	0.9

for the early detection of epithelial ovarian cancer, we specifically chose a set of samples that would represent the histological heterogeneity of the disease. We included tumors of serous, mucinous, clear cell, and endometrioid histotypes. In addition, we compared these tumor samples with pooled scrapings of normal ovarian surface epithelium, rather than to whole ovaries (12), short-term cultures (13, 14), or immortalized ovarian surface epithelium (15).

We identified 86 genes that were at least 3-fold up-regulated in cancers when compared with normal epithelial scrapings. We also found HE4 (WAP four-disulfide core domain 2; Refs. 12–14), CD24 antigen (12), ceruloplasmin (14), claudin 3 (14), claudin 4 (14), enolase 1 α (12), eyes absent homologue 2 (16), karyopherin α2 (12), mammaglobin 2 (16), mucin 1 transmembrane (12–14), preferentially expressed antigen in melanoma (17), and tumor-associated calcium signal transducer 1 (12) to be highly expressed in ovarian tumors as compared with normal ovarian epithelium. The concordance of results by several groups of investigators is reassuring.

Four genes (NOTCH3, E2F3, RACGAP1, and HN1) separated all 42 cancers from the five pools of normal ovarian epithelial cells with mean fold changes of 3.6 (lower bound = 2.9), 3.1 (lower bound = 2.3), 4.6 (lower bound = 3.7), and 3.8 (lower bound = 3.0), respectively. RACGAP1 has been shown to be down-regulated by estrogen (18), providing a plausible connection to the estrogen deficiency associated with postmenopausal women and ovarian cancer. Semiquantitative RT-PCR of these genes performed on the same samples confirmed the elevated levels of expression detected on the microarrays. When the individual expression values in the tumor and normal samples were examined for each gene, as seen in Fig. 1, the separation between them is small. Therefore, although a perfect separation was achieved with each of these four genes, the distance between tumor and normal may not be sufficiently great for clinical use. Whereas the mean fold change is significant, individual tumor values may overlap with normal values resulting in increased sensitivity but decreased specificity. Use of multiple markers in combination might increase specificity if

