Microarray Analysis Reveals Distinct Gene Expression Profiles among Different Histologic Types of Endometrial Cancer

John I. Risinger, G. Larry Maxwell, G. V. R. Chandramouli, Amir Jazaeri, Olga Aprelikova, Tricia Patterson, Andrew Berchuck, and J. Carl Barrett

Laboratory of Biosystems and Cancer, National Cancer Institute, Bethesda, Maryland 20892 [J. J. R., G. L. M., G. V. R. C., A. J., O. A., T. P., J. C. B.]; Walter Reed Army Medical Center, Washington, DC 20013 [G. L. M.]; Department of Obstetrics and Gynecology/Division of Gynecologic Oncology, Duke University, Durham, North Carolina 27710 [A. B.]

Abstract

Previous studies of oncogene and tumor suppressor gene alterations have suggested that differences exist in the molecular pathogenesis of the various histological types of endometrial cancer. To elucidate further the molecular events involved in endometrial carcinogenesis, we examined global expression patterns of 16 nonendometrioid cancers (13 serous papillary and 3 clear cell), 19 endometrioid cancers, and 7 age-matched normal endometria using cDNA microarrays. Unsupervised analysis of gene expression identified 191 genes that exhibited >2-fold differences (P < 0.001) between the histological groups. Many genes were similarly dysregulated in both nonendometrioid and endometrioid cancers relative to normal endometria. Gene expression differences in only 24 transcripts could distinguish serous from endometrioid cancers, the two most common subgroups. These data provide the basis for investigation of previously unrecognized novel pathways involved in the development of endometrial cancers.

Introduction

Two subtypes of endometrial carcinoma have been described based on both clinical and histopathologic variables (1). Type I endometrioid cancers account for the majority of cases, and these cancers are usually well differentiated and E2 in histology. These cancers are associated frequently with a history of unopposed estrogen exposure or other hyperestrogenic risk factors, such as obesity. Patients, with type I endometrial cancer, typically have early stage disease and a favorable prognosis with appropriate therapy. In contrast, Type II endometrioid cancers are often poorly differentiated, non-E, and are not associated with hyperestrogenic factors. These cancers are more likely to be metastatic at presentation and often recur despite aggressive clinical interventions.

Molecular genetic evidence indicates that endometrial carcinoma likely develops as the result of a multistep process of oncogene activation and tumor suppressor gene inactivation (2). Our group and others have described some of these changes and demonstrated that these molecular alterations appear to be specific for Type I (E) and Type II (non-E) cancers. Type I cancers are characterized by mutation of PTEN (3–5), KRAS2 (6, 7), and CTNNB1 (8, 9) defects in DNA mismatch repair (Ref. 10; as evidenced by the microsatellite instability phenotype) and a near diploid karyotype (11). Type II cancers often contain mutations of TP53 (12–14) and Her-2/neu (15) and are usually nondiploid (11).

The molecular pathogenesis of endometrial cancer is incompletely understood. Although alterations in several genes noted above have been described, none are present in the majority of cases. In addition, some endometrial cancers lack evidence of alterations in any of these genes. In a recent survey of our databases, we found that 44 of 87 endometrial cancers (50%) were lacking mutations in PTEN, TP53, CTNNB1, or the microsatellite instability phenotype, although about half of these cancers was advanced stage. This suggests the existence of unrecognized pathways that can lead to the development of endometrial cancer. To elucidate further the molecular pathogenesis of endometrial cancers, we have used cDNA microarrays to examine patterns of gene expression in E and non-E cancers and normal endometrium.

Materials and Methods

Specimens and Nucleic Acid Isolation

Flash frozen endometrial cancers were obtained from patients undergoing hysterectomy at Duke University Medical Center. None of the patients had received preoperative chemotherapy or radiation. In addition, samples of normal endometrium (N) were obtained from patients undergoing hysterectomy for benign gynecological diseases. Tissues were obtained at Duke with Institutional Review Board-approved informed consent, and this study was approved by the NCI Institutional Review Board. Endometrial cancers were examined by a gynecologic pathologist to confirm the histologies as being PS, E, or CC. Tissue samples were subjected to RNA isolation using TRizol and an additional purification using the RNeasy Kit (Qiagen, Valencia, CA) following the manufacturer’s recommendation. We examined 19 histologically normal endometria that were age matched to the cancers and able to isolate a sufficient quantity of dissected glandular epithelium in only seven that were used in this study. After isolation of RNA, the integrity of each RNA sample was verified by denaturing gel electrophoresis.

