

# Gene Expression in the Urinary Bladder: A Common Carcinoma *in Situ* Gene Expression Signature Exists Disregarding Histopathological Classification

Lars Dyrskjød,<sup>1</sup> Mogens Kruhøffer,<sup>1</sup> Thomas Thykjaer,<sup>1</sup> Niels Marcussen,<sup>2</sup> Jens L. Jensen,<sup>3</sup> Klaus Møller,<sup>4</sup> and Torben F. Ørntoft<sup>1</sup>

<sup>1</sup>Molecular Diagnostic Laboratory, Department of Clinical Biochemistry, Aarhus University Hospital, Skejby, Aarhus N; <sup>2</sup>University Institute of Pathology, Aarhus University Hospital, Aarhus C; <sup>3</sup>Department of Theoretical Statistics, Department of Mathematical Sciences, Ny Munkegade, Aarhus C; and <sup>4</sup>Department of Urology, Aarhus University Hospital, Skejby, Aarhus N, Denmark

## ABSTRACT

The presence of carcinoma *in situ* (CIS) lesions in the urinary bladder is associated with a high risk of disease progression to a muscle invasive stage. In this study, we used microarray expression profiling to examine the gene expression patterns in superficial transitional cell carcinoma (sTCC) with surrounding CIS (13 patients), without surrounding CIS lesions (15 patients), and in muscle invasive carcinomas (mTCC; 13 patients). Hierarchical cluster analysis separated the sTCC samples according to the presence or absence of CIS in the surrounding urothelium. We identified a few gene clusters that contained genes with similar expression levels in transitional cell carcinoma (TCC) with surrounding CIS and invasive TCC. However, no close relationship between TCC with adjacent CIS and invasive TCC was observed using hierarchical cluster analysis. Expression profiling of a series of biopsies from normal urothelium and urothelium with CIS lesions from the same urinary bladder revealed that the gene expression found in sTCC with surrounding CIS is found also in CIS biopsies as well as in histologically normal samples adjacent to the CIS lesions. Furthermore, we also identified similar gene expression changes in mTCC samples. We used a supervised learning approach to build a 16-gene molecular CIS classifier. The classifier was able to classify sTCC samples according to the presence or absence of surrounding CIS with a high accuracy. This study demonstrates that a CIS gene expression signature is present not only in CIS biopsies but also in sTCC, mTCC, and, remarkably, in histologically normal urothelium from bladders with CIS. Identification of this expression signature could provide guidance for the selection of therapy and follow-up regimen in patients with early stage bladder cancer.

## INTRODUCTION

Cancer of the urinary bladder is among the five most common malignancies worldwide and is one of the most prevalent cancer diseases in Western countries (1). About 80% of the patients present initially with superficial disease, which comprise Ta tumors located in the mucosa only, submucosa invasive T<sub>1</sub> tumors, and carcinoma *in situ* (CIS) lesions (flat, high-grade dysplasia lesion). The patients presenting with Ta and T<sub>1</sub> tumors experience frequent tumor recurrences (50–70%) and to a lesser extent disease progression to a muscle invasive stage (10–30%; Ref. 2). The patients presenting with isolated or concomitant CIS lesions have a high risk of disease progression to a muscle invasive stage (3). The CIS lesions may have a widespread manifestation in the bladder (field disease) and are believed to be the most common precursors of invasive carcinomas (4, 5). The ability to predict which tumors are likely to recur or progress would have great impact on the clinical management of patients with superficial disease, because it would be possible to treat high-risk

patients more aggressively (*e.g.*, radical cystectomy or adjuvant therapy). No clinically useful markers exist currently that identify these patients. Although many prognostic markers have been investigated, the most important prognostic factors are still disease stage, grade of atypia, and especially the presence of areas with CIS (6–8). The golden standard for detection of CIS is urine cytology and histopathologic analysis of a set of selected site biopsies taken during routine cystoscopy examinations. However, these procedures are not sufficiently sensitive and attempts to improve detection of a CIS lesion with, for example, 5-aminolevulinic acid fluorescence imaging exist (9).