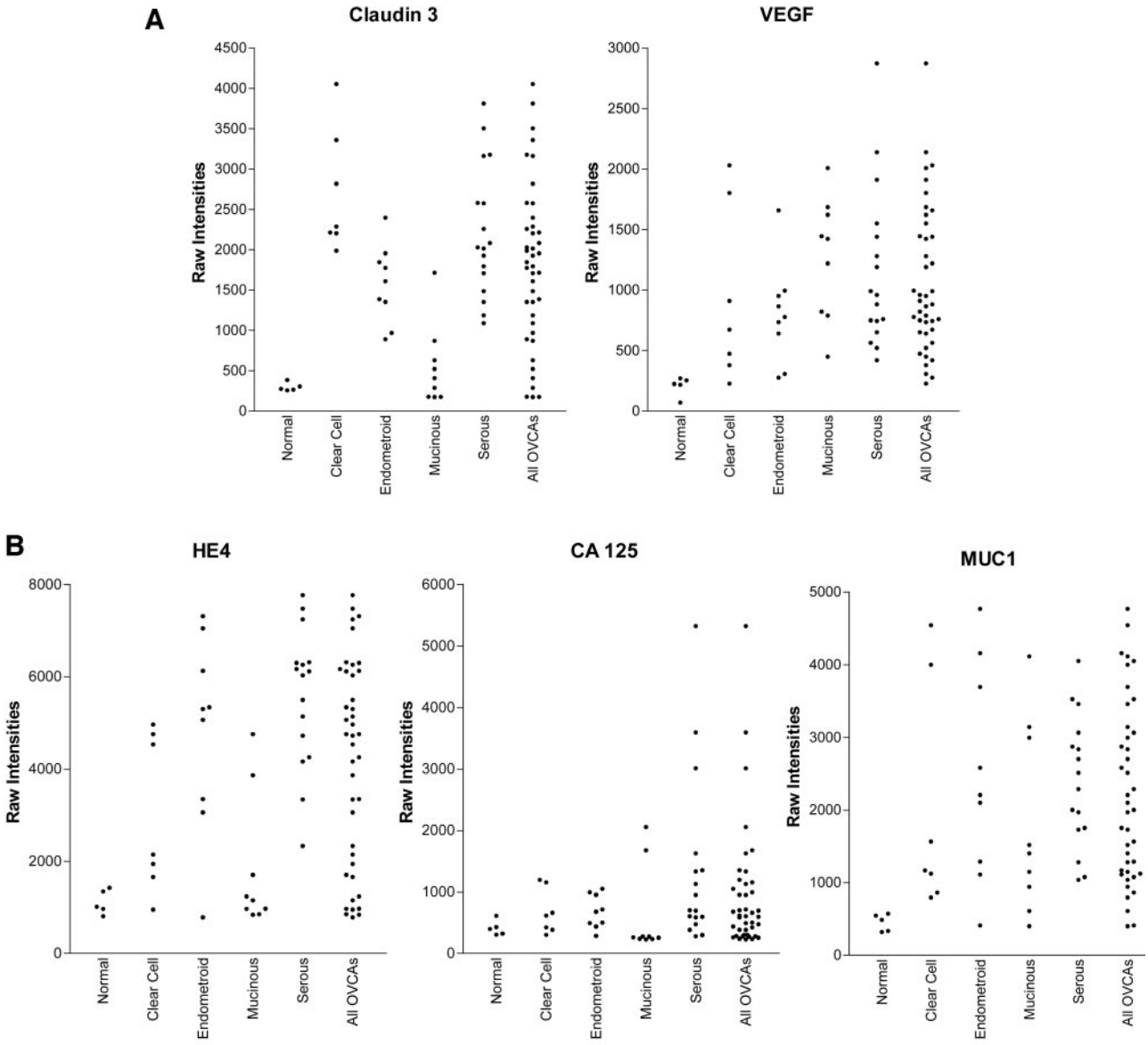

Fig. 3 Scatterplot of microarray data for vascular endothelial growth factor (*VEGF*) and CLDN3 (*Claudin 3*; *A*) and for HE4, CA 125, and MUC1 (*B*). Data are shown separated by histotype and combined ovarian carcinomas (*OVCAs*) for all four genes.

the markers could be captured using mathematical algorithms that optimize specificity. Furthermore, whereas the markers were selected for an ability to distinguish ovarian cancer from normal ovarian epithelium, the specificity may be decreased by the expression of these genes in other normal tissues or cancer cell lineages.

RDPA was applied to the microarray data to identify combinations of genes that would classify ovarian cancers of all histologies. The combination of two genes, CLDN3 (fold change = 6.3, lower bound = 4.3) and VEGF (fold change = 5.2, lower bound = 3.8), distinguished all of the ovarian tumors from normal surface epithelium. Semiquantitative RT-PCR of CLDN3 and VEGF performed on the same samples confirmed the elevated expression levels found on the microarrays. There was a wide separation of CLDN3 values in serous, endometrioid, and clear cell tumors as compared with normal. When immunohistochemical staining was performed on a larger set of tumors, elevated protein expression of CLDN3 and VEGF were again confirmed. CLDN3 is a member of a large family of integral membrane proteins important for tight junction formation and function. CLDN3 and CLDN4 have been shown to be highly expressed in ovarian cancer as well as other epithelial cancers. A recent report has demonstrated increased cytoplasmic staining of CLDN3 in ovarian cancers, suggesting a possible role as a signaling molecule in addition to its role regulating tight junction permeability (19). At present, we are determining whether significant amounts of CLDN3 are shed into the sera.