Microarray Processing

Total RNA was amplified linearly with a modification of the Eberwine method. Briefly, total RNA was reverse transcribed by using a 63 nucleotide synthetic primer containing the T7 RNA polymerase binding site, 5'-GGCCAGTGAATTGTAATACGACTCACTATAGGGAGGCGG(T)24-3'. Second-strand cDNA synthesis (producing double-stranded cDNA) was performed with RNase H, E. coli DNA polymerase I, and E. coli DNA ligase (Invitrogen, Carlsbad, CA). After cDNA was made blunt ended with T4 DNA polymerase (Invitrogen), it was purified by extraction with a mixture of phenol, chloroform, and isomyl alcohol and by precipitation in the presence of ammonium acetate and ethanol. The double-stranded cDNA was then transcribed with T7 RNA polymerase (T7 Megascript kit; Ambion, Austin, TX), yielding linearly amplified antisense RNA, which was purified with RNeasy mini columns (Qiagen). The cDNA microarray chips contained 9984 total features representing the human GEM2 set of clones (InCyte) and were manufactured at the NCI microarray facility. Four micrograms of amplified RNA were reverse transcribed and directly labeled with cyanine 5-conjugated dUTP (endometrial RNA) or cyanine 3-conjugated dUTP (Stratagene Universal Reference RNA). Hybridization was performed in 5 × saline sodium citrate.
containing 25% formamide for 14–16h at 42°C. Slides were washed, dried, and scanned. A detailed protocol for RNA amplification, cDNA probe labeling, and hybridization is available on the Internet. Genepix software (Axon Instruments, Inc., Union City, CA) was used to analyze the raw data, which were then uploaded to a relational database maintained by the Center for Information Technology at the NIH (Bethesda, MD).

**Data Analysis**

The microarrays were scanned on an Axon Instruments Genepix 4000A scanner (Axon Instruments, Foster City, CA) at wavelengths 635 and 532 nm for Cy5 and Cy3 dyes, respectively, to obtain images of 10-μm resolution. The quantification of spot intensities, qualities, and local background was performed automatically by Genepix software using spot diameter in the range 70–180 μm and a manual supervision for any inaccuracies in the automatic spot detection. Local background correction was applied to spot intensities before the calculation of expression ratios. The spots having a minimum signal level of 250 counts and well defined with ≥70% of the pixels above SD were used for expression analysis. The expressions were normalized by median centering the logarithmic expression ratios of Cy3 and Cy5 signals within each array. The logarithmic ratio versus logarithmic signal scatters and Cy5 signal versus Cy3 signal scatters did not indicate any abnormalities in distributions. Statistical analysis was performed using Biometric Research Branch Array Tools software (NCI) using logarithmic values of expression ratios to the base 2.

A set of 9431 cDNA clones with good spot quality from a total of 9984 spotted on microarray was used to study differential gene expression. The significance of expression differences between tissues was computed by F tests at a two-tailed significance of expression differences between tissues was computed by F tests and hybridization is available on the Internet.³ GenePix software (Axon Instruments, Inc., Union City, CA) was used to analyze the raw data, which were then uploaded to a relational database maintained by the Center for Information Technology at the NIH (Bethesda, MD).

**Hierarchical Clustering**. The set of genes differentially expressed > 2-fold was clustered by the similarity of their expression profiles. Hierarchical clustering was performed on logarithmic values of expressions using 1-ρ as distance metric, where ρ is the correlation coefficient of any two gene expressions (16). The expression data were shown relative to average normal endometrial expression. The cluster is color coded using red for up-regulation from normal endometria and green for down-regulation.

**Multidimensional Scaling**. The similarities of gene expression profiles of samples were studied by a multidimensional scaling procedure (Young and Hamer details available on the Internet).⁴ Each sample was considered as an n-dimensional vector where each gene represents a dimension, and n is the number of genes. The dissimilarities of samples were computed using 1-ρ as distance metric, where ρ is the correlation coefficient of the two samples. Each sample is represented as a coordinate in a three-dimensional space as shown in Fig. 1. Samples with similar gene expressions on microarray are placed at closer proximity compared with the dissimilar ones.

