Comparative hybridization of an array of 21 500 ovarian cDNAs for the discovery of genes overexpressed in ovarian carcinomas

Michel Schummer a,*, WaiLap V. Ng a, Roger E. Bumgarner a, Peter S. Nelson a, Bernhard Schummer b, David W. Bednarshi a, Laurie Hassell a, Rae Lynn Baldwin c, Beth Y. Karlan c, Leroy Hood a

a Department of Molecular Biotechnology, University of Washington, Box 357730, Seattle, WA 98195, USA
b Institut für Pharmakologie und Toxikologie, Fakultät für Klinische Medizin der Universität Heidelberg, Maxbacherstr. 14-16, 69110 Mannheim, Germany
c Department of Obstetrics and Gynecology, Cedars-Sinai Medical Center, University of California, Los Angeles, School of Medicine, Los Angeles, CA 90048, USA

Received 28 January 1999; received in revised form 2 July 1999; accepted 28 July 1999; Received by I. Verma

Abstract

Comparative hybridization of cDNA arrays is a powerful tool for the measurement of differences in gene expression between two or more tissues. We optimized this technique and employed it to discover genes with potential for the diagnosis of ovarian cancer. This cancer is rarely identified in time for a good prognosis after diagnosis. An array of 21 500 unknown ovarian cDNAs was hybridized with labeled first-strand cDNA from 10 ovarian tumors and six normal tissues. One hundred and thirty-four clones are overexpressed in at least five of the 10 tumors. These cDNAs were sequenced and compared to public sequence databases. One of these, the gene HE4, was found to be expressed primarily in some ovarian cancers, and is thus a potential marker of ovarian carcinoma. © 1999 Elsevier Science B.V. All rights reserved.

Keywords: Cancer maker, DNA array; Differential expression; HE4

1. Introduction

Ovarian cancer is the leading cause of gynecological cancer death in the United States. The American Cancer Society estimates that in 1998, some 25 400 women will develop ovarian cancer and 14 500 will die from it (American Cancer Society, 1998). The overall 5 year survival rate is about 46%, and has remained essentially unchanged for 25 years. Ovarian cancer is ranked fifth in cancer mortality among women, and raises concerns both with women and physicians because of its generally poor prognosis. Cancers diagnosed at an early stage have a 5 year survival rate of 92% in contrast to a 25% 5 year survival rate for patients with disseminated disease at diagnosis. Seventy-five per cent of epithelial ovarian cancers are diagnosed at advanced stages. This is in part due to the lack of symptoms early in the disease course, and the absence of a sensitive and specific screening test for early disease detection. Currently available ovarian cancer markers such as CA-125 are neither sensitive nor specific enough for population screening to detect early, treatable ovarian cancers (Jacobs et al., 1993).

We describe the use of ‘high-density cDNA array hybridization’ (HDAH) to identify transcripts that show high expression levels in ovarian cancer tissues as compared to ovarian surface epithelium (OSE). This technology has been used in a variety of experiments to identify transcripts (Schena et al., 1998), whose expression patterns differ in two tissues (e.g. normal and cancer). Our objective is to find (1) transcripts that are overexpressed in tumor as contrasted with normal ovarian tissue and (2) cDNAs encoding proteins that could be useful diagnostic markers (e.g. secreted or cell-surface pro-
teins). Two general types of assays are possible: (1) protein assays for secreted proteins or on the surface of cells that metastasize into the circulation, and (2) PCR assays from genes uniquely expressed in blood-borne (or ascites-borne) tumor cells. Hybridizing 21,500 randomly selected cDNAs from normal and neoplastic ovarian tissues with probes from 10 ovarian tumor and six normal tissues, we identified 134 clones with higher expression signals in ovarian tumors as opposed to normal tissues. These clones were sequenced, and in some cases, their expression pattern was confirmed by RT-PCR and Northern blot analysis. The expression pattern of one of these clones, HE4, suggests that it may be a potential candidate diagnostic marker for ovarian cancer.

2. Materials and methods

2.1. Tissues and cells

We used the following tissues for our experiments: ovarian surface epithelium short-term culture (Karlan et al., 1995), early passages (OSE); normal ovary consisting of primarily stromal cells (N002, N005, N006, N019 and N035); two benign ovarian tumors (T017B, an endometrioid polyp, and T018B, a serous cystadenoma); one borderline early stage serous carcinoma, LMP (T028L); late-stage, high-grade papillary serous ovarian adenocarcinomas (T001–T006, T008–T011, T014–T016 and T021); two early-stage ovarian adenocarcinomas (one serous: T007 and one mucinous: T037); one late-stage, high-grade serous ovarian adenocarcinoma post-chemotherapy (T012); two late-stage, high-grade serous ovarian adenocarcinoma with massive metastases (T013M and T026M); peripheral blood lymphocytes (PB1L and PB2L); Fetal ovaries: pool of 25 fetal ovaries (52–103 days); bone marrow, cerebellum, kidney, liver and placenta (Clontech, Palo Alto, CA).