Complete genome screenings for diagnostic and prognostic markers using DNA microarray technology have demonstrated improvements recently in the ability to diagnose and predict disease outcome in cancer patients (10–15). In a recent study, we found a large difference in gene expression patterns between superficial transitional cell carcinoma (sTCC) with and without surrounding CIS using microarrays (13). Superficial tumors with surrounding CIS lesions (4 cases) showed, in most cases, gene regulations similar to those observed in muscle invasive tumors. This indicated that transitional cell carcinoma (TCC) adjacent to CIS may share the more aggressive genotype associated with the CIS lesions, probably because of the mono- or oligoclonal nature of many bladder tumors (16). In the present project, we used a much larger material for microarray expression profiling of the expression patterns associated with sTCC with surrounding CIS compared with sTCC with no surrounding CIS and to muscle invasive carcinomas (mTCC). We compared our results with the gene expression patterns found in normal bladder biopsies and in biopsies from cystectomy specimens with CIS lesions. Our results showed a large difference in expression patterns between TCC with and without surrounding CIS. We found expression similarities among TCC with surrounding CIS, biopsies harboring CIS, and invasive carcinomas. Furthermore, the expression patterns identified in the CIS lesions were also present in histologically normal biopsies adjacent to the CIS lesions. The present data support the value of microarray-based gene expression signatures because these identify clinically important cellular properties.

## MATERIALS AND METHODS

**Clinical Samples.** Five different groups of clinical samples were used in this study. The first group contained 15 tumor biopsies from sTCC without surrounding CIS. Samples were selected based on the following criteria: (a) Ta tumors with no CIS in selected site biopsies at any visit; and (b) no previous muscle invasive tumor. The second group contained 13 tumor biopsies from sTCC with surrounding CIS. Samples were selected based on the following criteria: (a) Ta or T<sub>1</sub> tumors with CIS in selected site biopsies at any visit; and (b) no previous muscle invasive tumors. Nine samples of Ta and 2 samples of T<sub>1</sub> were obtained from 11 patients. The third group contained 13 tumor biopsies from mTCC. The fourth group contained 9 biopsies of normal bladder mucosa from patients without a bladder cancer history. The fifth group contained 10 biopsies from 5 cystectomy specimens. One histologically normal biopsy and 1 biopsy with CIS from the 5 different cystectomy specimens were obtained. A grid was placed in the bladder for orientation, and biopsies were taken from eight positions covering the bladder surface. At each position, 3

Received 11/18/03; revised 3/17/04; accepted 3/29/04.

**Grant support:** The Karen Elise Jensen Foundation, The Danish Cancer Society, the John and Birthe Meyers Foundation, The University and County of Aarhus, and the Danish Research Council.

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.

**Requests for reprints:** Torben F. Ørntoft, Molecular Diagnostic Laboratory, Department of Clinical Biochemistry, Aarhus University Hospital, Skejby, DK-8200 Aarhus N, Denmark. Phone: 45-89495100; Fax: 45-89496018; E-mail: orntoft@kba.sks.au.dk.

