Progression of Barrett's Metaplasia to Adenocarcinoma Is Associated with the Suppression of the Transcriptional Programs of Epidermal Differentiation

Erik T. Kimchi,[1] Mitchell C. Posner,[1] James O. Park,[1] Thomas E. Darga,[2] Masha Kocherginsky,[3] Theodore Karrison,[3] John Hart,[4] Kerrington D. Smith,[1] James J. Mezhir,[1] Ralph R. Weichselbaum,[2] and Nikolai N. Khodarev[2]

Departments of [1]Surgery, [2]Radiation and Cellular Oncology, [3]Health Studies, and [4]Pathology, University of Chicago, Chicago, Illinois

Abstract

We did expressional profiling on 24 paired samples of normal esophageal epithelium, Barrett's metaplasia, and esophageal adenocarcinomas. Matching tissue samples representing the three different histologic types were obtained from each patient undergoing esophagectomy for adenocarcinoma. Our analysis compared the molecular changes accompanying the transformation of normal squamous epithelium with Barrett's esophagus and adenocarcinoma in individual patients rather than in a random cohort. We tested the hypothesis that expressional profiling may reveal gene sets that can be used as molecular markers of progression from normal esophageal epithelium to Barrett's esophagus and adenocarcinoma. Expressional profiling was done using U133A GeneChip (Affymetrix, Santa Clara, CA), which represents approximately two thirds of the human genome. The final selection of 214 genes permitted the discrimination of differential gene expression of normal esophageal squamous epithelium, Barrett's esophagus, and adenocarcinoma using two-dimensional hierarchical clustering of selected genes. These data indicate that transformation of Barrett's esophagus to adenocarcinoma is associated with suppression of the genes involved in epidermal differentiation, including genes in *1q21* loci and corresponding to the epidermal differentiation complex. Correlation analysis of genes concordantly expressed in Barrett's esophagus and adenocarcinoma revealed 21 genes that represent potential genetic markers of disease progression and pharmacologic targets for treatment intervention. PCR analysis of genes selected based on DNA array experiments revealed that estimation of the ratios of *GATA6* to *SPRR3* allows discrimination among normal esophageal epithelium, Barrett's dysplasia, and adenocarcinoma. (Cancer Res 2005; 65(8): 3146-54)

Introduction

Barrett's esophagus is a specialized intestinal metaplasia of normal squamous to columnar epithelium. Barrett's esophagus is thought to be a premalignant transformation and has been identified in 80% to 100% of esophageal adenocarcinoma of the distal esophagus (1). The etiology of Barrett's esophagus is poorly understood, but chronic gastroesophageal reflux is considered a major contributing factor (2). The presence of Barrett's esophagus increases the risk of developing adenocarcinoma 40- to 125-fold (3). The incidence of adenocarcinoma has increased 3.5-fold over the past three decades, which exceeds that of all other types of cancer (4, 5). Patients with adenocarcinomas of the esophagus present with advanced disease, and 5-year survival is ~25% (6). Currently, endoscopic surveillance is the only method of identifying patients with early-stage esophageal cancers arising in Barrett's esophagus. In spite of endoscopic intervention, the infrequent association of cancers in Barrett's esophagus makes endoscopy a relatively low-yield intervention, especially considering the costs (7). However, identification of biological markers of Barrett's esophagus progression might identify high-risk patients in whom endoscopy might have a higher yield (8). Expressional profiling represents one such method (9–12). Molecular markers, which identify patients at the higher risk for subsequent transformation of Barrett's esophagus to adenocarcinoma, have yet to be identified. We examined differential gene expression patterns in resected esophageal specimens composed of normal esophageal epithelium, Barrett's esophagus, and adenocarcinoma obtained from the same individual patients. We selected and analyzed gene expression obtained from DNA microarrays and reported that transition from normal esophageal epithelium to Barrett's esophagus and esophageal adenocarcinoma is associated with the suppression of epidermal differentiation programs, including genes localized in the epidermal differentiation complex (EDC). The EDC represents the chromosomal region in *1q21* containing >40 genes responsible for epidermal differentiation. Major families of these genes are S100 proteins, small proline-rich proteins, precursors of the cornified envelope, and intermediate filament-associated proteins (see refs. 13, 14 for reviews). In our investigations, we focused on two members of the EDC, *S100A2* and *SPRR3*. In quantitative reverse transcription-PCR (QRT-PCR) experiments, we found that *SPRR3* combined with two transcriptional factors (*GATA6* and *HOXB7*) can successfully discriminate among normal epithelium, Barrett's dysplasia, and Barrett's esophagus–associated adenocarcinomas. These data show that selection of a limited number of markers based on DNA array experiments and their translation into PCR format allows discrimination of normal esophageal epithelium, Barrett's esophagus, and adenocarcinomas with relatively simple techniques.

Materials and Methods

Clinical samples. Samples of normal, Barrett's, and adenocarcinoma were obtained from fresh pathologic specimens of patients with known Barrett's esophagus and esophageal adenocarcinoma who had undergone esophagectomy. These specimens were processed by pathology within

Note: Supplementary data for this article are available at Cancer Research Online (http://cancerres.aacrjournals.org).

Requests for reprints: Nikolai N. Khodarev, Department of Radiation and Cellular Oncology, University of Chicago, Chicago, IL 60637. Phone: 773-834-3282; Fax: 773-702-1968; E-mail: nikolai@rover.uchicago.edu.

15 minutes of resection. Samples representative of the various gross histologic types were obtained from experienced gastrointestinal pathologists. These samples were labeled and snap frozen in liquid nitrogen and stored at −80°C for future RNA extraction.

Preparation of RNA and hybridizations. RNAs were purified by combination of column chromatography and TRIzol (Life Technologies, Carlsbad, CA) purification as described in ref. 15. Preparation of labeled cRNA and hybridization with U133A chips was done according to the manufacturer's instructions (Affymetrix, Santa Clara, CA). Data were acquired using MAS 5.0 software (Affymetrix) and exported to MS Excel.

Submission of DNA array data. Data were submitted to the Microarray Analysis and Data Management System database of the University of Chicago, constructed according to the Minimum Information About a Microarray Experiment recommendations and can be accessed through the Web address of the Microarray Analysis and Data Management System database (http://madam.bsd.uchicago.edu:8080/) using public log on. Data were also submitted to the GEO database (National Center for Biotechnology Information), with the accession no. GSE1420.

Data analysis. Throughout this section, patients are denoted by letter $K = 1, \ldots 8$ and genes by letter j.

For data normalization, the expression levels of each array were multiplied by $\tilde{M}/M$, where M is the median expression of the array and $\tilde{M}$ is the overall median expression level. This rescaling makes median expression levels equal across all arrays. For data filtration, we first excluded genes based on present (P) or absent (A) calls as defined by MAS 5.0. Genes were excluded if they were absent for all three tissue types in three or more patients. The genes were further filtered based on signal intensities using receiver operating characteristic analysis as described previously (16, 17). The total number of remaining genes was 8,636.