In order to minimize the effect of variance in tissue collection on the RNA quality and hence the hybridization patterns, we ensured that tissue collection would adhere to the following guidelines. After surgery, a tissue section was taken for the pathologist’s examination and an adjacent section was snap-frozen in liquid nitrogen.

All ovarian tumor tissue specimens were examined for their tumor cell content (which was above 80%) and the absence of necrosis. RNA preparations of all tissues or at increasing stringency (20 min, 2 × SSC, 0.5% SDS, 65 °C; 20 min, 0.5 × SSC, 0.5% SDS, 65 °C; 2 × 20 min, 0.2 × SSC, 0.5% SDS, 65 °C).

2.2. Minipreparation of 21,500 ovarian clones

Five cDNA libraries were created from ovarian tissues and cell cultures (OSE, T007, T008, T010 and T012) using the ZAP-cDNA synthesis kit (Stratagene). Examining the cDNA clones using PCR, the insert sizes were found to average between 1.2 and 1.5 kb. From each library, 96 clones were randomly chosen, sequenced and analyzed by similarity analysis against the non-redundant and EST database. The low number of mitochondrial and ribosomal sequences, the limited number of clones with no insert, and the significant cDNA diversity indicated that the libraries were of high quality. Using a 96-deep-well plate-based minipreparation assay (Ng et al., 1996), we picked 21,500 transformants (8600 from the OSE cDNA library and 3225 each from the four tumor cDNA libraries), extracted the cDNAs and transferred them to 384-well microtiter plates.

2.3. Dotting the 21,500 clones onto nylon membranes

Using a hand-held arraying tool with a 384-pin printhead developed in our laboratory (Schummer et al., 1997), we dotted the 21,500 cDNAs onto 16 sets of 14 nylon membranes of 7.5 × 12 cm, which held each of the 1536 clones. The cDNA was denatured and immobilized on the membrane as previously described (Schummer et al., 1997).

2.4. Labeling and hybridization protocol

Each set of membranes was hybridized with a complex probe consisting of 32P-labeled first-strand cDNA. Briefly, 5 μg of poly(A)+ RNA or 30 μg of total RNA were reverse-transcribed using Superscript II reverse transcriptase (Life Technologies) and oligo-dT12, primers with 30 μCi of alpha-32P-dCTP (3000 Ci/mmol) and unlabeled dATP, dGTP, dTTP at 1 mM each; after 20 min, unlabeled dCTP was added to a final concentration of 1 mM, and the reaction was continued for another 40 min. This unpurified probe was hybridized to 12 membranes under conditions described previously (Schummer et al., 1997). The membranes were washed at increasing stringency (20 min, 2 × SSC, 0.5% SDS, RT; 20 min 0.5 × SSC, 0.5% SDS, 65 °C; 2 × 20 min, 0.2 × SSC, 0.5% SDS, 65 °C).

2.5. Software for spot detection

After hybridization and washing, the membranes were exposed to a phosphor storage screen, and the hybridization patterns were captured as 16-bit TIFF.
images using a PhosphorImager (Molecular Dynamics, Sunnyvale, CA). Nine nylon membranes were imaged simultaneously on a 35 × 45 cm screen. The resulting file was processed using a software package developed in our laboratory. The TIFF image was split into nine smaller images, each representing one of the arrayed membranes. Briefly, the user defined the outer dimensions of each membrane by placing a cursor into each of the nine array images. Subsequently, the computer superimposed a grid, approximating the positions of the 1536 dots. By five passes of center-of-mass finding, the computer determined the exact center of each of the 1536 dots. It integrated the area of an experimentally determined number of pixels around each center that covered the area of the largest hybridization signal present on the membranes. The intensities of all pixels in the area were integrated. Local background was calculated by choosing one pixel with the lowest intensity out of four pixels situated halfway between one dot and its four diagonal neighbors. Both values were stored in a tab-delimited text file together with the coordinates of the spot on the array.

2.6. Single pass 5' sequencing, database analysis and sequence comparison

Sequencing was performed on plasmid DNA and PCR products using previously described methods (Ng et al., 1996). The single-pass sequences were edited to remove vector and poly(A) sequences. Edited sequences were compared with those in the EST (dBEST) and non-redundant nucleotide and protein databases (GenBank) at the National Center for Biotechnology Information (NCBI) using the Baylor College of Medicine Search Launcher batch client server 'Search Tissues comprise many different cell populations. Each type of cell in a tissue exhibits its particular gene expression pattern. Since most ovarian tumors arise from epithelial cells, the comparison of tumors against ovarian surface epithelium should provide a useful comparison. Two qualifications must be made: (1) ovarian surface epithelial cells in a short-term culture will probably have some differences in expression patterns from in-vivo ovarian epithelial cells, and (2) tumors may have intermixed normal cells from the ovary. In order to detect genes that are overexpressed in one cell type or tissue versus another, one needs to know the limitations of the detection system, notably (1) the upper and lower limits of detection (signal-to-noise ratio) which — translated into the number of mRNA molecules detectable per cell — should be suitable for the proposed study, and (2) the measured level of variation in signal intensity on identical membranes interrogated with identical probes. The latter will determine a factor above which overexpression can be regarded as significant.