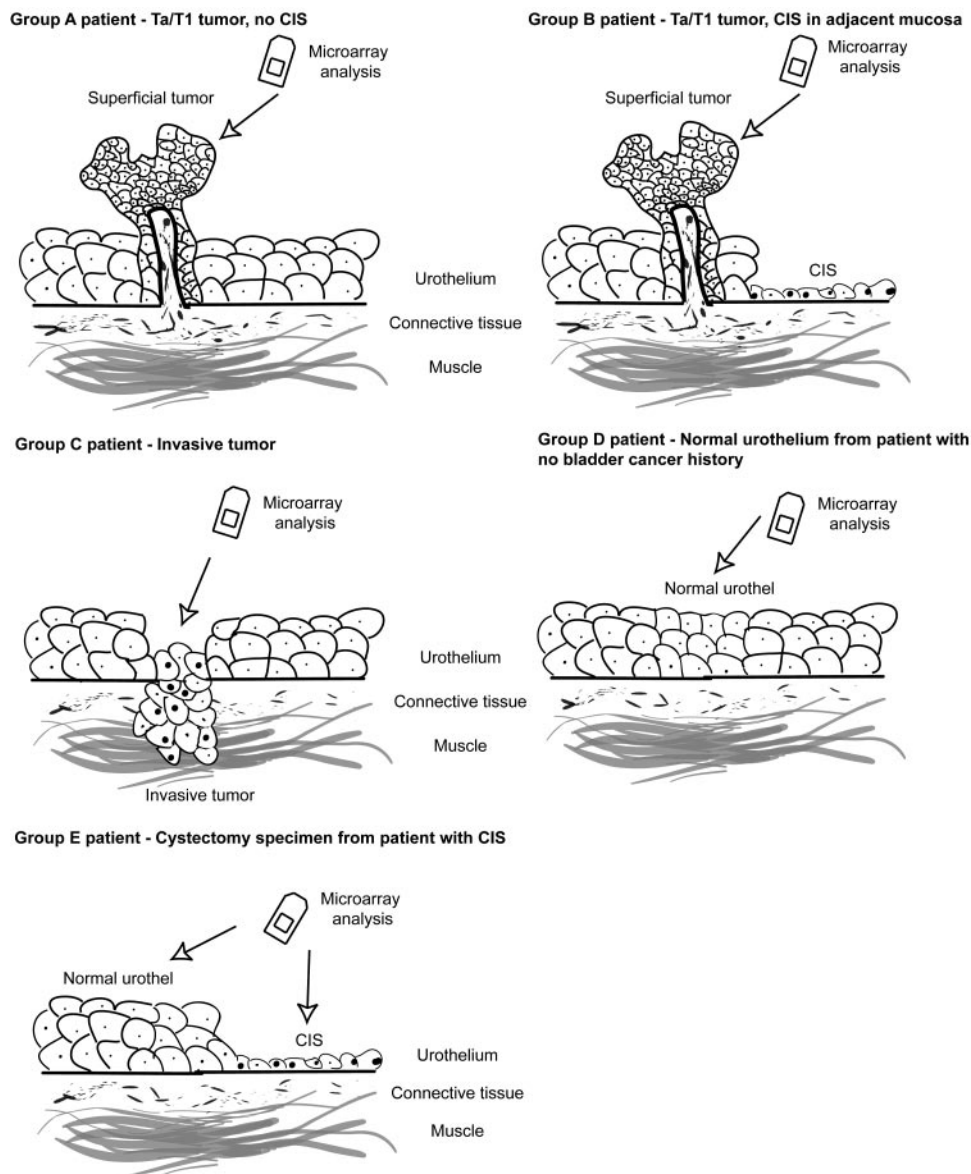


Fig. 1. Schematic overview of the different tissue samplings performed in the study.

biopsies were taken, 2 for pathological examination and 1 in between these for RNA extraction for microarray expression profiling. Samples used for RNA extraction were assumed to have CIS if CIS was detected in both adjacent biopsies. The histologically normal samples were assumed to be normal if both adjacent biopsies were normal. Histopathology scoring of CIS was defined as lesions with dysplasia grade III. The resected tumor material was randomly divided into two halves; one half was used for RNA extraction, and another half was used for histology. Fig. 1 shows an overview of the different groups of tissue used for expression profiling in this study.

All of the samples were obtained directly from surgery after removal of tissue for routine pathological examination. The samples were submerged immediately in a guanidinium thiocyanate solution for RNA preservation and stored at  $-80^{\circ}\text{C}$ . Informed consent was obtained in all of the cases, and the protocols were approved by the Scientific Ethical Committee of Aarhus County.

**The cRNA Preparation, Array Hybridization, and Scanning.** Purification of total RNA, preparation of cRNA from cDNA, and hybridization and scanning were performed as described previously (13). The labeled samples were hybridized to Affymetrix U133A GeneChips.

**Expression Data Analysis.** After scanning, all of the data were normalized using the Robust Multi-array Analysis (RMA) normalization approach in the Bioconductor Affy package to the R project for statistical computing (17). Variation filters were applied to the data to eliminate nonvarying and, pre-

sumably, nonexpressed genes. For gene set 1, this was done by including only genes with a minimum expression  $>200$  in at least 5 samples and genes with maximum/minimum expression intensities  $\geq 3$ . The filtering for gene set 2 included genes with only a minimum expression of 200 in at least 3 samples and genes with maximum/minimum expression intensities  $\geq 3$ . Average linkage hierarchical cluster analysis was carried out using the Cluster software with a modified Pearson correlation as similarity metric (18). We used the TreeView software for visualization of the cluster analysis results (18). Genes were log-transformed, median-centered, and normalized to have equal variance before clustering. We used GeneCluster 2.0<sup>5</sup> for the supervised selection of markers and permutation testing. The algorithms used in the software are based on methods described previously (14, 19). Classifiers for CIS detection were built using the same methods as described previously (13). We used EASE software in the search for overrepresented functional categories within the gene clusters (20).

## RESULTS

We used high-density oligonucleotide microarrays for gene expression profiling of  $\sim 22,000$  genes in biopsies from 28 superficial

<sup>5</sup> Internet address: <http://www-genome.wi.mit.edu/cancer/software/genecluster2/gc2.html>.

