Identification of Gene Expression Patterns in Superficial and Invasive Human Bladder Cancer¹

Thomas Thykjaer, Christopher Workman, Mogens Kruhøffer, Karin Demtröder, Hans Wolf, Lars Dyrgkjøt Andersen, Casper Møller Frederiksen, Steen Knudsen, and Torben F. Ørntoft²

Molecular Diagnostic Laboratory, Departments of Clinical Biochemistry [T. T., L. D. A., C. M. F., K. D., T. F. Ø.] and Urology [M. K., H. W.], Aarhus University Hospital, DK 8200 Skejby, Aarhus, Denmark; and Center for Biological Sequence Analysis, Technical University of Denmark, DK 2100 Lyngby, Denmark [C. W. S. K.]

ABSTRACT

Multiple transcriptional events take place when normal urothelium is transformed into tumor tissue. These can now be monitored simultaneously by the use of oligonucleotide arrays, and expression patterns of superficial and invasive tumors can be established. Single-cell suspensions were prepared from bladder biopsies (36 normal, 29 tumor). Pools of cells were made from normal urothelium and from pT² grade I and II and pT² grade III and IV bladder tumors. From these suspensions, and from 10 single-tumor biopsies, labeled cRNA was hybridized to oligonucleotide arrays carrying probes for 6500 genes. The obtained expression data were sorted according to a weighting scheme and were subjected to hierarchical cluster analysis of tissues and genes. Northern blotting was used to verify the array data, and immunohistology was used to correlate between RNA and protein levels. Hierarchical clustering of samples correctly identified the stage using both 4076 genes and a subset of 400 genes covarying with the stages and grades of tumors. Hierarchical clustering of gene expression levels identified several stage-characteristic, functionally related clusters, encoding proteins that were related to cell proliferation, oncogenes and growth factors, cell adhesion, immunology, transcription, proteinases, and ribosomes. Northern blotting correlated well with array data. Immunohistology showed a good concordance between transcript level and protein staining. The study indicates that gene expression patterns may be identified in bladder cancer by combining oligonucleotide arrays and cluster analysis. These patterns give new biological insight and may form a basis for the construction of molecular classifiers and for developing new therapy for bladder cancer.

INTRODUCTION

Multiple molecular events take place when normal epithelia are transformed into tumor tissue and later acquire an invasive potential. Previously, these events have been examined at the single-gene level, however, the newly developed array technology has made it possible to simultaneously monitor thousands of genes during tumor evolution and progression (1). To generate meaningful information from such large datasets, new methods have been developed that cluster closely related genes or tissues based on their expression profiles (2, 3). The purpose is two-fold: to identify genes having a coordinated expression change and thereby obtain insight into possible regulatory patterns (4–7), and to classify biological samples based on expression (5, 8). The latter may then form a basis for building of molecular class predictors, potentially useful for tumor staging and prediction of clinical outcome as recently demonstrated (5, 9, 10). Two recent studies (3, 4) that used two-way clustering of thousands of genes were able to separate biopsies into those originating from cancerous and noncancerous tissue, and cell lines from in vivo tumors. Both studies was it possible to identify clusters of functionally related genes, some associated with proliferation, IFN-regulation, ribosomes or stromal components like lymphocytes, and smooth muscle cells.

The bladder cancer disease is characterized by two basically different disease courses: one in which patients suffer from multiple recurrences of superficial tumors, and one in which the patients' primary tumor is muscle invasive (11, 12). The group with superficial recurrences includes patients with the relatively benign superficial pT² tumors (~5% of which will progress to muscle invasive cancer), and patients with superficially invasive pT² tumors (~25% of which will progress to muscle invasive disease; Ref. 13). At present, no marker can predict which of the three disease courses a patient with a superficial tumor will follow: no recurrence, superficial recurrence, or invasive cancer. With the present study, we have aimed at identifying gene clusters that characterize the superficial and the muscle invasive diseases, as the first step in a detailed gene expression-based characterization of bladder cancer.

DNA arrays have been used primarily on relatively homogeneous biological samples such as yeast cultures, cell lines, and leukemias. The application on complex mixtures of cells such as tissue biopsies may create data that are more difficult to interpret because of the many different cell types providing signals (3). One approach to this problem is immunohistological or in situ hybridization detection of the cell type providing a specific signal, as demonstrated with the STAT1 gene product in breast cancer (4). We used immunohistology to demonstrate the cellular origin of eight proteins and to correlate the level of protein to the level of transcript detected on the array. Furthermore, to reduce the number of transcripts from stromal tissue and to enrich for tumor cells, tumors were disintegrated into single-cell suspensions by a method previously used for preparation of bladder tumor cells for flow cytometry (14, 15).