Next, the significance analysis of microarrays approach (18) was used to identify genes significantly overexpressed and underexpressed in the three pairwise comparisons of Barrett's/normal, adenocarcinoma/normal, and Barrett's/adenocarcinoma. Significance analysis of microarrays identifies genes with statistically significant differences between groups by assigning each gene a score based on the difference in gene expression between two groups (e.g., normal and Barrett's) relative to the adjusted pooled SD of the multiple measurements from both groups. Permutations of the measurements are then used to estimate the false discovery ratio (FDR), the percentage of genes identified by chance. As the cutoff point, we used a Δ value, such that the estimated median number of falsely discovered (called) genes was <1, and required at least a 2-fold expression ratio. In contrast to using a cutoff point of a fixed FDR level, this approach resulted in different cutoff Δs and FDR levels for the three comparisons: Barrett's/normal (Δ = 1.270, FDR = 0.33%), adenocarcinoma/normal (Δ = 1.555, FDR = 0.121%), and adenocarcinoma/Barrett's (Δ = 0.892, FDR = 0.876%). Based on these criteria, we selected 447 genes significantly expressed in adenocarcinomas compared with normal epithelium and 200 genes significantly expressed in Barrett's esophagus compared with normal epithelium. We also found 85 genes significantly different between adenocarcinoma and Barrett's esophagus, among which 45 overlapped with genes significantly different in adenocarcinoma versus normal epithelium. Next, expression ratios of all genes between two tissue groups were compared with the reference "same-to-same" distribution to identify genes for which the ratios are larger than expected. In a simple case with two normal samples, the "same-to-same" distribution is the distribution of $L_j = \log_2 N_{j1}/N_{j2}$ over all genes j (17). We extend this concept to a situation with more than two arrays by considering

$$L_j = \log_2 \left[\left(\frac{N_{j1} \cdot N_{j2} \ldots \cdot N_{j,\frac{K}{2}}}{N_{j,\frac{K}{2}+1} \cdot N_{j,\frac{K}{2}+2} \ldots N_{jK}} \right)^{1/(K/2)} \right]$$

where K is an even number of normal samples and N_{jk} is the expression level of gene j. For every gene j, we consider the $C = (KK/2)$ possible ways the samples can be separated into two equal size groups, obtaining $C = 70$ possible combinations for each gene based on the eight arrays, hybridized with RNA from normal tissues. For each of the 70 distributions, we compute quantiles $q_{0.005}$, $q_{0.025}$, $q_{0.975}$, and $q_{0.995}$, corresponding to nonparametric 95% and 99% confidence limits. Averaging these over the 70 combinations provides cutoff points for where the bulk of the same-to-same log ratios occur. For each gene j Barrett's/normal and adenocarcinoma/normal ("different-to-same") log ratio

$$L_j = \log_2 \left[\left(\frac{T_{j1} \cdot T_{j2} \ldots \cdot T_{jK}}{N_{j1} \cdot N_{j2} \ldots \cdot N_{jK}} \right)^{1/K} \right]$$

is then compared with the reference "same-to-same" distributions, and genes with expression ratios outside the cutoff limits are considered to be differentially expressed. Using the geometric mean rather than the nonstandardized ratio allows us to compare directly the distributions of the "same-to-same" and "different-to-same" hybridizations, adjusting for the fact that the "different-to-same" ratios are based on K pairs of tumor and normal expression levels and that the "same-to-same" ratios are based only on K normal expression levels. Thus, L_j can be naturally interpreted as average per-patient log ratio.

Functional selection and prognosticators analysis. For the selection of functionally significant groups of genes, we used OntoExpress software described by Draghici et al. (19) and available at http://vortex.cs.wayne.edu:8080/ontoexpress/servlet/UserInfo. We selected functional groups containing at least three genes and used a binomial distribution with a significance level of ≤ 0.05. Combining results of functional and expression-based selections, we finally selected 214 genes. Two-dimensional hierarchical clustering of these genes was done based on the estimation of the Euclidean distances by Ward's method using $\log_2 X_{ijk}/\tilde{X}_j^{(N)}$, the log-transformed expression levels normalized to the average expression level in the normal tissues, $\tilde{X}_j^{(N)}$. Samples T5 and N8 were removed as outliers. For clustering and data presentation, we used JMP and TreeView software as described previously (20).

To define genes that correlate with the progression of Barrett's esophagus to adenocarcinoma, we considered the 96 genes expressed in both Barrett's esophagus and adenocarcinoma. These genes were separated into two groups based on average between-patient expression: the first group contained genes that were up-regulated from normal to Barrett's esophagus and further from Barrett's esophagus to adenocarcinoma and the second group was defined similarly for down-regulated genes. All other potential patterns were excluded from this analysis. The significance of the difference in expression from normal to Barrett's esophagus and from Barrett's esophagus to adenocarcinoma in each group was evaluated by a one-sided paired t test using a $P < 0.05$ cutoff (taking into account that we already preselected genes that are either up-regulated or down-regulated in each group).

Quantitative reverse transcription-PCR. cDNA was synthesized using SuperScript II reverse transcriptase (Invitrogen Life Technologies, Carlsbad, CA) following the manufacturer's instructions. cDNA was diluted 1:10 in sterile nuclease free water (Ambion, Austin, TX). Quantitative PCR was done on an ABI 7700 system (Applied Biosystems, Foster City, CA) using SYBR Green PCR reagents in a 25 μL reaction mixture containing 2.5 μL of 10× SYBR Green PCR buffer, 0.25 μL of 10 mmol/L primers, 2 μL deoxynucleotide triphosphate mix, 3 μL of 25 mmol/L $MgCl_2$, 0.25 μL AmpErase, 0.125 μL Amplitaq Gold, and 2.5 μL of the 1:10 diluted cDNA.

Primers for selected genes were designed based on Unigene reference sequences using PrimerExpress software (Applied Biosystems). For the internal control, we used glyceraldehyde-3-phosphate dehydrogenase (GAPDH). PCR was done for 40 cycles at 95°C for 15 seconds and 60°C for 1 minute after initial incubations at 50°C for 2 minutes and 95°C for 10 minutes.

All samples were amplified in triplicate reactions. The expression of each individual gene was calculated based on the difference between amplification of the individual mRNA template and the internal control (GAPDH) mRNA template. These differences were measured by delta ct (dct) values as described in the manufacturer's instructions (Applied Biosystems). dct values were calculated as $(ct_X - ct_{GAPDH})$, where ct_X is the ct value of the specific gene X and ct_{GAPDH} is the amplification of the

internal control. Fold induction relative to GAPDH was calculated as 2^{-dct} and therefore was equal to $2^{-(ctX\ -\ ctGAPDH)}$. Ratios of gene X relative to gene Y in the same samples were calculated as $R_{X/Y} = 2^{-(ctX\ -\ ctY)}$. To transform these ratios to a range above 1, we multiplied them by a factor of 100,000. Finally, we transformed the data to $\log_{10}$ format. The final expressional value was calculated as:

$$EV_{X/Y} = \log_{10}[10^5 \times \{2^{-(ct_X - ct_Y)}\}].$$

Results

Discrimination among normal esophageal epithelium, Barrett's metaplasia, and adenocarcinomas based on expressional profiling. Based on the results of the statistical analysis, we selected genes differentially expressed (either up-regulated or down-regulated) in Barrett's esophagus and adenocarcinoma. Figure 1 represents these results as a Venn diagram. Compared with normal esophageal epithelium, we identified 200 genes differentially expressed in Barrett's esophagus and 447 genes differentially expressed in the Barrett's esophagus–associated adenocarcinoma. The criteria and techniques for the selection of the differentially expressed genes are described in Materials and Methods. The comparison of genes differentially expressed in Barrett's esophagus and adenocarcinoma showed that 96 genes were commonly overexpressed in Barrett's esophagus and adenocarcinoma. Adenocarcinoma differentially expressed 351 genes that were not expressed in Barrett's esophagus, and Barrett's esophagus differentially expressed 104 genes that were not expressed in adenocarcinoma. Nonoverlapping genes (351 in adenocarcinoma and 104 in Barrett's esophagus) were used in the subsequent selection of significant functional groups using OntoExpress software and these functions are presented in Table 1. We also selected genes by comparison of the "same-to-same" and "different-to-same" hybridizations as described in Materials and Methods using 99% confidence intervals based on the nonparametric quantile analysis. Combining both approaches, we selected 214 genes, presented in Fig. 2, which represent the two-dimensional hierarchical clustering of selected genes and show the actual discrimination among normal samples, Barrett's esophagus, and adenocarcinoma. The list of these genes is presented in Supplementary Table S1. In the first dimension (vertical), the data separate into three expressional clusters according to the dendrogram presented in Fig. 2 (*right*). Cluster 1 (80 genes) contains the genes up-regulated in adenocarcinoma compared with normal epithelium. Cluster 2 (63 genes) contains the genes that are sequentially suppressed in Barrett's esophagus and adenocarcinoma compared with the normal epithelium. Cluster 3 (71 genes) contains the genes most drastically suppressed in adenocarcinoma compared with normal epithelium and Barrett's esophagus.

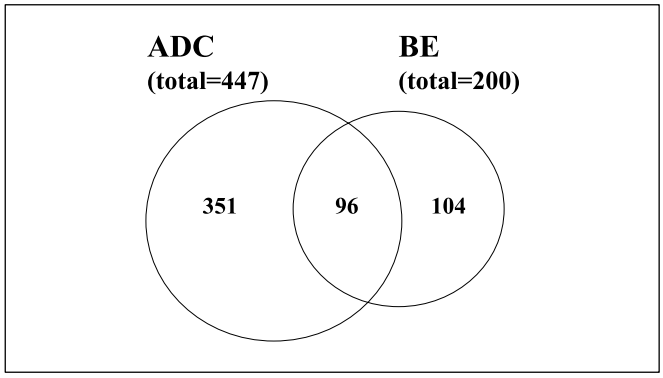

Figure 1. A Venn diagram representing genes selected as differentially expressed in Barrett's esophagus (*BE*) and adenocarcinoma (*ADC*) relative to the normal esophageal epithelium.

Expressional patterns of normal epithelium, Barrett's esophagus, and adenocarcinoma include different functional groups of genes. Figure 3 identifies the major functional groups associated with the three major expressional clusters presented in Fig. 2. The first expressional cluster contains functional groups of genes associated with immune response, cell-cell signaling and cell-extracellular matrix interactions, control of cell cycle/growth/proliferation, and regulation of transcription and receptor activity (Fig. 3*A*).

A detailed description of these groups in the terms of the Gene Ontology (GO) database is presented in Table 1. The second expressional cluster also includes control of the cell cycle/proliferation but also contains the specific functions of intracellular transport, bile acid transport, and aldehyde and lipid metabolism (Fig. 3*B*). The third expressional cluster contains functional groups of genes that may specifically lead to the development of adenocarcinoma. They include ectoderm development/epidermal differentiation, cytoskeleton, control of cell shape and cell-to-cell and cell-to-extracellular matrix interactions, Ca^{2+} binding and metabolism, and a group of proteases and protease inhibitors. Many of these genes are specifically associated with epidermal differentiation and malignant transformation (see Discussion).

Analysis of genes common to Barrett's and adenocarcinoma. Ninety-six genes were commonly expressed in Barrett's esophagus and adenocarcinomas. The list of these genes is presented in Supplementary Table S2. We assumed that these genes might be helpful prognostic and diagnostic markers and/or potential therapeutic targets if they satisfy several conditions. These conditions are as follows: (*a*) common differential expression (up-regulation or down-regulation) in Barrett's esophagus and adenocarcinoma relative to normal esophageal epithelium, (*b*) significance of differences between expression in Barrett's esophagus versus normal epithelium and adenocarcinoma versus Barrett's esophagus, and (*c*) identical expressional changes (e.g., up-regulation in Barrett's esophagus relative to normal and further up-regulation in adenocarcinoma relative to Barrett's esophagus). Based on these assumptions (see Materials and Methods), we selected the 21 genes described in Table 2. Three of these genes (*HLA-F*, *HOXB7*, and *RRBP1*) were presented by two probe sets. These genes are progressively up-regulated or down-regulated in Barrett's esophagus and adenocarcinoma and have significant differences between expressional levels in Barrett's esophagus versus normal epithelium and adenocarcinoma versus Barrett's esophagus.

Quantitative reverse transcription-PCR experiments. For confirmation of our DNA array-based expressional data, we used QRT-PCR as described in Materials and Methods. Individual samples of RNA were profiled using six gene primer pairs for *GATA6*, *HOXB7*, *TCF3*, *S100A2*, *SPRR3*, and *SCCA1*. These genes were selected based on potential prognostic value and levels of differential expression on the arrays. *S100A2* and *SPRR3* were also selected as components of the EDC (14). Data from these experiments are presented in Figs. 4 and 5. Fig. 4*A* shows a trend in three transcriptional factors (*GATA6*, *HOXB7*, and *TCF3*) toward up-regulation in the transition from normal esophageal