**Classification**. Binary class comparison and prediction was performed on PS and E pairs. The genes distinguishing PS and E were selected by F test (P < 0.001). The class prediction was performed by computing compound covariate of gene expressions (17). The compound covariate is defined as \( \Sigma \), where \( t_i \times (x_{ij} - m_i) \), where \( x_{ij} \) is the expression of gene \( i \) in sample \( j \), \( m_i \) is the midpoint of two classes for gene \( i \), \( t_i \) is the t score, and the summation is overall genes selected for classification. The prediction was performed by leaving one sample out at a time for cross-validation and using all other samples for classification. The cross-validated misclassification error rate was estimated by 10,000 random permutations of class labels as < 0.01%. The class prediction using 232 differentially expressed genes between PS and E could successfully predict 94% of the samples.

**Data availability**. The complete dataset can be accessed on the Internet.⁵

---

⁴ Internet address: http://forrest.psych.unc.edu/teaching/p208a/mds/mds.html.
⁵ Internet address: http://home.ccr.cancer.gov/risingerdata1102.
Quantitative Real-time PCR

The relative expression of genes shown in Fig. 2 was determined as follows. The concentrations of genes PEG3, STAT2, REV3L, FOXO1A, MLLT7, and glyceraldehyde-3-phosphate dehydrogenase for all samples were determined using the standard curve method for normalization. Sequences for primers and probes are available on request. The gene expressions were then compared with average threshold PCR amplification cycle time of normal endometrial samples. Fig. 2 shows relative gene expressions (on logarithmic scale to base 2) compared with normal endometria. The SE values are shown as error bars.

Results and Discussion

The endometrium is one of the most dynamic tissues present in placental mammals and poses interesting questions regarding experimental approaches with gene expression array technologies, because levels of ovarian steroids would be predicted to drastically affect gene expression. Previous global expression profile analysis has examined the normal endometria of normal cycling women and in one case compared these data to a limited set of E endometrial cancers (18–20). We chose to examine a set of cancers and normal tissues matched for age and typical of the median age of onset for endometrial cancers, which is generally after the period of normal cycling. We analyzed the RNA from 42 endometrial samples and hybridized them with a universal reference RNA against a cDNA gene chip with 9984 features, most of which are known genes. RNA preparation and processing of arrays were stringently controlled for quality and reproducibility. After array processing and scanning, 9431 features provided gene expression data sufficient for analysis (expression was present in either the endometrial or universal RNA).

Initially, to seek similarities of expression profiles, we classified the 35 endometrial cancers and 7 normal endometria using unsupervised multidimensional scaling. Multidimensional scaling using all of the 9431 genes on arrays revealed distinctively separated clusters for PS, CC, E, and normal endometria, indicating consistent global expression changes for each type of sample. This analysis indicated that gene
expression patterns between samples were somewhat variable, as expected, but differences were sufficient to cluster the majority of cases into their respective subgroups (Fig. 1). These data are in accordance with the differing morphological characteristics of these samples yet still show the heterogeneity predicted of any cancer. Somewhat surprisingly, two PS cancers clustered close to the normal endometrial samples. Additionally, the distinct clustering of the three CC cancers indicates that their underlying biology may be distinct from the majority of serous papillary and E lesions.

Next, we compared the relative gene expressions based on the individual histology. The significance of expression differences of genes between samples was computed using F tests. The significant differential gene expression from normal endometrial tissue (N) to PS, E, or CC histology was separately determined with a two-tailed F test (P < 0.001). At this statistical significance level, 293 genes in PS, 281 genes in E, and 67 genes in CC were found to differ significantly from normal endometria. F tests comparing PS and E, PS and CC, and N and CC revealed 232, 16, and 113 differentially expressed genes at the same significance level. These data indicate the least significant differences in the expression profiles were between PS and CC, whereas the greatest differences were between PS and normal endometria.