Two different expression measures, based on fold change and numerical difference between transcript level in tumor and normal urothelium, were tested for the clustering of tissues and genes. Of these, fold change proved to be the most informative. With both methods, however, several important functionally related clusters of genes were identified. These include clusters related to cell proliferation, oncogenes and growth factors, cell adhesion, immunology, transcription, and proteinases, as well as a large ribosomal gene cluster.

MATERIALS AND METHODS

Material. Thirty-nine bladder tumor biopsies (29 formed the first tumor set, 10 the second) were sampled from patients after the removal of the necessary amount of tissue for routine pathology examination. Grading according to Bergkvist et al. (16) was made by one pathologist. In the first hypothesis-generating tumor set, single-cell suspensions were made by immediately disintegrating biopsies on ice with a scalpel and a syringe followed by filtering through a 100 µm filter, as described previously (14, 15). The cells were inspected and counted under a microscope, and if RBCs occurred frequently, the sample was discarded to reduce signals from leukocytes in peripheral blood. When pools were made, RNA corresponding to a similar
number of cells was used from each of six tumors all having the same stage and grade. From the first tumor set, a total of four tumor pools were made (pT1 grade I pool, pT1 grade II pool, pT2+ grade III pool, and pT2+ grade IV pool), and the single tumors examined were 335 (stage pT1 grade I), 837 (pT1 grade II), 901 (pT1 grade III), 320 (pT1 grade III), and 713 (pT2+ grade III). For generation of a second dataset to test the reproducibility of the first set on biopsies, the tumor biopsies examined were 709, 928, 930, 934, 968 (pT1 grade II), and 875, 937, 1005, 1078, 1133 (pT2+ grade III).

Cells from normal bladder mucosa biopsies from 36 patients with prostate hyperplasia or urinary incontinence were pooled to obtain a normal urothelial reference. Informed consent was obtained in all of the cases, and protocols were approved by the local scientific ethical committee.

### RNA Preparation

Total RNA was isolated using the RNAzol B RNA isolation method (WAK-Chemie Medical GmbH). Poly(A)+ RNA was isolated by an oligo-dT selection step (Oligotex mRNA kit; Qiagen). One µg of mRNA was used as starting material for the cDNA preparation. The first and second strand cDNA synthesis was performed using the SuperScript Choice System (Life Technologies) according to the manufacturer’s instructions, except using an oligo-dT primer containing a T7 RNA polymerase promoter site. Labeled cRNA was prepared using the Megascript In Vitro Transcription kit (Ambion). Biotin-labeled CTP and UTP (Enzo) were used in the reaction together with unlabeled NTPs. After the IVT reaction, the unincorporated nucleotides were removed using RNeasy columns (Qiagen).

### Array Hybridization and Scanning

On the basis of previously published methods (1, 17–20), 10 µg of cRNA was fragmented at 94°C for 35 min in a fragmentation buffer containing 40 mM Tris-acetate (pH 8.1), 100 mM KCl, and 30 mM MgOAc. Prior to hybridization, the fragmented cRNA was combined in a 6× SSPE-T hybridization buffer (1 mM NaCl, 10 mM Tris (pH 7.6), and 0.005% Triton) was heated to 95°C for 5 min and subsequently to 40°C for 5 min before loading onto the Affymetrix probe array cartridge. The probe array was then incubated for 16 h at 40°C constant rotation (60 rpm). The washing and staining procedure was performed in the Affymetrix Fluidics Station. The probe array was exposed to 10 washes in 6× SSPE-T at 25°C followed by 4 washes in 0.5× SSPE-T at 50°C. The biotinylated cRNA was stained with a streptavidin-phycocerythrin conjugate (10 µg/mL; Molecular Probes, Eugene, OR) in 6× SSPE-T for 30 min at 25°C followed by 10 washes in 6× SSPE-T at 25°C. The probe arrays were scanned at 560 nm using a confocal laser-scanning microscope with an argon ion laser as the excitation source (Hewlett Packard GeneArray Scanner G2500A). The readings from the quantitative scanning were analyzed by the Affymetrix Gene Expression Analysis software. For comparison from array to array, these were scaled to a global intensity of 150, as published previously (17, 19, 20).