Table 1. Selected functional groups for Barrett's and adenocarcinomas

Barrett		Adenocarcinomas			
GO biological process		GO biological process		GO molecular function	
GO ID	Function name	GO ID	Function name	GO ID	Function name
0000074	Regulation of cell cycle	0001558	Regulation of cell growth	0003700	Transcription factor activity
0001501	Skeletal development	0006081	Aldehyde metabolism	0003821	Class II MHC antigen
0006812	Cation transport	0006355	Regulation of transcription, DNA dependent	0004029	Aldehyde dehydrogenase (NAD) activity
0006915	Apoptosis	0006461	Protein complex assembly	0004263	Chymotrypsin activity
0006935	Chemotaxis	0006629	Lipid metabolism	0004295	Trypsin activity
0006955	Immune response	0006886	Intracellular protein transport	0004601	Peroxidase activity
0007160	Cell-matrix adhesion	0006899	Nonselective vesicle transport	0004867	Serine protease inhibitor activity
0007166	Cell surface receptor linked signal transduction	0006944	Membrane fusion	0004930	G-protein-coupled receptor activity
0007229	Integrin-mediated signaling pathway	0006979	Response to oxidative stress	0005152	Interleukin-1 receptor antagonist activity
0007267	Cell-cell signaling	0007048	Oncogenesis	0005198	Structural molecule activity
0008151	Cell growth and/or maintenance	0007398	Ectoderm development	0005200	Structural constituent of cytoskeleton
0008152	Metabolism	0007417	Central nervous system development	0005509	Calcium ion binding
0009653	Morphogenesis	0008284	Positive regulation of cell proliferation	0005524	ATP binding
		0008544	Epidermal differentiation	0005525	GTP binding
		0016049	Cell growth	0008237	Metallopeptidase activity
		0019883	Antigen presentation, endogenous antigen	0016301	Kinase activity
GO molecular function		0019885	Antigen processing, endogenous antigen via MHC class	0016853	Isomerase activity
0004716	Receptor signaling protein tyrosine kinase activity	0045786	Negative regulation of cell cycle	0030106	MHC class I receptor activity
0004872	Receptor activity	Null	Cell shape and cell size control	0045012	MHC class II receptor activity
0004895	Cell adhesion receptor activity	0006470	Protein amino acid dephosphorylation	0004033	Aldo-keto reductase activity
0008201	Heparin binding	0006805	Xenobiotic metabolism	0005488	Binding
		0006810	Transport	0008014	Calcium-dependent cell adhesion molecule activity
		0006955	Immune response	0015125	Bile acid transporter activity
		0007155	Cell adhesion	0017017	Mitogen-activated protein kinase phosphatase activity
		0007156	Homophilic cell adhesion	0047115	*Trans*-1,2-dihydrobenzene-1,2-diol dehydrogenase activity
		0007267	Cell-cell signaling		

epithelium to adenocarcinomas. In contrast, three genes related to keratinocyte differentiation (*S100A2*, *SPRR3*, and *SCCA1*) trended toward down-regulation during the transition from normal squamous to malignant epithelium (Fig. 4*B*). These data are consistent with the results from the entire set of tissue types in the microarray analysis as shown in Fig. 4*C* and *D*. To select expressional markers most closely correlated with premalignant and malignant changes, we calculated *P*s and correlation coefficients for single genes and their combinations, and these are presented in Table 3, where we see that each single marker can significantly discriminate normal esophageal epithelium from adenocarcinoma. However, only *HOXB7* can discriminate normal tissues from Barrett's. Neither of these markers alone can discriminate Barrett's from adenocarcinomas and simultaneously normal tissue from Barrett's. We therefore decided to check whether combinations of markers may provide better discrimination than each marker alone. As is shown in Table 3, combinations of *GATA6/SPRR3*, *HOXB7/SPRR3*, and *GATA6+HOXB7/SPRR3* discriminate better than each of the genes alone. For these gene combinations, the *t* test provides highly significant values for discrimination between each condition: Barrett's esophagus and normal, adenocarcinoma and normal, and Barrett's esophagus and adenocarcinoma (P = 0.0012, 1.462×10^{-6}, and 0.0013, respectively). Additionally, mixed-effects ANOVA models were used to determine whether there are differences in expression of *GATA6/SPRR3*, *HOXB7/SPRR3*, and *GATA6+HOXB7/SPRR3* combinations between the three groups, accounting for the presence of intrasubject correlation due to the presence of several subjects with multiple samples. These analyses confirmed that the expression levels of these combination markers are significantly

different among normal, Barrett's, and tumor tissue types (data not shown). In addition, for the combinations listed, the correlation between expression and tumor progression is higher than for either gene alone. However, increasing the number of combined markers may not improve correlation with clinical diagnosis. As is shown in Table 3, double markers (*GATA6/SPRR3* and *HOXB7/SPRR3*) did as well as triple markers (*GATA6+HOXB7/SPRR3*). These differences suggested the use of the *GATA6/SPRR3* ratio as a potential molecular marker of transformation. The performance of this ratio as a marker is presented in Fig. 5, which shows that the ratio of *GATA6* to *SPRR3* progressively increases along the spectrum from normal epithelium to Barrett's dysplasia to adenocarcinomas. In the range of the 95% confidence interval (estimated by normal epithelium values), specificity of the test (measured by percentage of negative cases) is equal to 89% (8 of 9). Sensitivity of the test (measured by percentage of positive cases) is equal to 28.6% (2 of 7) for Barrett's esophagus and 100% (7 of 7) for adenocarcinoma. With a cutoff level equal to 67% confidence interval, specificity of the test is also equal to 89%, sensitivity to Barrett's esophagus is 86% (6 of 7), and sensitivity to adenocarcinoma is 100% (7 of 7). These data show that selection of the markers based on DNA array experiments allows significant separation of pathologic conditions and that the number of these markers can be reduced to several specific genes that can make such an estimation feasible in large populations.

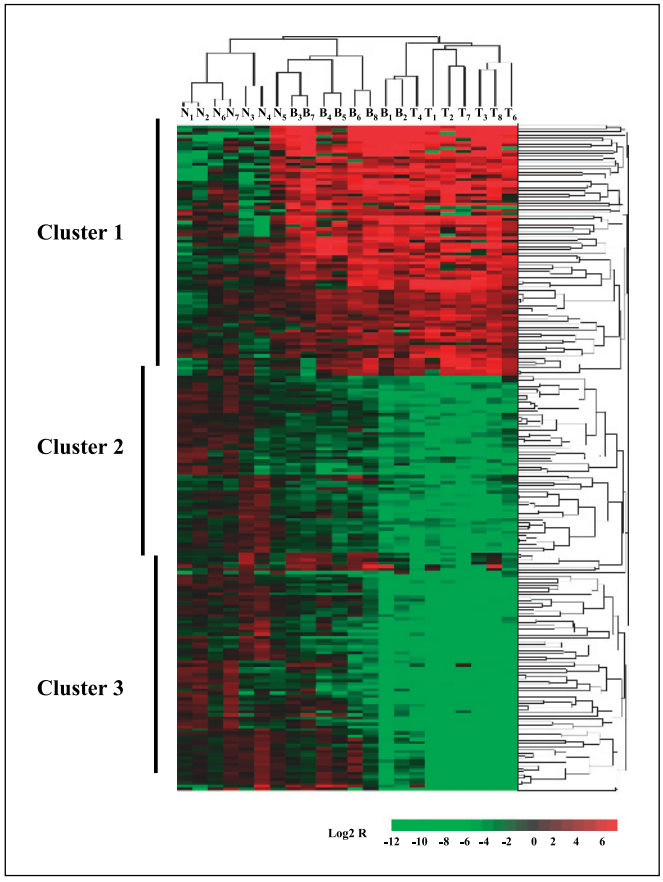

Figure 2. Hierarchical clustering of genes selected for discrimination of the normal esophageal epithelium (*N*), Barrett's esophagus (*B*), and Barrett's esophagus–associated adenocarcinomas (*T*). *X axis,* samples of different tissues obtained from the same patients (e.g., N_1, B_1, and T_1 correspond to the normal esophageal epithelium, Barrett's esophagus, and Barrett's esophagus–associated adenocarcinoma obtained from patient 1). *Y axis,* genes included in the three major expressional clusters indicated by *thick lines* on the *left*. Genes included in these expressional clusters are presented in Fig. 3 and Supplementary Table S1. The color code is in the *bottom right*.