2.8. Northern blot

A HE4 PCR product of 500 bp was cloned into a pCR2.1 vector using the TA cloning kit (Invitrogen, San Diego, CA). A digoxigenin-labeled riboprobe was prepared from this vector using a Genius RNA DIG labeling kit (Boehringer Mannheim, Germany). The probe was hybridized overnight at 68°C in DIG Easy Hyb buffer and washed in 2 × SSC, 0.1% SDS for 15 min at room temperature; 2 × SSC, 0.1% SDS for 20 min at 68°C; and 0.1 × SSC, 0.1% SDS for 2 × 15 min at 68°C. The hybridized RNA was visualized using the DIG detection kit (Boehringer Mannheim), and the membrane was exposed to X-ray film for 15 min.

3. Results and discussion

Tissues comprise many different cell populations. Each type of cell in a tissue exhibits its particular gene expression pattern. Since most ovarian tumors arise from epithelial cells, the comparison of tumors against ovarian surface epithelium should provide a useful comparison. Two qualifications must be made: (1) ovarian surface epithelial cells in a short-term culture will probably have some differences in expression patterns from in-vivo ovarian epithelial cells, and (2) tumors may have intermixed normal cells from the ovary. In order to detect genes that are overexpressed in one cell type or tissue versus another, one needs to know the limitations of the detection system, notably (1) the upper and lower limits of detection (signal-to-noise ratio) which — translated into the number of mRNA molecules detectable per cell — should be suitable for the proposed study, and (2) the measured level of variation in signal intensity on identical membranes interrogated with identical probes. The latter will determine a factor above which overexpression can be regarded as significant.
3.1.1. Determination of detection limit and dynamic range

The sensitivity of the array technology determines the number of detectable mRNA molecules in a cell. In order to determine the mean signal-to-noise ratio, we hybridized 14 identical arrays containing 1536 identical cDNAs coding for the green fluorescent protein (GFP) with first-strand cDNA probes made from human liver poly(A)⁺ RNA in which a GFP mRNA was added in decreasing concentrations (14 different concentrations ranging from one transcript in 200 to one in 200 000). As depicted in Fig. 1, the probe with the highest GFP concentration yielded a mean value of 8300 ± 416 dpm (decays per minute) per pixel, and the mean background value was determined as 90 ± 18 dpm/pixel. With background subtraction, this represents a dynamic range of 456 (background-subtracted signal divided by background fluctuation: 8210/18 = 456) or 2.5 orders of magnitude. We established a lower limit of sensitivity of 1 GFP RNA in 20 000 liver RNAs, a result similar to those in other studies (Pietu et al., 1996). Based on an estimated 10⁵–10¹⁰ transcripts per average eukaryotic cell (Bishop et al., 1974), the membrane-based HDASH can detect a minimum of between five and 50 mRNA molecules in a cell and a maximum of 500–5000. The lower limit falls in the low to medium class of transcripts, and the upper limit lies in the highly expressed gene class (Zhang et al., 1997). This detection range should be sufficient for the identification of overexpressed genes.

3.1.2. Normalization of the hybridization signals

In order to compare hybridization signatures of two identical membranes that have been hybridized with different probes in two separate incubations, one needs to normalize the signals to a standard. Although we adhere to a strict protocol, slight variations can be introduced by minute differences in probe labeling, probe purification, hybridization and wash conditions and exposure time. We normalized the background-subtracted intensities of one membrane by setting the median to 1. Assuming that among the 1536 clones present on one membrane, the majority does not alter its expression (Zhang et al., 1997), we believe that this is justified.