### Northern Blotting

Total RNA, 0.5–4 µg per lane, was separated in 1.5% agarose-formaldehyde gels, transferred onto Zeta-Probe nylon membrane (Bio-Rad) by positive pressure (Posiblotter; Stratagene) and immobilized by baking.

### Immunohistochemistry

Four-µm sections were cut from the paraffin-embedded tissue blocks, mounted, and deparaffinized by incubation at 80°C for 10 min, followed by immersion in heated oil (Etsisol 312; Estichem A/S, Copenhagen, Denmark) at 60°C for 10 min and rehydration. Antigen retrieval was achieved in TEG buffer using microwaves at 900 W. The tissue sections were cooled in the buffer for 15 min before a brief rinse in tap water. Endogenous peroxidase activity was blocked by incubating the sections with 1% H2O2 for 20 min, followed by three rinses in tap water, 1 min each. Then the sections were soaked in PBS buffer for 2 min. The next steps were modified from the descriptions given by Oncogene Science Inc., in the Mouse Immunohistochemistry Detection System (XHIC01; UniTect, Uniondale, NY). Briefly, the tissue sections were incubated overnight at 4°C with primary antibody (against β2-microglobulin (Dako), Cytokeratin 8, Cystatin C (both from Europa), junB, CD59, E-cadherin, apo-E, Cathepsin E, (all from Santa Cruz), followed by three rinses in PBS buffer, 5 min each. Afterward, the sections were incubated with biotinylated secondary antibody for 30 min, rinsed three times with PBS buffer, and subsequently incubated with avidin-biotinylated horseradish peroxidase complex (AECl) for 30 min, followed by three rinses in PBS buffer. Staining was performed by incubation with 3-aminoethylcarbazole for 10 min. The tissue sections were counterstained with Mayer’s hematoxylin, washed in tap water for 5 min, and mounted with glycerol-gelatin. Positive and negative controls were included in each staining round with all of the antibodies.

### Cluster Analysis

The scaled AvgDif measures (1) calculated by Affymetrix software were extracted from each array assay. Only the 4067 genes scored as present in at least one of the tissues were considered. For calculation, all of the AvgDif measures below 20 were set to 20. For each tumor tissue and each gene, the AvgDif from normal urothelium was either subtracted, to define the difference, or divided, and natural logarithm applied to define the “log-fold” relative measure.

For tissue clustering, the relative expression measures for each tumor tissue (log-fold or difference) were used to cluster tissues by a hierarchical method using Euclidean distance between tissues X and Y (average-linkage, agglomerative hierarchical clustering (UPGMA), our UNIX implementation of the algorithm described in Ref. 2).

Euclidean distance was calculated as follows:

$$d(X, Y) = \sqrt{\sum_{i=1}^{n} (x_i - y_i)^2}$$

where $x_i$ and $y_i$ are the component genes of the expression vectors $X, Y$; and $n$ is the dimensionality of $X, Y$ (number of genes).

Tissue dendrograms were constructed with the PHYLIP program DRAWTGRAM from the newick tree files generated by the clustering program. As a new approach for gene clustering, a weighting scheme for the seven observed grades of cancer was used to select 200 genes positively covarying and 200 genes negatively covarying to increasing stage. Only genes with an AvgDif of ≥100 in at least one tissue were used for the selection of the 400 covarying genes. An integer value between 1 and 8 was assigned to each stage, incrementing by one for each grade within stages and incrementing by 2 for increasing stage.

Covariance measure between gene A and the increasing stage vector (or gene) B:

$$\text{Covariance}(A, B) = \frac{\sum_{i=1}^{n} (a_i - \text{mean}(A))(b_i - \text{mean}(B))}{n - 1}$$

where $n$ is the dimensionality of vectors $A$ and $B$ (number of tissues). Pearson correlation distance was also tested for this ranking but was found to be too sensitive to the specific shape of the arbitrary weight vector chosen. Furthermore, Pearson correlation distance selected vectors with small overall fold change (small vector length). The covariance factor for the top 200 positively covarying genes ranged from 5.7 down to 1.6, whereas the 200 most negatively covarying genes ranged from −6.3 up to −1.8.

For gene clustering the same hierarchical method (average linkage agglomerative hierarchical clustering) as used for tissue clustering was applied, but in this case a normalized Euclidean distance (vector angle) was used to cluster the top 400 positively and negatively covarying genes for both relative expression measures.