Discussion

Several gene families are consistently down-regulated in Barrett's and adenocarcinoma compared with normal esophageal epithelium and/or adenocarcinoma compared with Barrett's esophagus (Fig. 3*B* and *C*). These genes include aldo-keto reductases, aldehyde dehydrogenases (located predominantly in expressional cluster 2), dual-specificity phosphatases, annexins, chloride channels, keratins and genes involved in the formation of desmosomes, and the cornified envelope of squamous epithelium (see refs. 21–23 for reviews). The genes involved in these processes are S100 proteins, small proline-rich proteins, involucrin, periplacin, desmocollins, and desmoglein 3. Many of these genes are in the specific chromosomal loci called the *EDC* (13, 14). The *EDC* is the chromosomal loci containing >40 genes involved in epidermal differentiation and localized in *1q21* (see refs. 13, 14, 24). Excluded from the genes presented in Figs. 2 and 3 were the 96 genes commonly expressed in Barrett's esophagus and adenocarcinoma (see also Supplementary Tables S1 and S2). Among these, two genes are also involved in epidermal differentiation: *transglutaminase-3* (25) and *NICE-1*, with the latter also localized in *1q21* (14). Our data indicate that the formation of Barrett's esophagus–associated adenocarcinoma is accompanied by the suppression of the genetic programs of epidermal differentiation, including genes clustered in the EDC, which raises questions regarding the mechanisms of suppression of the EDC. One explanation may be the deletion or loss of heterozygosity of *1q21*, although data from the literature are controversial. Loss of heterozygosity of *1q21* was shown in six gastrinomas (26), but in the majority of other cases, cytogenetics, comparative genomic hybridization, and fluorescence *in situ* hybridization analysis reveal amplifications of *1q21* or translocations involving this locus (27–31). Recent data indicate that activity of *EDC* genes may be controlled by highly coordinated changes of local chromatin structure, specific spatial distribution of chromatin regions, and temporally and spatially ordered transcriptional regulation (13, 32, 24). These data show that in interphase nuclei of keratinocytes (but not lymphoblasts) the EDC is positioned external to chromosome 1 (13). The same situation was shown for the transcriptionally active *MHC* locus on chromosome 6 (33). Interestingly, that suppression of histone deacetylases led to the "open" conformation of the EDC and increased expression of some genes. Silencing of genes in this locus was connected with CpG methylation (24). Such coordinated changes of conformation may be tightly coordinated with transcriptional control. From that point of view, our data provide interesting correlations between transcriptional activity of the EDC and expression of different transcriptional and chromatin remodeling factors, mostly grouped in expressional cluster 1 (see Figs. 2 and 3*A*; Table 2). For example, the evolutionarily conserved CBX3 protein involved in the condensation of heterochromatin regions (34) increased 1.3-fold in Barrett's esophagus and 2.8-fold in adenocarcinoma relative to normal epithelium, consistent with the possibility of condensation of the EDC region (see Fig. 3*A*; Supplementary Table S1). *HOX* genes have been implicated in human skin development (35). *HOXA7* has been shown recently to regulate transglutaminase-1 and suppress keratinocyte

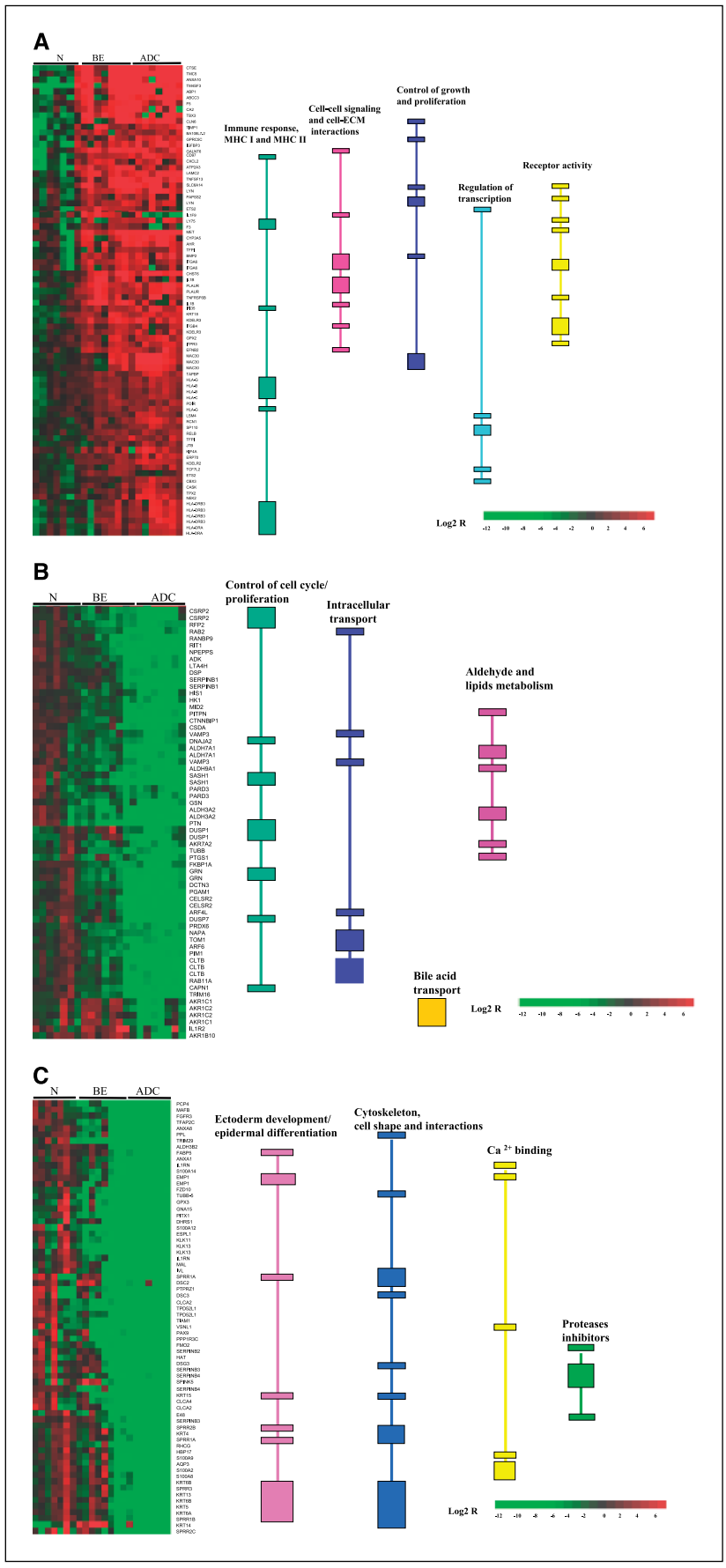

Figure 3. Correlation between expressional clusters and functional groups of genes. The color panel represents up-regulated or down-regulated genes (the color code is the same as on Fig. 2). The colored boxes on the *right* of the clusters represent specific functional groups. Major groups are indicated. Positions and sizes of the boxes correspond to the genes presented in the cluster and belonging to this specific functional group. *Lines above the color panel* show the normal, Barrett's, and adenocarcinoma samples. *A,* expressional cluster 1; *B,* expressional cluster 2; *C,* expressional cluster 3 (see Fig. 2).