3.1.3. Determination of variation in signal intensity

Two factors influence the accuracy of the hybridization detection for one particular cDNA on a membrane: the amount of cDNA on the membrane (governed by the dotting procedure) and the amount of labeled cDNA that remains bound to the target cDNA on the membrane after hybridization (governed by the efficiency of the probe labeling reaction and the hybridization and washing kinetics). We determined the variation of amounts of DNA spotted by our arraying tool to be ±14% (data not shown). Since the probe consists of a complex mixture of cDNAs, the arrayed DNA is in vast excess of the probe cDNA, and thus the variations caused by the spotted cDNA can be regarded as negligible. In order to assess the probe-to-probe variance, we hybridized four replica membranes containing 1536 ovarian cDNAs with four ³²P-labeled first-strand cDNA probes independently generated from one batch of total RNA prepared from liver tissue. We compared the background-subtracted intensities of one cDNA across the four membranes and calculated the standard deviation, thus generating 1536 values. We ranked the clones by their expression and determined three means of standard deviations, one for the upper, the middle and the lower third, corresponding roughly to the high, medium high and low, expression categories of transcripts. The mean of the standard deviations amounted to ±15%, ±24% and ±40% respectively, which averages to ±26% for all clones. Using the following equation, we calculated the threshold value for a ratio to be regarded as significant: \[1 + \text{standard deviation} \times 10\%\]. In order to be above this threshold of significance, a highly expressed gene needs to display a ratio of 1.35, a medium expressed gene a value of 1.63 and the least expressed gene a value of 2.33. These measurements would suggest that a threshold of significance, which is a function of intensity, should be used and that the threshold will vary from 1.35 for the most highly expressed genes to 2.33 for the least expressed genes. However, the measurements performed here are at best a surrogate system for estimating...
error in the tumor data, i.e. the above experiments control for hybridization, filter and analysis variation but do not control for labeling and other sample-handling variation in the tumor samples. With limited tissue available for each tumor, it is not possible to perform replicate measurements on all our samples to generate similar significance curves for the actual data. Hence, we chose to use a ratio of 2.5 or more as the threshold of significance for our tumor data. We recognize that this criterion will result in the exclusion of genes that are differentially regulated at a statistically level. However, given that our goal is to develop genes that may serve as serum markers for ovarian cancer, and given the limitations of currently available assay systems for serum marker testing, a factor of 2.5 differential expression is appropriate.

3.2. Screening of 21,500 ovarian clones

An ideal array of cDNAs would contain a single copy of every gene expressed by the tissues to be compared. Since the identification of all human genes is incomplete, we chose to array randomly selected cDNAs derived from a wide spectrum of ovarian tissues including normal ovarian epithelium, early stage ovarian carcinomas, and late-stage pathologically aggressive ovarian carcinomas. We chose to array 8600 clones in form of purified plasmids from an OSE library [short-term culture of ovarian surface epithelial cells (Karlan et al., 1995)], and 3225 each from four ovarian cancer cDNA libraries from increasing malignancy, totaling 21,500 arrayed clones. We created 16 replicate sets of these arrays, each set consisting of 14 membranes of 7 × 12 cm holding 1336 clones. Each of the membrane sets was hybridized with a 32P-labeled first-strand cDNA probe made from the RNA of an early-stage serous ovarian tumor (T007), eight late-stage serous ovarian tumors (T004, T006, T009, T010, T011, T014, T015, T016), one recurrent ovarian tumor (T012), ovarian surface epithelium (N001S), liver, placenta, bone marrow, cerebellum, and kidney. Two types of comparative experiments were carried out: (1) normal and tumor ovarian tissues were contrasted, and (2) ovarian tissues were compared against a variety of normal tissues. The first comparisons would reveal the tumor-specific cDNAs and the second the ovarian-specific cDNAs (at least with respect to the five different normal tissues). It was not our purpose to analyze early-to-late stage differences or tumor stratification as the limited number of cancerous tissues would not allow this. Our objective was to determine whether it is possible to use this technique to detect genes that are overexpressed in ovarian carcinomas respective to normal ovary and other tissues.

3.3. Differential transcript expression

Using the spot-finding and detection software developed in our laboratory, we determined the hybridization intensities for each clone and calculated their ratios. Comparing the 10 hybridizations with ovarian tumor tissues to those with OSE, the vast majority (>93%) of the clones displayed tumor-to-OSE ratios of less than a factor 2.5, and therefore were considered unchanged; about 7% of the clones exhibited a tumor-to-OSE ratio of more than 2.5, 0.9% a ratio of greater than 5.0, and 0.5% a ratio of greater than 10.0. Thus, most transcripts were expressed at similar levels in normal and tumor tissues, a finding that has been reported in colorectal and pancreatic cancers (Zhang et al., 1997).

No clone exhibited a 2.5-fold difference in expression in more than six of the ovarian tumors relative to OSE. Given the difference in tumor stages (one was an early stage tumor, and one a recurrent late stage tumor, the rest being late-stage ovarian adenocarcinomas) and the fact that the same stages, if they represent different stratified types, do not necessarily reflect high degrees of similarity on the molecular level, given the inter- and intra-tissue heterogeneity (possible proximity of section to areas of necrosis, differences in histology and pathology between tumors and across tumor sample), we did not expect to see a particular clone exhibit high tumor- to-OSE ratios in all tumors. Sixteen clones showed overexpression in at least six ovarian cancers, but 14 of these 16 were also expressed in at least one non-ovarian tissue. In order to obtain a reasonable number of clones with overexpression in ovarian tumors and not in non-ovary tissues, we chose clones that fulfilled the following criteria: ratios greater than 2.5 in at least five out of the 10 tumors compared to OSE, and ratios below 2.5 in bone marrow, cerebellum, kidney, liver, and placenta compared to OSE. We were able to identify 134 clones that fulfilled these criteria. Sequencing of the partial cDNA clones revealed 60 that matched sequences in the non-redundant (nr) GenBank database. Of these, 17 matched to mitochondrial and ribosomal genes, and 43 matched to 37 other characterized genes (Table 1). Forty-seven clones matched only to sequences in the EST database, and 24 clones did not match any sequence in GenBank and were classified as novel. Three clones of 254, 312 and 323 bp length matched entirely to SINE and LINE sequences and were thus classified as repeats (see Table 1).