The angle between expression vectors $A$ and $B$:

$$d(A, B) = \cos^{-1}\left(\frac{\sum_{i=1}^{n} (a_i b_i)}{(\text{length}(A) \text{length}(B))}\right)$$

where $\text{length}(X) = \sqrt{\sum_{i=1}^{n} (x_i)^2}$

1 The abbreviations used are: pT2+, stages pT2, pT1, and pT1+, log-fold, natural logarithm of fold changes; AvgDif, average difference between perfect match probes and mismatch probes on array.
Gene dendrograms were constructed by the same method as the tissue dendrograms.

RESULTS

We performed scaling to a global chip intensity of 150 units per probe set. The scaling made it possible to compare individual experiments with each other. To verify the reproducibility, double determinations were made in selected cases and showed a good correlation (Fig. 1A).

A scatter plot of the noninvasive pT1 grade 1 tumor and the invasive highly abnormal grade IV pT2+ tumor showed a minor subtraction of the gene transcripts to deviate much from those in the normal urothelium. The large majority of transcripts were within a narrow range in both tumors and normal urothelium (Fig. 1, B and C). The number of deviating genes was higher in the most abnormal tumor. With the purpose of confirming the array data with another method, we performed Northern blotting of four genes on the same samples of RNA as used for array hybridization. A standardized amount of RNA was run in each lane, followed by blotting with a labeled RNA probe, and quantitation of the obtained band (Fig. 2). The oligonucleotide array and the Northern blot gave similar results (Fig. 2) corresponding to previous studies (19, 20).

Cluster Analysis. The level of a gene transcript in different tumors can be thought of as a pattern that can be related to patterns of other gene transcripts. If the expression of one gene is very similar to the expression of another gene in several samples, they will have a high similarity measure. Gene expression patterns (vectors) that are similar are grouped together in clusters by the methods described in Ref. 2. Briefly, clustering is done by finding the most similar pair of vectors and merging them. The merged pair is replaced by a single vector that is the average of the two, and the distances from the new cluster vector to all of the other vectors are calculated. Because clusters can represent many genes, the averaging of cluster vectors is weighted by the size of the respective merged clusters or singletons. In this way, the cluster vector maintains the average of all of the vectors it represents (the cluster centroid). The same procedure can be applied to cluster
tissues based on expression levels over the different genes considered. We based clustering analysis on either the 4067 transcripts being scored as present in at least one of the samples, or based on those 400 transcripts that covaried best with a simple weighting scheme adding increasing values to increasing grades of atypia and sequentially increasing from stage to stage (pTa grade I, 1; pT grade II, 2; pT grade III, 3; pT grade III, 5; pT grade III, 7; pT grade IV, 8). Without a weighting scheme, the gene clusters formed lacked functionally related genes, a finding also seen when two-way clustering was used (data not shown).

**Tissue Clusters.** Different algorithms, applied to either log-fold change or differences in transcript levels across the different samples, were applied to all of the transcripts or only those covarying with the weighting scheme (Fig. 3). The log-fold-change based method, applied to the covarying genes correctly, identified the relationship between the samples (Fig. 3A), clustering the superficial pools close together and far from the invasive pools. The use of the full dataset of 4076 genes led to a reduced separation of superficial and invasive tumors (Fig. 3, C and D), as did the use of the average difference method (Fig. 3, B and D). However, separation of superficial and invasive tumors was still possible. We repeated the tissue clustering based on the 400 covarying genes on a second set of tumor biopsy samples (5 samples of pTa, and 5 of pT2+), and these were also correctly identified (Fig. 3E), which indicated that the 400 covarying genes could be of general use for classification. This is remarkable because the second set of tumors consisted of tissue biopsies and not single-cell suspensions.

On the basis of these dendrograms, we conclude that the log-fold-change-clustering algorithm works well, that the dataset obtained from the oligonucleotide arrays reflects the stage of the tumors, and that objective information on the stage and grade of a tumor can be obtained from a mathematical analysis of gene expression data. Because the set of 400 covarying genes showed optimal performance in tissue clustering, we used this set of genes when analyzing gene clusters.