We also performed RDPA on a set of 11 previously re-

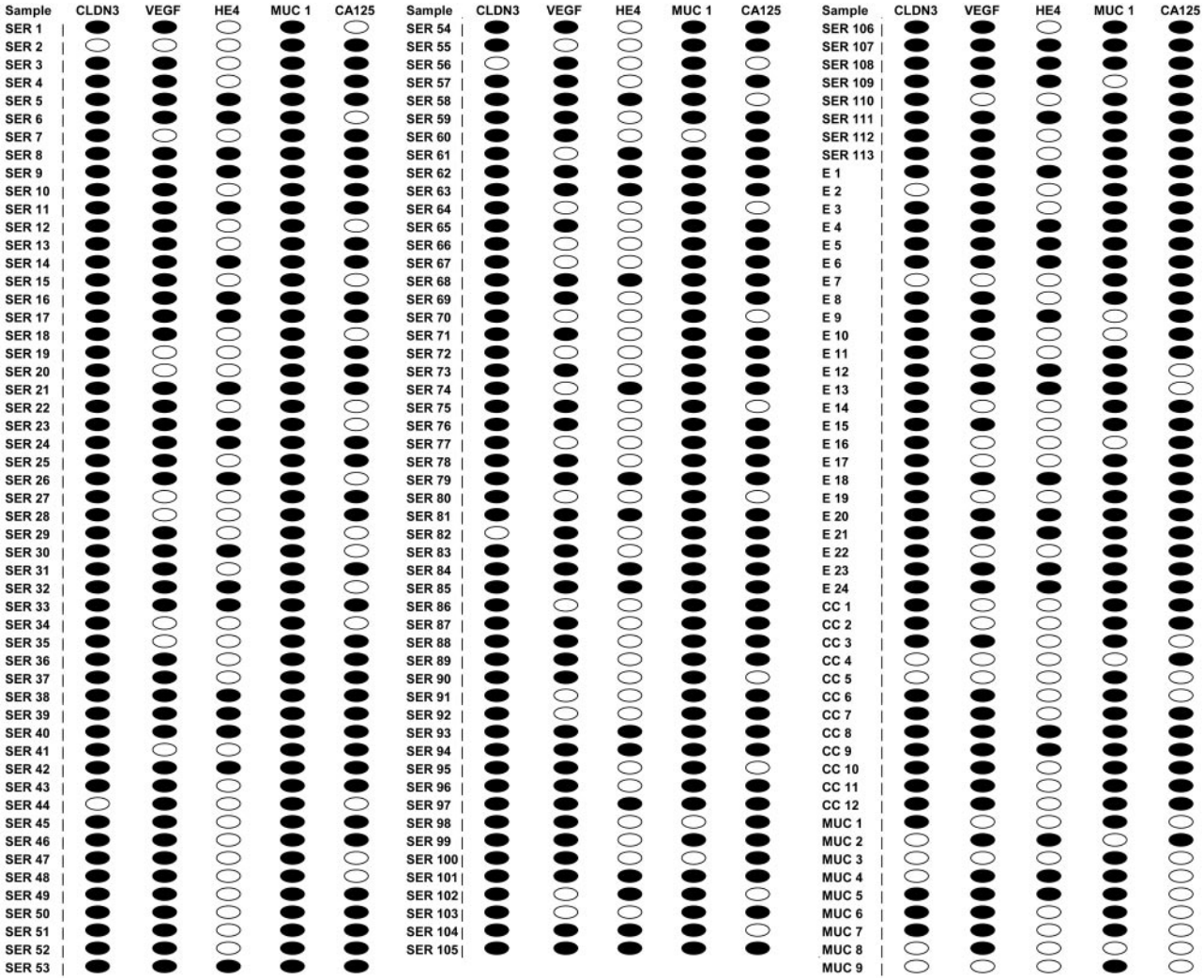

Fig. 4 Immunohistochemical staining results for claudin 3 (*CLDN3*), vascular endothelial growth factor (*VEGF*), WAP four-disulfide core domain 2 (*HE4*), mucin 1, transmembrane (*MUC1*), and mucin 16 (*CA125*) protein expression in ovarian cancer tumors. Representation of protein expression across 158 ovarian cancer samples. *SER*, serous histotype; *E*, endometrioid histotype; *CC*, clear cell histotype; *MUC*, mucinous histotype. *Open oval*, scoring of 0 to 1 plus; *filled oval*, scoring of 2 to 3 plus.

ported potential tumor markers to identify a set of markers that would identify all histologies of ovarian cancer from normal. HE4 (FC = 3.9, LB = 3.0), CA125 (FC = 2.2, LB = 1.5), and MUC1 (FC = 5.1, LB = 3.9) in combination separated the tumors from normal. From Table 4, values of HE4 in serous tumors and endometrioid tumors show a significant separation from normal. Whereas RDPA was able to separate each of the histologies of ovarian cancers from normal, it was interesting to note that the mucinous tumors should be separated from other histologies because of low CA125 values; this was also confirmed by immunohistochemistry and semiquantitative RT-PCR. This is consistent with previous reports that indicate that