Table 2. Genes with progressive changes of expression in Barrett's and adenocarcinomas

ID	Symbol	Name	Ratio (Barrett's/normal)	Ratio (tumor/normal)
Up-regulated genes				
201301_s_at	*ANXA4*	Annexin A4	2.28	3.13
201954_at	*ARPC1B*	Actin-related protein 2/3 complex, subunit 1B	3.20	5.42
214439_x_at	*BIN1*	Ridging integrator 1	2.23	3.26
202901_x_at	*CTSS*	Cathepsin S	3.08	5.26
210002_at	*GATA6*	GATA-binding protein 6	6.27	10.77
221875_x_at	*HLA-F*	MHC, class I, F	2.28	3.35
204806_x_at	*HLA-F*	MHC, class I, F	2.17	3.20
204779_s_at	*HOXB7*	Homeo box B7	3.56	5.82
216973_s_at	*HOXB7*	Homeo box B7	2.71	4.42
201422_at	*IFI30*	IFN-γ-inducible protein 30	2.23	4.11
212110_at	*KIAA0062*	SLC39A14: solute carrier family 39 (zinc transporter), member 14	5.26	7.86
203943_at	*KIF3B*	Kinesin family member 3B	2.27	3.43
218376_s_at	*NICAL*	NEDD9 interacting protein with calponin homology and LIM domains	2.03	3.12
219622_at	*RAB20*	RAB20, member RAS oncogene family	2.90	4.66
201206_s_at	*RRBP1*	Ribosome binding protein 1	4.02	5.80
201204_s_at	*RRBP1*	Ribosome binding protein 1	2.46	3.34
213811_x_at	*TCF3*	Transcription factor 3	2.84	4.45
208998_at	*UCP2*	Uncoupling protein 2	3.52	6.57
Down-regulated genes				
210020_x_at	*CALML3*	Calmodulin-like 3	0.40	0.11
203585_at	*ZNF185*	Zinc finger protein 185	0.46	0.17
213005_s_at	*KANK*	Kidney ankyrin repeat-containing protein	0.49	0.24
211734_s_at	*FCER1A*	Fc fragment of IgE, high affinity I, receptor for α subunit	0.25	0.14
201848_s_at	*BNIP3*	BCL2/adenovirus E1B 19-kDa protein-interacting protein 3	0.43	0.26
219100_at	*FLJ22559*	Hypothetical protein	0.48	0.29

differentiation (36). *HOXB7* was identified as the promoter of ovarian carcinomas (37) and identified recently as an important factor in the tumor-associated angiogenic switch (38). *TCF3* (see Table 2) was identified as a suppressor of epidermal differentiation (39) and *GATA* factors as regulators of epithelial differentiation and skin development (40, 41). In *Caenorhabditis elegans*, genes encoding *GATA* factors (*elt-5* and *elt-3*) control differentiation of epidermal cells throughout animal development (42). In addition to the transcriptional factors, several other genes indicated in Table 2 are related to epidermal differentiation and

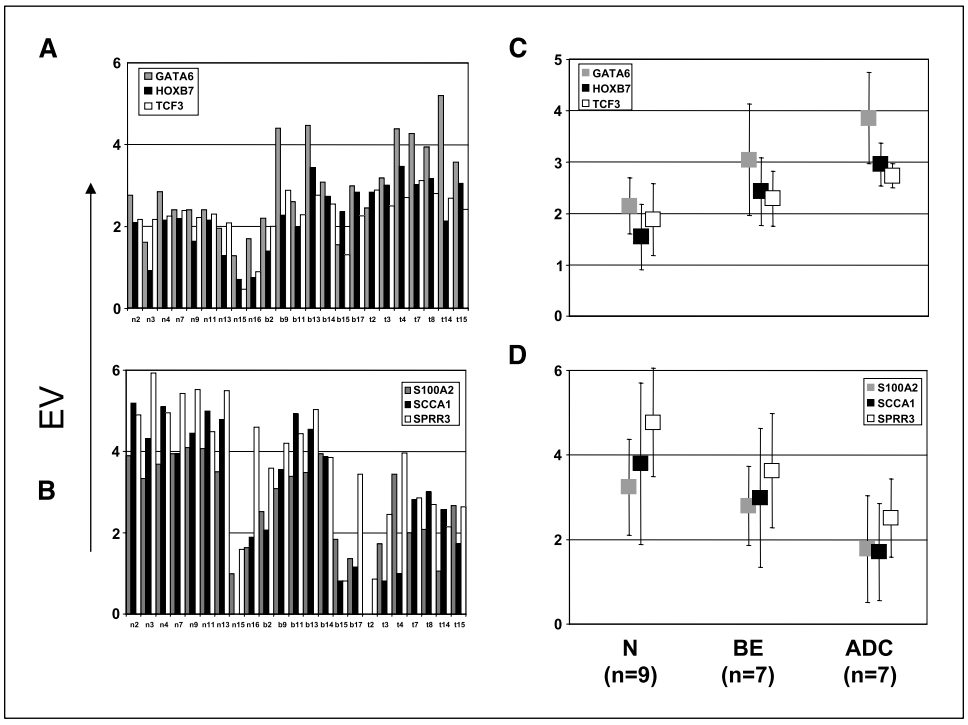

Figure 4. PCR screening of individual patients. RNA was purified from surgical samples, and QRT-PCR for *GATA6*, *HOXB7*, *TCF3*, *S100A2*, *SCCA1*, and *SPRR3* was done as described in Materials and Methods. Patients 9, 11, and 13 to 17 were enrolled after DNA arrays profiling experiments. *A,* genes up-regulated in the progression from normal to Barrett's esophagus to adenocarcinoma. *B,* genes down-regulated in the progression. Four samples (*n12*, *n14*, *n17*, and *t17*) failed to amplify specific gene products by PCR and were excluded. *Y axis,* expressional value (*EV*) was calculated relative to GAPDH. *C* and *D,* corresponding average values. *Bars,* SD.

oncogenesis. *RRBP1* is involved in the epithelial-mesenchymal transformation (43), and *KANK* has been reported as a tumor suppressor (44). Its orthologue, *VAB-19*, is critical for epidermal differentiation in *C. elegans* (45). Based on the progressive behavior of these genes and significance of differences, we suggest that they may be not only useful markers but also potential targets for pharmacologic intervention.