To further understand the underlying biology of these groupings, the 191 genes that statistically differed between one or more of the four subgroups and exhibited at least a 2-fold gene expression difference were analyzed using hierarchical clustering (Fig. 2a). We grouped these and depicted the 20 most highly up-regulated and 20 most highly down-regulated genes (Fig. 2b). Many genes shown previously to be involved in carcinogenesis are present on these lists, providing some validity to our array analysis, including the platelet-derived growth factor-α and AXL genes, which were included in a previous array study (20). To further verify the quality of our array data, a subset of five genes (PEG3, STAT12, REV3L, FOXO1A, and MLLT7) was examined in all 42 tissue specimens using quantitative real-time PCR (TaqMan; Fig. 2c). These genes were chosen based on their universal down-regulation in both E and non-E cancers and on their potentially interesting roles in carcinogenesis.

Defects in DNA mismatch repair characterize only ~25% of E endometrial cancers (10), suggesting that other unrecognized DNA repair pathways are impaired in endometrial cancers. To examine this concept, we scanned our array data for genes involved in DNA repair processes. Microarray analysis indicated that most PS cancers under-expressed the DNA polymerase ζ catalytic subunit REV1L, as did a significant portion of the E lesions. Loss of this activity could result in a hypermutagenic phenotype as the result of defective translesion repair synthesis, a major route of DNA resynthesis after DNA repair. Real-time PCR analysis confirmed the down-regulation of this gene in endometrial cancers (Fig. 2c).

Endometrial cancers likely evolve in part as a result of many epigenetic defects, including loss of imprinting and promoter hypermethylation, e.g., most endometrial cancers with the microsatellite instability phenotype do not contain mutation of a mismatch repair gene but instead contain hypermethylated MLH1 promoter alleles that silence transcription from this locus (21, 22). Evidence for other hypermethylated loci in endometrial cancers involves the estrogen and progesterone receptors and the adenomatous polyposis coli tumor suppressor gene (23–25). Our array and a previous array study that included only E cancers (20) indicate many down-regulated genes, some of which may be silenced by similar epigenetic mechanisms. We chose to examine the PEG3 gene as one of these down-regulated transcripts in which to verify our array data. PEG3, a kruppel type zinc-finger transcription factor, is transcribed from a paternally imprinted locus on chromosome 19. PEG3 is down-regulated in gliomas by a hypermethylation mechanism (26), and reintroduction of PEG3 suppresses tumor formation in glioma cell lines (27). Our microarray and real-time PCR analysis show a dramatic reduction in PEG3 mRNA almost universally in endometrial cancers, perhaps indicating identification of a novel unrecognized pathway in endometrial carcinogenesis. Several genes like PEG3 are candidates for epigenetic down-regulation by promoter hypermethylation in endometrial cancers.

Several genes that may impact on a central signaling pathway involving the mammalian counterparts of the C-elegans insulin/IGF-I cell survival or longevity pathway were identified on our array. PTEN, an essential lipid phosphatase that has activity in many signal transduction pathways, is a key component of this pathway and is mutated in many E endometrial cancers (3–5). PTEN negatively regulates signals through its lipid phosphatase activity (28). We hypothesize that many risk factors for endometrial cancer act by elevating growth factors like estrogen and IGF-I, which serve to constitutively activate antiapoptosis pathways. Normally, negative regulators of these pathways like PTEN suppress carcinogenesis, and endometrial cancers can only arise clonally from rare cells that escape the tight regulation of these pathways either through gene mutations or epigenetic changes. Mutation of PTEN has been demonstrated in histologically normal and hyperplastic endometrium, suggesting that alteration of this tumor suppressor gene may be an early event in carcinogenesis of E endometrial cancer (29, 30).

Interestingly, array data indicated the down-regulation of two forkhead transcription factors, MLLT7 and FOXO1A, also known as AFX and FKHR. These transcripts represent two of the three mammalian homologues to the C-elegans DAF-16 forkhead transcription factor involved in insulin/IGF-I signaling and cell longevity. PTEN, a negative upstream regulator of this pathway, affects downstream signaling presumably through augmentation of cell survival and antiapoptosis programs. This activity is mediated in part through the phosphorylation of the v-AKT homologues, which are augmented in the absence of functional PTEN. Phosphorylated AKT acts on FKHRs by directing their ubiquitination in the cytoplasm and subsequently allowing the transcription of antiapoptotic genes. Interestingly, our array data indicate that these two forkhead transcripts are down-regulated in endometrial cancers, suggesting an additional mechanism by transcriptional repression. This observation would result in the diminution of the activity of DAF-16 homologues. The exact downstream changes that might result from this diminution and contribute to endometrial carcinogenesis are obscure. Independent real-time PCR analysis confirms the down-regulation of these genes in our cancer set and the need to further explore this observation (Fig. 2c).