Table 4 Immunohistochemical staining results for five genes[a]

Histotype	*CLDN3*		*VEGF*		*HE4*		*MUC1*		*CA 125*	
Serous	109/113	92%	88/113	78%	41/113	36%	109/113	92%	88/113	78%
Endometroid	22/24	92%	17/24	71%	12/24	50%	21/24	88%	22/24	92%
Clear cell	10/12	83%	8/12	67%	2/12	17%	11/12	92%	9/12	75%
Mucinous	4/9	44%	6/9	56%	3/9	33%	7/9	78%	1/9	11%
Total	145/158	92%	119/158	75%	58/158	37%	148/158	94%	120/158	76%

[a] Percentage of protein expression.

CA125 is poorly expressed by many mucinous cancers. Immunohistochemical staining of ovarian cancer samples using the HE4 antibody was suboptimal, with only 36% of serous tumors and 50% of endometrioid tumors showing positive HE4 staining. This may relate to loss of antigenic activity during fixation and embedding, as Hellstrom *et al.* recently reported a serum assay for HE4 using this same antibody to detect antigen in sera from women with ovarian cancer. Using sera from 37 ovarian cancer patients, they found that HE4 demonstrated similar specificity and sensitivity to CA-125, with fewer false positive values for nonmalignant ovarian disease (11). Recent studies have also identified HE4 to be overexpressed in some breast tumors (20). HE4 is an anti-proteinase that was initially identified in the male epididymis (21). The highly elevated expression levels seen in serous and endometrioid ovarian tumors in our study and other studies (13, 14), as well as the relative specificity to ovarian cancer, makes HE4 a leading candidate for a tumor marker.

One critical requirement for an effective tumor marker is its presence in early-stage disease. One of the main shortcomings of CA125 is that 50% of stage 1 ovarian cancers do not have an elevated CA125. Our previous study has shown that a subset of genes that are overexpressed in late-stage serous cancers are also overexpressed in early-stage serous cancers (22). In our present study, we included nine early-stage serous cancers and eight late-stage serous cancers. According to microarray data, all of the candidate markers except CA125 showed elevated expression levels in both early- and late-stage serous tumors.

Ultimately, serum validation of these markers in women with early-stage disease will be necessary to determine optimal marker candidates. It is not known whether the genes identified in the microarray analysis are released into the circulation, which is a limitation of our study. In addition, presence of protein in ovarian cancer, as detected by immunohistochemistry, may not be accompanied by presence of the protein in the circulation. Genes that may not show striking overexpression of transcript may have significant serum levels. Using too stringent microarray criteria for choosing candidate genes may exclude these candidates. For example, whereas mesothelin showed low expression values in our study, Scholler *et al.* (23) have shown that mesothelin can be detected in serum of ovarian cancer patients. Finally, the markers presented here may identify genes overexpressed in solid tumors and may not be ovarian cancer-specific. We believe our study is a first step toward identifying new and complementary markers for ovarian cancer. Clearly, further work for individual genes will be necessary.

In conclusion, we believe that it is unlikely that a single marker for epithelial ovarian cancer will be clinically useful, given the biological heterogeneity of the disease. Our data, however, suggest that a limited number of markers in combination might identify all ovarian cancers. However, specificity may be decreased by the expression of these genes in other normal tissues or cancer cell lineages. For effective screening with serum tumor markers, expression of protein must also be accompanied by appropriate release into the circulation. Work is currently on-going to develop assays with the candidate genes and combinations identified in this study.

REFERENCES

1. Greenlee RT, Murray T, Bolden S, Wing PA. Cancer statistics, 2000. CA Cancer J Clin 2000;50:7–33.

2. American Cancer Society. Cancer Facts and Figures. Atlanta, GA: American Cancer Society; 2003.

3. Fishman DA, Bozorgi K. The scientific basis of early detection of epithelial ovarian cancer early detection program (NOCEDP). In: Stack MS, Fishman DA, editors. Ovarian cancer. Boston, MA: Kluwer Academic Publishers; 2002. p. 3–28.

4. Schwartz DR, Wu R, Kardia SL, et al. Novel candidate targets of beta-catenin/T-cell factor signaling identified by gene expression profiling of ovarian endometrioid adenocarcinomas. Cancer Res 2003;63:2913–22.