Two recent publications where DNA arrays or subtractive libraries were used for analysis addressed the issue of differential gene expression in Barrett's esophagus, adenocarcinoma, and squamous cell carcinoma formation compared with normal esophageal epithelium and are germane considering our observations. Barrett et al. compared normal esophageal, gastric and duodenal squamous epithelium, and Barrett's epithelia (12). Of the 203 genes included in the squamous epithelium cluster, 37 matched with our selection (Fig. 2; Supplementary Table S1). According to the observations of Barrett et al., genes associated with keratinocyte differentiation were up-regulated in the squamous epithelial cluster compared with duodenum and gastric epithelium. According to our observations, these genes are down-regulated in adenocarcinoma. In another report, human esophageal squamous cell carcinoma and normal esophageal tissues were used for construction of subtractive λ-libraries, which were screened by custom nylon arrays (9). Fifteen genes identified in this study as down-regulated were common or homological to the genes found in our current report. These genes also belong to the small proline-rich proteins, annexins, S100 proteins, keratins, and transglutaminase-3. These data show that results from different laboratories with different platforms for expressional profiling analysis lead to the consistent selection of genes involved in epidermal differentiation. In contrast to the observations with normal squamous epithelium where previous studies show up-regulation of differentiation-associated genes in squamous epithelium (12), our data show marked down-regulation of the same genes in Barrett's esophagus–associated adenocarcinoma. We suggest that such dedifferentiation is important for tumor clones to release from tight junctions with the surrounding cells through desmosomes and cornified envelope. This is a potential mechanism for invasive growth and formation of metastasis. Understanding of the regulatory mechanisms of suppression of epidermal differentiation programs and the EDC are critical for potential control of adenocarcinoma development. Our data show that development of Barrett's esophagus–associated adenocarcinoma is connected with the suppression of the epidermal differentiation program and associated with activation or alteration of expression of specific regulatory genes. Identification of these genes may provide insights into the etiology and development of Barrett's esophagus and esophageal adenocarcinomas. These findings may also provide a set of markers to determine pathologic stage and risk of adenocarcinoma development as well as define new potential therapeutic targets. QRT-PCR experiments provide further specification of such markers. Six selected genes (*GATA6*, *HOXB7*, *TCF3*, *S100A2*, *SCCA1*, and *SPRR3*) showed significant correlation during progression from normal tissue to Barrett's esophagus and adenocarcinoma and were able to be quantified by QRT-PCR in RNA samples extracted directly from surgical specimens. All are directly involved in or associated with epidermal differentiation. Our experiments showed that the ratio of *GATA6* to *SPRR3* can significantly discriminate normal from malignant esophageal epithelium. Most importantly, our data suggest that this ratio may vary in Barrett's dysplasia, perhaps reflecting risk of Barrett's transformation. Further investigations with a larger number of patients with clinical follow-up may reveal whether it can be used for estimation of the risk of Barrett's progression to adenocarcinoma.

These investigations show that the selection of markers based on DNA array experiments may provide molecular criteria for discrimination of pathologic conditions of esophageal epithelium. They also show that these data may be used for selection of a limited number of markers that can be screened in large populations by conventional techniques like PCR.

Table 3. *P*s and Pearson's correlation coefficients

	*P*s			*R*s
	Normal/ adeno carcinoma	Normal/ Barrett's	Barrett's/ adeno carcinoma	
GATA6	0.0014	0.0797	0.1510	0.6909
HOXB7	0.0001	0.0183	0.1045	0.7657
TCF3	0.0063	0.2048	0.0769	0.5797
S100A2	0.0332	0.4123	0.1131	−0.4752
SCCA1	0.0171	0.3794	0.1184	−0.4551
SPRR3	0.0011	0.1116	0.1014	−0.6177
GATA6/ SPRR3	1.4662×10^{-6}	0.0012	0.0013	0.8732
HOXB7/ SPRR3	4.0369×10^{-6}	0.0092	0.0211	0.8176
GATA6+HOXB7/ SPRR3	1.6406×10^{-6}	0.0028	0.0034	0.8628

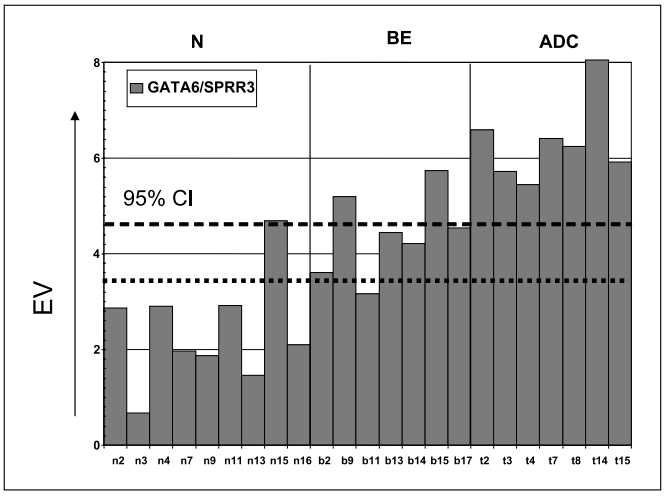

Figure 5. Ratios of *GATA6* to *SPRR3* in individual samples. *Dashed line,* 95% confidence interval (mean ± 1.96 SD). *Dotted line,* mean ± 1 SD.

Acknowledgments

Received 7/22/2004; revised 12/17/2004; accepted 2/9/2005.

Grant support: NIH grant PO1 CA 71933 (R.R. Weichselbaum) and Claire and Dennis Nardoni Fellowship in Surgical Oncology (E.T. Kimchi).

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.

We thank Edwardine Labay for active help with the preparation of this article.