The expression patterns of two of these clones, which code for S-adenosyl homocysteine hydrolase and HE4, are shown in Fig. 2. For both genes, the calculated overexpression by signal intensities in the cancer tissues can be confirmed by visual inspection of the hybridized membranes. It is obvious, however, that by visual inspection alone, these clones would have probably escaped our scrutiny since their expression is rather weak compared to neighboring clones. The overexpression of the 17 clones with similarity to mitochondrial sequences and ribosomal proteins can...
be attributed to the higher metabolic activity of the tumors. Ribosomal protein sequences have been found to be more highly expressed in colon carcinomas (Pogue-Geile et al., 1991). Likewise, five other genes linked to metabolic pathways such as elongation factor 1 gamma and initiation factor 4A1 were overexpressed in ovarian cancer tissues. It is notable that these 22 clones displayed an average tumor-to-OSE ratio of 5.22 ± 2.4, whereas the remaining 38 clones with homology to known genes had a lower average ratio of 4.11 ± 1.8. This underscores the fact that the degree of overexpression alone is not necessarily indicative of a clone that can be used as a marker protein.

In order to estimate the quality of the HDAH in identifying cancer related genes, and since we were realistically capable of processing only a limited number of clones, we focused on the 43 previously characterized clones, as opposed to the 47 clones that match only ESTs or those 24 that do not match any sequence in GenBank. Of the 43 clones with homology to the 37 characterized genes, 10 genes are expressed in epithelial tissues: 14.3.3, B44, HD3, Muc1, Ornithine decarboxylase, Progesterone binding protein, Rhod, and Ryuvocin (GenBank Accession Nos listed in Table 1). This coincides with the fact that the vast majority of ovarian cancers, including all those used for HDAH, arise from the ovarian surface epithelium (Berchuck et al., 1996).

Thirteen of the 37 genes (35%) are known to be overexpressed in various cancers, including lung, breast and colon. Six of these 13 are expressed in ovarian carcinomas, their expression not being restricted to ovarian tissues. The thirteen genes are 14.3.3 [lung cancer (Nakanishi et al., 1997)], beta-actin [AML (Blomberg et al., 1987) and colorectal carcinomas (Naylor et al., 1992)], B44 [breast cancer (Couto et al., 1996)], CD44 [ovarian cancer cell lines (Stickeler et al., 1997)], Clip-Restin [Hodgkin disease and anaplastic large-cell lymphoma (Delabie et al., 1992)], Collagen COL1A2 [ovarian cystadenoma (Kauppila et al., 1996)], E16 [colorectal carcinoma, adenocarcinomas from breast and endometrium (Wolf et al., 1996)], Insulin-like growth factor BP 3 [breast cancer (Ng et al., 1998)], Muc1 [epithelial ovarian cancer (Dong et al., 1997)], Procollagen-alpha [ovarian cystadenocarcinoma (Kauppila et al., 1996)], putative Progesterone binding protein [ovarian cancer (Isola et al., 1990)], RhoA proto oncogene [ras activation (Khoshavi-Far et al., 1995)], and MDC15, a metalloproteinase [some metalloproteinases are elevated in ovarian tumor cell cultures (Fishman et al., 1997)].
is indeed capable of narrowing down the pool of 21,500 randomly selected clones to a few epithelium- and cancer-related genes.

3.4. Confirmation of overexpression of four selected clones by RT-PCR-based transcript quantitation

Any clone with its expression restricted to ovarian carcinomas can be potentially used as a marker without knowing its function. Early detection of ovarian cancer, however, requires that the assay be suitable for routine screening of women, which means that it must be affordable, non-invasive and with a high degree of specificity. Only a serum-based assay can deliver this. Therefore, knowing whether a protein is secreted or membrane-bound maximizes the chance that the protein or its degradation product will be found in the blood either as freely circulating protein or bound to the membrane of a cell that has detached from the tumor. In both cases, an antibody can be used to detect the protein in the blood. A circulating cancer cell can be detected by an RT-PCR assay or fluorescence-activated cell sorting.

In an attempt to find out whether one of the 43 clones that match characterized genes would be a potential candidate for a marker protein in a serum-based assay, we examined which of the clones codes for a cell surface protein such as Her2/neu, used as a target in breast cancer treatment (Baselga et al., 1998) or a secreted protein such as Prostate Specific Antigen (PSA) which is used in prostate cancer diagnosis (Rittenhouse et al., 1998).