**Gene Clusters.** The data obtained from gene cluster analysis are presented as colored images in which genes with similar expression patterns are clustered next to each other on the vertical axis, and the samples according to stage and grade on the horizontal axis (Fig. 4). The two different clusterings, log-fold change and difference, gave completely different clusters across the set of samples (Fig. 4). To validate the reproducibility of the selected 400 genes for log-fold-based gene clustering, we applied the same set of genes to a new set of samples according to stage and grade on the horizontal axis (Fig. 4). The three clusters (Fig. 4, second column from the left), a similar pattern of expression was detected in the set of biopsies, indicating the robustness of the selected 400 genes for clustering. The difference between suspensions of single cells and biopsies was reflected as a slightly higher variation from tumor to tumor in the biopsies compared with the suspensions.

In the log-fold-based cluster analysis, the top 200 positively covarying genes can be divided into five different clusters containing functionally related genes (Fig. 4A). The cluster shown at the top (Fig. 4A) contains genes related to cell proliferation such as cyclins A and E, PCTAIRE-1, and SWI/SNF. The next cluster mainly contains oncogenes and growth factors. Genes in both of these clusters are expressed at a level close to that seen in normal urothelium in superficial tumors (Fig. 4A, black) and increase in higher stage tumors (yellow). The three clusters at the lower part show a reduced expression level in the superficial tumors compared with normal (cyan) and then, in most cases, increase above the normal urothelial level in invasive tumors (shades of yellow). These clusters contain a set of immunologically related genes, including different MHCs and immunoglobulins, cancer-related genes like src-like kinase and Fas/Apo-1, and finally another immunologically related cluster at the bottom of the figure (Fig. 4A).

The 200 negatively covarying genes (Fig. 4B) could be divided into three different clusters based on log-fold change and function of the genes. The upper cluster contains genes related to cell adhesion like laminins, integrins, and P-cadherin (Fig. 4B). They all show a reduced level of expression in the invasive tumors as evidenced by the cyan coloring to the right. The small cluster in the middle contains four genes related to transcription, and, finally, the lowest cluster in the

---

For a list of standard gene names see data at: www.MDL.DK/sdata.html.
figure contains five proteinases, including cathepsin E (two different probe sets for the same gene) and metalloproteinase as well as a protease inhibitor. The lower clusters are characterized by an increase in level in superficial tumors (yellow) followed by a return to approximately normal levels in invasive tumors.

In the difference-based cluster analysis, the top 200 covarying genes that showed a positive covariance contained only a few clusters having a functional relation. The upper cluster (Fig. 4C) contained five genes related to cell proliferation like the microtubule-associated protein and oncoprotein 18/5athmin. The next cluster was a set of immunology-related genes such as MHC and LEKR-2. Both of these clusters showed an increased expression level in invasive tumors compared with normal urothelium. The cluster at the lower end showed a reduced level in superficial tumors and a return to normal or increased level in invasive tumors. This cluster contained many immunology-related genes like MHC, HLA, and immunoglobin genes. Finally, for genes that showed a negative covariance based on difference (Fig. 4D), most were attributable to the clustering of ribosomal genes. The upper and lower ribosomal clusters show genes that are up-regulated in expression in superficial tumors and down-regulated or unaltered in invasive tumors. The middle ribosomal cluster is generally expressed at a lower level than in normal urothelium. Other genes that seemed to cluster were a closely related set to single-cell suspensions.

Correlation between mRNA Levels Detected on Arrays and Protein Levels Detected by Immunohistochemistry. From the single tumors examined on arrays, tissue sections were cut and used for immunostaining. It was then possible to correlate the level of transcript measured on the chip to the extent of immunostaining for the translation product (Fig. 5, genes marked with * in Fig. 4). On the basis of the transcript levels, we selected a group of proteins for immunostaining that were supposed to show variation from sample to sample and that covered a broad range of expression levels. Finally, an antibody should be commercially available.

Several of the proteins were expressed not only by urothelial cells but also by leukocytes, endothelial cells, or histiocytes (Fig. 5). The level of protein identified by immunostaining, disregarding the cell type expressing the protein, seemed to correlate well with the transcript level measured on the microarray (Fig. 5).

DISCUSSION

The use of array technology for monitoring the expression of thousands of genes makes it possible to obtain a more complete understanding of the many events that characterize the different stages of a cancer disease. Furthermore, stage characteristic expression patterns may be used for molecular classification of tumors. The basis for this is reproducible and simultaneous measurement of transcripts, which we demonstrate is possible with oligonucleotide arrays, that showed a good correlation with Northern blots.