References

1. Stein HJ, Siewert JR. Barrett's esophagus: pathogenesis, epidemiology, functional abnormalities, malignant degeneration, and surgical management. Dysphagia 1993; 8:276–88.
2. Lagergren J, Bergstrom R, Lindgren A, Nyren O. Symptomatic gastroesophageal reflux as a risk factor for esophageal adenocarcinoma. N Engl J Med 1999;340: 825–31.
3. Shaheen N, Ransohoff DF. Gastroesophageal reflux, Barrett esophagus and esophageal cancer. JAMA 2002; 287:1972–81.
4. Shaheen NJ, Crosby MA, Bozymski EM, Sandler RS. Is there publication bias in the reporting of cancer risk in Barrett's esophagus? Gastroenterology 2000; 119:333–8.
5. Devesa SS, Blot WJ, Fraumeni JF Jr. Changing patterns in the incidence of esophageal and gastric carcinoma in the United States. Cancer 1998;83:2049–53.
6. Swisher SG, Hunt KK, Holmes EC, Zinner MJ, McFaddwn DW. Changes in the surgical management of esophageal cancer from 1970 to 1993. Am J Surg 1995; 169:609–14.
7. Goldblum JR, Lauwers GY. Dysplasia arising in Barrett's esophagus: diagnostic pitfalls and natural history. Semin Diagn Pathol 2002;19:12–9.
8. McManus DT, Olaru A, Meltzer SJ. Biomarkers of esophageal adenocarcinoma and Barrett's esophagus. Cancer Res 2004;64:1561–9.
9. Luo A, Kong J, Hu G, et al. Discovery of Ca^{2+}-relevant and differentiation-associated genes downregulated in esophageal squamous cell carcinoma using cDNA microarray. Oncogene 2004;23:1291–9.
10. Xu Y, Selaru FM, Yin J, et al. Artificial neural networks and gene filtering distinguish between global gene expression profiles of Barrett's esophagus and esophageal cancer. Cancer Res 2002;62:3493–7.
11. Dahlberg PS, Ferrin LF, Grindle SM, et al. Gene expression profiles in esophageal adenocarcinoma. Ann Thorac Surg 2004;77:1008–15.
12. Barrett MT, Yeung KY, Ruzzo WL, et al. Transcriptional analyses of Barrett's metaplasia and normal upper GI mucosae. Neoplasia 2002;4:121–8.
13. Williams RR, Broad S, Sheer D, Ragoussis J. Subchromosomal positioning of the epidermal differentiation complex (EDC) in keratinocyte and lymphoblast interphase nuclei. Exp Cell Res 2002;272: 163–75.
14. Marenholz I. Identification of human epidermal differentiation complex (EDC)-encoded genes by subtractive hybridization of entire YACs to a gridded keratinocyte cDNA library. Genome Res 2001;11:341–55.
15. Khodarev NN, Yu J, Nodzenski E, et al. Method of RNA purification from endothelial cells for DNA array experiments. Biotechniques 2002;32:316–20.
16. Khodarev NN, Park J, Kataoka Y, et al. Receiver operating characteristic analysis: a general tool for DNA array data filtration and performance estimation. Genomics 2003;81:202–9.
17. Khodarev NN, Beckett M, Labay E, Darga T, Roizman B, Weichselbaum RR. STAT1 is overexpressed in tumors selected for radioresistance and confers protection from radiation in transduced sensitive cells. Proc Natl Acad Sci U S A 2004;101: 1714–9.
18. Tusher VG, Tibshirani R, Chu G. Significance analysis of microarrays applied to the ionizing radiation response. Proc Natl Acad Sci U S A 2001; 98:5116–21.
19. Draghici S, Khatri P, Martins RP, Ostermeier GC, Krawetz SA. Global functional profiling of gene expression. Genomics 2003;81:98–104.
20. Khodarev NN, Kataoka Y, Murley JS, Weichselbaum RR, Grdina DJ. Interaction of amifostine and ionizing radiation on transcriptional patterns of apoptotic genes expressed in human microvascular endothelial cells (HMEC). Int J Radiat Oncol Biol Phys 2004;60: 553–63.
21. Seery JP. Stem cells of the oesophageal epithelium. J Cell Sci 2002;115:1783–9.
22. Kitajima Y. Mechanisms of desmosome assembly and disassembly. Clin Exp Dermatol 2002;27:684–90.
23. Kalinin AE, Kajava AV, Steinert PM. Epithelial barrier function: assembly and structural features of the cornified cell envelope. Bioessays 2002;24: 789–800.
24. Elder JT, Zhao X. Evidence for local control of gene expression in the epidermal differentiation complex. Exp Dermatol 2002;11:406–12.
25. Hitomi K, Presland RB, Nakayama T. Analysis of epidermal-type transglutaminase (transglutaminase 3) in human stratified epithelia and cultured keratinocytes using monoclonal antibodies. J Dermatol Sci 2003;32: 95–103.
26. Chen YJ, Vortmeyer A, Zhuang Z, Huang S, Jensen RT. Loss of heterozygosity of chromosome 1q in gastrinomas: occurrence and prognostic significance. Cancer Res 2003;63:817–23.
27. Koon N, Zaika A, Moskaluk CA, et al. Clustering of molecular alterations in gastroesophageal carcinomas. Neoplasia 2004;6:143–9.
28. Pantou D, Tsarouha H, Papadopoulou A, et al. Cytogenetic profile of unknown primary tumors: clues for their pathogenesis and clinical management. Neoplasia 2003;5:23–31.
29. Neglia M, Bertoni L, Zoli W, Giulotto E. Amplification of the pericentromeric region of chromosome 1 in a newly established colon carcinoma cell line. Cancer Genet Cytogenet 2003;142:99–106.
30. Wong N, Chan A, Lee SW, et al. Positional mapping for amplified DNA sequences on 1q21-q22 in hepatocellular carcinoma indicates candidate genes overexpression. J Hepatol 2003;38:298–306.
31. Sy SM, Wong N, Lee TW, et al. Distinct patterns of genetic alterations in adenocarcinoma and squamous cell carcinoma of the lung. Eur J Cancer 2004;40: 1082–94.
32. Mahy NL, Perry PE, Bickmore WA. Gene density and transcription influence the localization of chromatin outside of chromosome territories detectable by FISH. J Cell Biol 2002;159:753–63.
33. Volpi EV, Chevret E, Jones T, et al. Large-scale chromatin organization of the major histocompatibility complex and other regions of human chromosome 6 and its response to interferon in interphase nuclei. J Cell Sci 2000;113:1565–76.
34. Garcia-Cao M, O'Sullivan R, Peters AH, Jenuwein T, Blasco MA. Epigenetic regulation of telomere length in mammalian cells by the Suv39h1 and Suv39h2 histone methyltransferases. Nat Genet 2004;36:94–9.
35. Stelnicki EJ, Komuves LG, Kwong AO, et al. HOX homeobox genes exhibit spatial and temporal changes in expression during human skin development. J Invest Dermatol 1998;110:110–5.
36. La Celle PT, Polakowska RR. Human homeobox HOXA7 regulates keratinocyte transglutaminase type 1 and inhibits differentiation. J Biol Chem 2001;276: 32844–53.
37. Naora H, Yang YQ, Montz FJ, Seidman JD, Kurman RJ, Roden RB. A serologically identified tumor antigen encoded by a homeobox gene promotes growth of ovarian epithelial cells. Proc Natl Acad Sci U S A 2001; 98:4060–5.
38. Care A, Felicetti F, Meccia E, et al. HOXB7: a key factor for tumor-associated angiogenic switch. Cancer Res 2001;61:6532–9.
39. Merrill BJ, Gat U, DasGupta R, Fuchs E. Tcf3 and Lef1 regulate lineage differentiation of multipotent stem cells in skin. Genes Dev 2001;15:1688–705.
40. Capo-chichi CD, Roland IH, Vanderveer L, et al. Anomalous expression of epithelial differentiation-determining GATA factors in ovarian tumorigenesis. Cancer Res 2003;63:4967–77.
41. Kaufman CK, Zhou P, Pasolli HA, et al. GATA-3: an unexpected regulator of cell lineage determination in skin. Genes Dev 2003;17:2108–22.
42. Koh K, Rothman JH. ELT-5 and ELT-6 are required continuously to regulate epidermal seam cell differentiation and cell fusion in *C. elegans*. Development 2001; 128:2867–80.
43. Basson CT, MacRae CA, Schoenberg-Fejzo M, et al. Identification, characterization, and chromosomal localization of the human homolog (hES) of ES/130. Genomics 1996;35:628–31.
44. Sarkar S, Roy BC, Hatano N, Aoyagi T, Gohji K, Kiyama R. A novel ankyrin repeat-containing gene (Kank) located at 9p24 is a growth suppressor of renal cell carcinoma. J Biol Chem 2002;277:36585–91.
45. Ding M, Goncharov A, Jin Y, Chisholm AD. *C. elegans* ankyrin repeat protein VAB-19 is a component of epidermal attachment structures and is essential for epidermal morphogenesis. Development 2003;130: 5791–801.