From the 43 clones with homology to the 37 known genes, we chose five that are expressed at the cell surface (progesterone binding protein, ryudocan, mucin1, E16, BA46) and one which is secreted (HE4). In addition, we included the gene 14-3-3, which is expressed in the cytoplasm but which, like HE4, appeared twice in our selected clones list. Beta actin is often used as a control for quantitative analyses because of its assumed uniformity in expression in a large array of tissues. Our HDAH results suggest, however, that beta actin is differentially expressed in some ovarian tumors. We therefore chose to verify beta actin expression as well. The characteristics of the eight chosen genes are summarized in Table 2. We used RT-PCR-based transcript quantitation to confirm overexpression in tumors relative to normal tissues.

Due to the small size of our tumor specimens (ranging from 200 to 400 mg per tissue), the RNA preparations used in the array hybridization were exhausted during library construction and probe preparation. Therefore, new ovarian adenocarcinomas matching the stage and grade of the original tumors were used for the RT-PCR analysis. We chose one early-stage, low-grade mucinous ovarian adenocarcinoma (T037) five late-stage, high-grade serous ovarian adenocarcinomas (T001-T006 and T021) and two metastatic ovarian serous adenocarcinomas (T013M and T026M). In order to incorporate different tumor histologies, we included two benign ovarian tissues (T017B and T018B) as well as a borderline ovarian tumor tissue (T028L). In addition, we tested the expression in four normal ovaries (N002, N005, N006 and N019), in a pool of fetal ovaries and in two batches of peripheral blood lymphocytes (PBL1 and PBL2).

The reason for analyzing the expression patterns of these genes in peripheral blood lymphocytes is to determine whether they are expressed in blood elements, for if they are, they would not be good candidates for a diagnostic probe in blood samples. The OSE, as well as the liver and placental tissue were the same as used for array hybridization. As a control for the quality of the RNA template, we included a gene that we found to be expressed at high levels in all tissues tested so far, S31iiif25 (GenBank Accession No. U61734, Trower et al., 1996). Fig. 3 shows the results of the RT-PCR. The quantitated intensities of the PCR bands are summarized in Table 2. While trying to match the tumor tissues in stage and grade, we did not expect an exact reproduction of the ratios from the HDAH analyses. In spite of these shortcomings, we were able to reproduce the tumor-to-OSE ratios observed in the HDAH for seven out of the eight genes, albeit only qualitatively. For the gene 14-3-3, the tumor-to-OSE ratios were low but still measurable. This discrepancy can be attributed to the difference in tumor samples used or to an erroneous reading of the HDAH signals. For three genes (BA46, E16 and ryudocan), a high placenta-to-OSE ratio stands in discordance with the HDAH results where they had been low. Since the placental RNA used in both cases was the same, and since our quadruple RT-PCR approach is more accurate than the HDAH method, we must conclude that in the HDAH, the placental values must have been misread for these three clones.

14-3-3 shows no tumor-to-OSE ratios above the threshold of significance of 2.5. It displays a mean ratio of 1.5 in four invasive and in one benign ovarian tumor, which does not compare well with the mean ratio of 4.4 determined in the HDAH.

BA46 shows tumor-to-OSE ratios above 2.5 in five tumors but also in one normal ovary and in placenta. In spite of its low expression in PBL (which, as noted in the beginning of this section, is a prerequisite for a serum marker), the relatively low mean ratios in RT-PCR and HDAH of 3.2 make it a second choice marker gene.

Beta actin shows tumor-to-OSE ratios above 2.5 in 10 out of the 12 tumors (a mean of 3.9 compared to 4.4 in the HDAH), but also in some normal tissues, including PBL. Although these numbers do not warrant the consideration as a tumor marker gene, they give cause to question the use of beta actin as a normalization standard.
Table 2  
HDAH (top) and RT-PCR ratios (bottom) of nine selected genes