To obtain a general information on gene expression at superficial and invasive stages of bladder cancer, we chose to pool tumors representing different stages. In that way individual differences between samples were smoothed, and the most stage-characteristic
patterns strengthened. We believe it is necessary to seek general characteristic patterns, because these may later form the basis for the building of molecular classifiers (5). In addition to the pools, we analyzed single superficial and invasive tumors. These showed an expression pattern similar to the pools representing their stage, and clustered close to these when tissue clustering was performed. It was interesting to observe that the pT1 tumor clustered close to the pT2+ tumors, indicating that pT1 tumors could be more familiar with muscle-invasive tumors than with superficial benign tumors.

Another approach that we used was the preparation of single cell suspensions to reduce the stromal component of the biopsies. The single-cell suspensions showed less expression of smooth muscle- and connective tissue-related genes (data not shown). The single-cell suspensions were prepared from cooled biopsies immediately after surgery using a procedure previously used for the preparation of bladder tumors for flow cytometry (14, 15). The single-cell suspensions, furthermore, have the advantages that the cells can be inspected under the microscope to ensure the presence of >90% urothelial cells and only a minimum of peripheral blood contamination, tumor pools can be made based on similar numbers of cells from each tumor, and the RNA-preserving guanidinium thiocyanate, used for storage, immediately disrupts the cells and inactivates RNases.

To reduce the number of genes used for clustering, we filtered the expression data based on different cutoff levels (data not shown). However, we found this approach to be much less informative than using the top 10% of genes covarying with a simple weighting scheme. The weighting scheme applied more weight to samples with increasing grade of atypia as well as to samples having a high stage. The top 10% covarying genes were then analyzed by clustering the tissues in the first dimension, then genes in the other dimension. Organizing genes with a similar expression pattern into clusters identified a number of functionally related genes in several clusters. The most significant of these were obtained by examining log-fold change of expression, which clustered genes related to cell cycle, oncogenes and growth factors, immunology, cell adhesion, transcription, and proteinases into separate clusters. In summary, this pattern suggests a high level of protein synthesis in superficial papillomas having increased transcription factor and ribosomal levels as well as potential tissue-degrading properties obtained by proteinase up-regulation. In the invasive tumors, an increased level of cell cycle-related transcripts were observed, which might partly reflect the increased level of growth factor and oncogene transcripts also seen in these. A loss of cellular adhesion proteins was found in invasive tumors and may be related to tissue invasion and metastasis. The invading tumor cells seem to be a challenge to the immune system as reflected by an increase in immunology-related proteins. Why we find a relatively lower level of proteinases in invasive tumors compared with superficial tumors is at present unexplained.

Because of space limitations, we will comment only on a few of the clusters in the following. A closely related cluster of positively covarying genes, reaching maximum expression in invasive tumors, consisted of genes related to the cell cycle. Most of these were genes that will increase the cellular proliferation. Several genes normally up-regulated in the G1 phase and during entry into S phase were increased, such as cyclins A and E, c-myb (induces cyclin A1), hepatocyte growth factor, and MCM2/BM28 (21–25). Other genes related to the spindle checkpoint as TTK (26, 27), and sister chromatid separation as E2-C/InchH10 (28), peaking at late G2/M phase were also up-regulated, as was PCTAIRE-1 (29), which is a member of the CDK family with an unknown function.

The only negative regulator in this cluster was the SWI/SNF complex, which represses the activation function of E2F1 in cooperation with Rb (30, 31). These data indicate that all of the components needed to direct the cell through the cell cycle are up-regulated in invasive bladder tumors, probably accounting for the increased number of cells in S phase that was previously observed in many flow cytometry studies (32). These up-regulated proteins could be targets for inhibitors aiming at reducing the cell cycle and thereby working as cytostatic pharmaceuticals.