Loss of tumor suppressor activities is a fundamental property of most cancer types. Our gene array analysis indicated a down-regulation of the STAT12 or SOCS2 transcript in endometrial cancers. SOCS2 possesses functional characteristics of a tumor suppressor gene. Specifically, SOCS2-deficient mice exhibit a giant phenotype as the result of nonabrogated growth (31). This phenotype is identical to mice given exogenous IGF-I or growth hormone. Subsequent studies based in part on this observation indicated that SOCS2 also directly interacts with the IGF receptor (32). This observation also potentially links this gene to the insulin survival pathway. Additionally, we also noted dysregulation of the YWHAZ (14–3–3 zeta) transcript in our array data. Limited evidence also implicates this gene to the insulin survival pathway by its association with IRS1 (33). Deregulation of growth factor pathways caused by inappropriate epithelial-stromal interactions have been proposed a potential mediator of endometrial carcinogenesis (34, 35). SOCS2 may also function in this capacity. Down-regulation of SOCS2, Forkhead transcription factors, and putative PTEN indicate significant convergence on the IGF-I cell survival pathway in this tumor type.

Determining the gene expression differences between E and serous cancers represents a potential step in determining the differences between these two most common histologies of endometrial cancer. A relatively few transcripts could distinguish between these two types of
endometrial cancer (at ≥2-fold difference in 24 genes and 1.5-fold difference in 75 genes). The entire list of 75 genes is depicted in Fig. 3, including the relative expression of all cancers in each histological group. We validated the expression of six of the transcripts, including the FOLR, dual specificity phosphatase 6, IGF-II, TFF3, AGR2, and ubiquitin COOH-terminal esterase-like 1. All six transcripts were analyzed by real-time PCR and confirmed our microarray analysis (Fig. 3b). Class prediction models could reproducibly define 30 of 32 cancers into either serous or E histology (data not shown). Interestingly, the two cancers that were not accurately classified in this model were the same two serous cancers that grouped with normal atrophic endometrium in multidimensional scaling analysis. These two cancers may represent a grouping of serous cancers that could be distinguished with a larger sample set or may represent anomalies because of some other undetermined factors. Despite these two samples, the successful analysis on such a limited number of samples suggests that class prediction models could be applied to future endometrial arrays with the idea of predicting clinical outcomes.

Quantitative PCR analyses are more reflective of the absolute level of gene expression differences than are cDNA arrays. This analysis validated our array data in all cases but indicated substantial gene expression differences in three of the transcripts that we chose to validate. In particular, the transcript for the intestinal trefoil protein, TFF3, was dramatically up-regulated in the cancers with E histology as was the AGR2 developmental gene (Fig. 3c). The trefoil peptides are implicated in other cancer types and down-regulation of TFF1 and TFF2 is often accompanied by up-regulation of TFF3, which mediates various cell adhesion and other signaling pathways (36–39). The observed 40-fold up-regulation of this gene deserves more study in E cancers.
Overexpression of the folate receptor and its fetal homologues was noted in the PS cancers. Real-time quantitative PCR assays revealed that overexpression of FOLR in PS cancers was even more striking than suggested by the microarray, with 9 of 13 PS tumors showing a significantly elevated level of expression compared with normal endometrium. These data are presented as a group in Fig. 3b, where these cancers express this gene at levels >60-fold, as compared with normal endometrium. Vaccines that target FOLR have been developed for use in ovarian carcinoma (40–42) and could potentially be used in the adjuvant treatment of PS endometrial cancers that overexpress this gene.

In summary, these array data will be useful for investigating pathways to be targeted by small molecules or for future gene reactivation strategies yet to be realized. As expected, many genes were disregulated in various histological groups, but our data also indicate many genes similarly disregulated in the various histologies. These data identify several additional pathways important in the development of endometrial cancer and suggest multiple avenues of investigation.

References


Received

11
Microarray Analysis Reveals Distinct Gene Expression Profiles among Different Histologic Types of Endometrial Cancer

John I. Risinger, G. Larry Maxwell, G. V. R. Chandramouli, et al.