| Gene name | Accession No. | Protein | T004 | T007 early | T008 | T009 | T010 | T011 | T012 recur | T013M | T014 | T015 | T016 | Liver | Placenta | PBL1 | PBL2 | Fetal | N002 | N004 | N006 | N019 | N035 | T017B | T018B | T028L | T037 early | T002 | T003 | T005 | T006 | T001 | T021 | T013M | T026M |
|-----------|---------------|---------|------|------------|------|------|------|------|-----------|-------|------|------|------|-------|---------|------|------|-------|------|------|------|------|------|-------|------|------|-------|------|------|------|------|------|------|------|------|------|
| 14.3.3    | X56468        | Cytopl. | 3.1  | 5.1        | 5.1  | 5.1  | 5.5  | 5.1  | 5.1       | 3.1   | 3.1  | 3.1  | 3.1  | 3.1   | 3.1     | 3.1  | 3.1  | 3.1   | 3.1  | 3.1  | 3.1  | 3.1  | 3.1  | 3.1   | 3.1  | 3.1  | 3.1   | 3.1  | 3.1  |
| 14.3.3    | X56468        | Cytopl. | 3.1  | 5.1        | 5.1  | 5.1  | 5.5  | 5.1  | 5.1       | 3.1   | 3.1  | 3.1  | 3.1  | 3.1   | 3.1     | 3.1  | 3.1  | 3.1   | 3.1  | 3.1  | 3.1  | 3.1  | 3.1  | 3.1   | 3.1  | 3.1  | 3.1   | 3.1  | 3.1  |
| β-actin   | X00351        | Membr.  | 5.4  | 8.5        | 8.5  | 8.5  | 8.5  | 8.5  | 8.5       | 8.5   | 8.5  | 8.5  | 8.5  | 8.5   | 8.5     | 8.5  | 8.5  | 8.5   | 8.5  | 8.5  | 8.5  | 8.5  | 8.5  | 8.5   | 8.5  | 8.5  | 8.5   | 8.5  | 8.5  |
| E16       | M01044        | Membr.  | 3.1  | 3.1        | 3.1  | 3.1  | 3.1  | 3.1  | 3.1       | 3.1   | 3.1  | 3.1  | 3.1  | 3.1   | 3.1     | 3.1  | 3.1  | 3.1   | 3.1  | 3.1  | 3.1  | 3.1  | 3.1  | 3.1   | 3.1  | 3.1  | 3.1   | 3.1  | 3.1  |
| HE4       | X63187        | Secreted| 3.1  | 3.1        | 3.1  | 3.1  | 3.1  | 3.1  | 3.1       | 3.1   | 3.1  | 3.1  | 3.1  | 3.1   | 3.1     | 3.1  | 3.1  | 3.1   | 3.1  | 3.1  | 3.1  | 3.1  | 3.1  | 3.1   | 3.1  | 3.1  | 3.1   | 3.1  | 3.1  |
| HE4       | X63187        | Secreted| 3.1  | 3.1        | 3.1  | 3.1  | 3.1  | 3.1  | 3.1       | 3.1   | 3.1  | 3.1  | 3.1  | 3.1   | 3.1     | 3.1  | 3.1  | 3.1   | 3.1  | 3.1  | 3.1  | 3.1  | 3.1  | 3.1   | 3.1  | 3.1  | 3.1   | 3.1  | 3.1  |
| Mucin1    | X55229        | Membr.  | 3.1  | 3.1        | 3.1  | 3.1  | 3.1  | 3.1  | 3.1       | 3.1   | 3.1  | 3.1  | 3.1  | 3.1   | 3.1     | 3.1  | 3.1  | 3.1   | 3.1  | 3.1  | 3.1  | 3.1  | 3.1  | 3.1   | 3.1  | 3.1  | 3.1   | 3.1  | 3.1  |
| ProgBP    | Y12711        | Membr.  | 3.1  | 3.1        | 3.1  | 3.1  | 3.1  | 3.1  | 3.1       | 3.1   | 3.1  | 3.1  | 3.1  | 3.1   | 3.1     | 3.1  | 3.1  | 3.1   | 3.1  | 3.1  | 3.1  | 3.1  | 3.1  | 3.1   | 3.1  | 3.1  | 3.1   | 3.1  | 3.1  |
| Ryu       | D31292        | Membr.  | 3.1  | 3.1        | 3.1  | 3.1  | 3.1  | 3.1  | 3.1       | 3.1   | 3.1  | 3.1  | 3.1  | 3.1   | 3.1     | 3.1  | 3.1  | 3.1   | 3.1  | 3.1  | 3.1  | 3.1  | 3.1  | 3.1   | 3.1  | 3.1  | 3.1   | 3.1  | 3.1  |

* Eight genes out of the 43 clones that match to 37 known genes were validated for their expression by RT-PCR (see Fig. 3). The values of the PCR bands were calculated using the software QuantityOne (BioRad, Hercules, CA). Titan RT-PCR amplifies the template semiquantitatively; therefore, the numbers in this table are merely indicative of a tendency and cannot be translated into copy numbers. The rows show the gene name, GenBank Accession No., protein localization, 10 tumor-to-OSE ratios that were observed in the HDAH (only ratios above 2.5; normal-to-OSE ratios are omitted for they lied all below 2.5), followed by 22 tissue-to-OSE ratios determined in the RT-PCR (for clarity, only ratios above 1 are displayed). The columns are duplicated for 14.3.3 and HE4 because two clones were selected for them by HDAH. Putative Progesterone binding protein (ProgBP): progesterone binding proteins can be found in low-grade breast cancers and in some ovarian cancer cell lines. The homologous rat sequence has a transmembrane region (Falkenstein et al., 1996), indicating that our clone might also be membrane-bound. Ruddy (abbreviated as Ryu.) is a cell-surface proteoglycan with a transmembrane domain; it is expressed in an extensive array of human tissues (Kojima et al., 1993). HE4 is an epidermal, epididymis-specific protease inhibitor that is thought to be involved in the maturation of spermatozoa (Keirnollf et al., 1991). HE4 is a cell-surface protein expressed in human breast carcinomas. It has been used successfully as a target for experimental breast cancer radioimmunotherapy (Costo et al., 1996). Beta actin is a cytoskeletal protein with differential expression in acute myeloid leukemia (Blomberg et al., 1987) and high expression in colorectal carcinomas (Naylor et al., 1992). E16 codes for an integral membrane protein that was isolated from peripheral blood lymphocytes (Gaugitsch et al., 1992). It is expressed in colorectal and other human carcinomas (Wolf et al., 1996).