We identified a cell adhesion cluster in the log-fold analysis that showed down-regulation of the transcripts in invasive tumors. The cluster contained two laminin and two integrin genes, a laminin receptor, two members of the cadherin family (P-cadherin 3 and FAT tumor suppressor), epican, factor H homologue, and the mucin MGC-24. A reduced cell adhesion has been reported to accompany progression of various epithelial cancers. Although still controversial, it seems that self-adhesion is reduced in invasive cancers, whereas other adhesive properties may be up-regulated simultaneously (33). The reduced expression of laminins and integrins has previously been described in invasive cancer (34) and both α5 and β1 integrins showed a progressive reduction with increasing neoplastic transformation in colorectal carcinomas (35). P-cadherin has shown reduced expression in poorly differentiated gastric carcinomas compared with well differentiated (36), and the E-cadherin is lost in invasive bladder tumors (37). The FAT gene is a new member of the human cadherin superfAMILY that closely resembles a Drosophila gene essential for controlling cell proliferation during Drosophila development (38). Loss of gene function in Drosophila causes hyperplastic tumor-like overgrowth of larval imaginal discs (39). Epican is a heparan/chondroitin sulfate proteoglycan form of CD44 that creates self-aggregation and adherence to keratinocytes when transfected into fibroblasts (40).

Factor H is a multifunctional protein that acts as a complement regulator but that also has functions outside the complement system because it binds to the cellular integrin receptor (CD11b/CD18) and interacts with cell surface glycosaminoglycans (41). It has several homologues with ill-defined functions. Finally, the MGC-24 gene encodes a mucin (MUC18/MCAM/CD146) that leads to increased homotypic adhesion in vitro when transfected into melanoma cell lines (42). In conclusion, it seems that muscle-invasive bladder cancer is characterized by loss of expression of several adhesion proteins, some of which are still poorly examined in relation to cancer.

The ribosomal gene cluster that was identified when using the difference calculation corresponded to a similar cluster recently identified by a deterministic-annealing algorithm in a set of colorectal tissue samples (3). In that case, the ribosomal transcripts were low in normal tissue and high in the colon tumor tissue. In the bladder, a high level was also detected in superficial tumors, whereas invasive bladder cancers had a normal or reduced level of ribosomal proteins. Eight of the ribosomal proteins clustering in the colon tumor tissue were identical to those in the bladder and may form a set of transcripts or proteins that change in cancer development.

Apart from the functionally related genes, most clusters also contained genes without any obvious relation. Why they occurred together is at present unknown. Coregulation attributable to utilization of the same transcription factors is one possible explanation. We did examine whether it was attributable to co-localization at the same chromosomal region but found no support for that hypothesis.

The level of a transcript determined in a tumor biopsy does not provide any information on the cellular origin of that transcript. Many transcripts, such as housekeeping transcripts, are produced...
in the majority of cells, whereas others, such as immunoglobulins, are confined to certain cell types. We used immunohistochemistry to demonstrate the complex cellular origin of some of the transcripts. Because many thousands of transcripts are examined by array technology, it seems impossible to localize the cellular origin of each signal. That would be necessary if new information on the biology of urthelial tumor cells per se was sought; however, another approach would be to regard the tumor cells and their environment as providers of equally important information. For example, the increase in immune-related transcripts in invasive tumors might provide information on the stage of the disease. Such a global picture on tumor markers may provide important classifiers and predictive markers in the near future. Interestingly, a good correlation between mRNA level and protein level was observed in eight selected cases. Because the protein, not the mRNA, most often is the functional effector, this is of obvious importance.

The level of a transcript gives no information on the functional status of the translation product, the protein, that could be damaged by post-translational modification. The level of a transcript gives no information on the functional status of the translation product, the protein, that could be damaged by post-translational modification. The level of a transcript gives no information on the functional status of the translation product, the protein, that could be damaged by post-translational modification.

ACKNOWLEDGMENTS

We thank Thomas Gingrich, Christine Harrington, Søren Brunak, and Anders Krogh for helpful discussions.

REFERENCES

Identification of Gene Expression Patterns in Superficial and Invasive Human Bladder Cancer

Thomas Thykjaer, Christopher Workman, Mogens Kruhøffer, et al.

Cancer Res 2001;61:2492-2499.

Updated version
Access the most recent version of this article at:
http://cancerres.aacrjournals.org/content/61/6/2492

Cited articles
This article cites 40 articles, 19 of which you can access for free at:
http://cancerres.aacrjournals.org/content/61/6/2492.full.html#ref-list-1

Citing articles
This article has been cited by 23 HighWire-hosted articles. Access the articles at:
http://cancerres.aacrjournals.org/content/61/6/2492.full.html#related-urls

E-mail alerts
Sign up to receive free email-alerts related to this article or journal.

Reprints and Subscriptions
To order reprints of this article or to subscribe to the journal, contact the AACR Publications Department at pubs@aacr.org.

Permissions
To request permission to re-use all or part of this article, contact the AACR Publications Department at permissions@aacr.org.