5. Schwartz DR, Kardia SL, Shedden KA, et al. Gene expression in ovarian cancer reflects both morphology and biological behavior, distinguishing clear cell from other poor-prognosis ovarian carcinomas. Cancer Res 2002;62:4722–9.

6. Bast RC Jr, Urban N, Shridhar V, et al. Early detection of ovarian cancer: promise and reality. In: Stack, MS, Fishman DA, editors. Ovarian cancer. Boston, MA: Kluwer Academic Publishers; 2002. p. 61–97.

7. Li C, Wong WH. Model-based analysis of oligonucleotide arrays: expression index computation and outlier detection. Proc Natl Acad Sci USA 2002;98:31–6.

8. Hastie T, Tibshirani R, Friedman J. The elements of statistical learning. New York: Springer; 2001.

9. Venables WN, Ripley BD. Modern applied statistics with S-Plus. New York: Springer-Verlag; 1994.

10. Chambers JM, Hastie TJ. Statistical models in S. Pacific Grove; CA: S. Wadsworth and Brooks, 1992.

11. Hellstrom I, Raycraft J, Hayden-Ledbetter M, et al. The HE4 (WFDC2) protein is a biomarker for ovarian carcinoma. Cancer Res 2003;63:3695–700.

12. Welsh JB, Zarrinkar PP, Sapinoso LM, et al. Analysis of gene expression profiles in normal and neoplastic ovarian tissue samples identifies candidate molecular markers of epithelial ovarian cancer. Proc Natl Acad Sci USA 2001;98(3):1176–81.

13. Schummer M, Ng WV, Bumgarner RE, et al. Comparative hybridization of an array of 21,500 ovarian cDNAs for the discovery of genes overexpressed in ovarian carcinomas. Gene (Amst) 2001;238(2):375–85.

14. Hough CD, Sherman-Baust CA, Pizer ES, et al. Large-scale serial analysis of gene expression reveals genes differentially expressed in ovarian cancer. Cancer Res 2000;60(22):6281–7.

15. Mok SC, Chao J, Skates S, et al. Prostasin, a potential serum marker for ovarian cancer: identification through microarray technology. J Natl Cancer Inst (Bethesda) 2001;93(19):1458–64.

16. Su AI, Welsh JB, Sapinoso LM, et al. Molecular classification of human carcinomas by use of gene expression signatures Cancer Res 2001;61(20):7388–93.

17. Giordano TJ, Shedden KA, Schwartz DR, et al. Organ-specific molecular classification of primary lung, colon, and ovarian adenocarcinomas using gene expression profiles. Am J Pathol 2001;159(4):1231–8.

18. Laufs U, Adam O, Strehlow K, et al. Down-regulation of Rac-1 GTPase by estrogen. J Biol Chem 2003;278(8):5956–62.

19. Rangel LB, Agarwal R, D'Souza T, et al. Tight junction proteins claudin-3 and claudin-4 are frequently overexpressed in ovarian cancer but not in ovarian cystadenomas. Clin Cancer Res 2003;9(7):2567–75.

20. Perou CM, Sorlie T, Eisen MB, et al. Molecular portraits of human breast tumours. Nature (Lond) 2000;406(6797):747–52.

21. Kirchhoff C, Habben I, Ivell R, Krull N. A major human epididymis-specific cDNA encodes a protein with sequence homology to extracellular proteinase inhibitors. Biol Reprod 1991;45(2):350–7.

22. Shridhar V, Lee J, Pandita A, et al. Genetic analysis of early- versus late-stage ovarian tumors. Cancer Res 2001;61(15):5895–904.

23. Scholler N, Fu N, Yang Y, et al. Soluble member(s) of the mesothelin/megakaryocyte potentiating factor family are detectable in sera from patients with ovarian carcinoma. Proc Natl Acad Sci USA 1999; 96(20):11531–6.

Clinical Cancer Research

Selection of Potential Markers for Epithelial Ovarian Cancer with Gene Expression Arrays and Recursive Descent Partition Analysis

Karen H. Lu, Andrea P. Patterson, Lin Wang, et al.

Clin Cancer Res 2004;10:3291-3300.