E16 shows tumor-to-OSE ratios above 2.5 in three tumors (with a mean of 5.3 compared to a mean of 4.4 in the HDAH). It also shows high ratios for two normal ovaries and placenta. The low expression in PBL and the high average ratios for the tumors make it a possible marker candidate.

HE4 shows a clear tumor-restricted expression, making its pattern resemble that in the HDAH. Most importantly, the results suggest that it is not expressed in peripheral blood lymphocytes. As noted in the beginning of this section, this accordingly represents a candidate for a serum marker assay. The difference in the mean rates of overexpression measured by RT-PCR (11 ×) and HDAH (4.1 ×) can be attributed either to...
Fig. 3. Expression monitoring by RT-PCR. Eight genes (plus one control) are tested in 23 tissues. Tissue names are on top; OSE is marked with an arrow. Tissues starting with an N are normal ovaries, and those starting with a T are ovarian tumors. The S31iii125 gene serves as a control. The number of PCR cycles is indicated behind each gene name (on the right). ProgBP stands for ‘putative progesterone binding protein’. The PCR bands are in the range of 420–660 bp. All reactions had been performed in four parallel sets with one set shown here.

the better signal-to-noise ratio in the RT-PCR or to the different tumor samples used.

Mucin 1 shows a high RT-PCR value in fetal ovaries, suggesting that this might be a fetal gene that is re-expressed in the tumor. It shows strong bands in three out of the 12 tumors, two of them metastatic and one an early stage tumor, resulting in a mean tumor-to-OSE ratio of 3.0. This result correlates with that of the array hybridization.

The Putative progesterone binding protein shows a high tumor-to-OSE ratio for only two tumors, one being similar in stage to a tumor used in the HDAH. All other tumors show medium high ratios but so do the normal tissues, including the PBL. The strong expression in the metastasizing tumor may indicate a role as a marker for tumor staging, prognosis or stratification.

The transcript of ryudocan displays a similar pattern of expression as HE4, and the mean the tumor-to-OSE ratio of 4.3 are is slightly higher than the one determined by HDAH (where it was 6). The presence of ryudocan mRNA in liver, PBL and placenta means that the protein might normally be found in the blood, thus making it a less suitable marker candidate.

3.5. Confirmation of overexpression of HE4 by Northern blot analysis

Of the eight genes tested in the RT-PCR, only HE4 shows a clear tumor-restricted expression pattern. To further confirm the cancer-restricted expression of HE4, we used a Northern blot (Northern Territory®).

Invitrogen, San Diego, CA) that contained total RNA from ovaries from four patients who had unilateral ovarian cancer. RNA from both the affected and the unaffected ovary was present on the blot (loaded adjacent to each other). Fig. 4 shows that HE4 is expressed in two ovarian carcinomas but not in the matching normal ovaries. HE4 cannot be detected in the tumors nor in the normal ovaries of two other patients. The ratios of HE4 expression between the unaffected and
the affected ovary was 6.1 for patient 1 and 4.5 for patient 2. Thus, HE4 is also a candidate for a tumor-staging, prognosis or stratification marker.

3.6. Conclusion

From the 21,500 clones, we chose 43 that were overexpressed in ovarian tumors by HDAH with homology to characterized genes. We chose eight genes for expression validation by RT-PCR. From these eight, seven genes displayed tumor-to-OSE ratios similar to those measured in the HDAH, albeit with different tumor tissues matching grade and stage. Seven of these eight display expression in normal tissues; only HE4 showed a clear tumor-restricted expression pattern. We conclude that the HE4 message is significantly overexpressed in a variety of ovarian tumors relative to normal tissues or OSE, thus making it a potential candidate for a marker protein.

The results support the validity of using HDAH combined with a second quantitation method for the identification of genes that are overexpressed in cancers as compared to normal tissues. We are preparing an antibody against HE4 to further analyze whether it indeed could be a diagnostic marker for ovarian cancer.

Acknowledgements

This work was supported by the Stowers Institute and by grants from the Deutsche Forschungsgemeinschaft, the National Institutes of Health (5RO1HG01713-02) the Marsha Rivkin Center for Ovarian Cancer Research and the National Science Foundation (BIR9214821/9423347). We would like to thank the members of the sequencing group in our laboratory.

References


Pfitz, G., Alborn, O., Gschwind, V., Lamy, B., Bots, F., Leroy, E.,...