Table 1 Clinical data on patient disease courses and results of molecular CIS<sup>a</sup> classification

Sample group <sup>b</sup>	Patient <sup>c</sup>	Previous tumors	Tumor analyzed	Subsequent tumors	CIS <sup>d</sup>	CIS classifier <sup>e</sup>
1	1060-1		Ta gr2	2 Ta	No	No CIS
1	1146-1		Ta gr2		No	No CIS
1	1216-1		Ta gr2		No	No CIS
1	1303-1		Ta gr2		No	No CIS
1	524-1		Ta gr2		No	No CIS
1	692-1		Ta gr2	2 Ta	No	No CIS
1	1264-1		Ta gr3	20 Ta	No	No CIS
1	1350-1		Ta gr3	1 Ta	No	No CIS
1	1354-1		Ta gr3	3 T1	No	No CIS
1	775-1		Ta gr3	1 Ta	No	No CIS
1	1066-1		Ta gr3	1 Ta	No	No CIS
1	1276-1		Ta gr3	2 T1	No	No CIS
1	1070-1		Ta gr3	1 Ta	No	No CIS
1	989-1		Ta gr3		No	No CIS
1	1482-1		Ta gr3	20 Ta	No	CIS
2	1345-2	1 T1	Ta gr3		Sampling visit	CIS
2	1062-2	1 Ta	Ta gr3	1 T1	Sampling visit	CIS
2	956-2	1 Ta	Ta gr3	1 Ta	Sampling visit	CIS
2	320-7	1 Ta, 2 T1	Ta gr3	2 Ta	Sampling visit	CIS
2	1330-1		Ta gr3		Sampling visit	CIS
2	602-8	5 Ta	Ta gr3	3 Ta	Sampling visit	CIS
2	763-1		Ta gr2	14 Ta	Sampling visit	CIS
2	1024-1		T1 gr3	2 Ta, 1 T1	Sampling visit	CIS
2	1182-1		Ta gr3	7 Ta	Subsequent visit	CIS
2	1093-1		Ta gr3	4 Ta, 1 T1	Subsequent visit	CIS
2	797-1		Ta gr3		Sampling visit	CIS
2	1337-1	1 Ta	T1 gr3		Sampling visit	CIS
2	1625-1		Ta gr2		Sampling visit	CIS
3	1015-1		T3b gr4		No	—
3	1385-1		T4a gr3		Sampling visit	—
3	1041-1		T4b gr3		No	—
3	1044-1		T4b gr3		ND	—
3	1055-1	1 Ta gr2	T3a gr3		No	—
3	1109-1		T2 gr3	1 T2-4	No	—
3	1124-1		T4a gr3	2 T2-4	No	—
3	1154-1		T3a gr3	1 Ta, 1 T2-4	No	—
3	1167-1	1 T2-4	T3b gr4	2 T2-4	ND	—
3	1178-1		T4b gr3		ND	—
3	1215-1		T4b gr3		ND	—
3	1271-1		T3b gr4		No	—
3	1321-1	1 T1	T3b gr?		ND	—

<sup>a</sup> CIS, carcinoma *in situ*; sTCC, superficial transitional cell carcinoma; ND, not determined.

<sup>b</sup> The tumor groups involved were sTCC without CIS (1), sTCC with CIS (2), and mTCC (3).

<sup>c</sup> The numbers indicate the patient number followed by a number indicating number of visit in the clinic.

<sup>d</sup> CIS in selected site biopsies in previous, present, or subsequent visits to the clinic.

<sup>e</sup> Molecular classification of the samples using 25 genes in cross-validation loops.

bladder tumors (13 tumors with surrounding CIS and 15 without surrounding CIS) and 13 invasive carcinomas. (See Table 1 for patient disease course descriptions.) Furthermore, expression profiles were obtained from 9 normal biopsies and from 10 biopsies from 5 cystectomy specimens (5 histologically normal biopsies and 5 biopsies with CIS). Fig. 2 shows representative histological compositions from each group of samples investigated. The scanning images obtained were quality checked for errors, and the gene expression data were generated and normalized using RMA (17). All of the samples were labeled and hybridized simultaneously in one session, and we did not observe any group-specific differences in overall array intensity (Supplementary Data, Fig. 1).<sup>6</sup> Histological examination of the superficial tumors used for microarray analysis did not show any systematic variation in content of inflammatory cells between tumors with or without surrounding CIS.

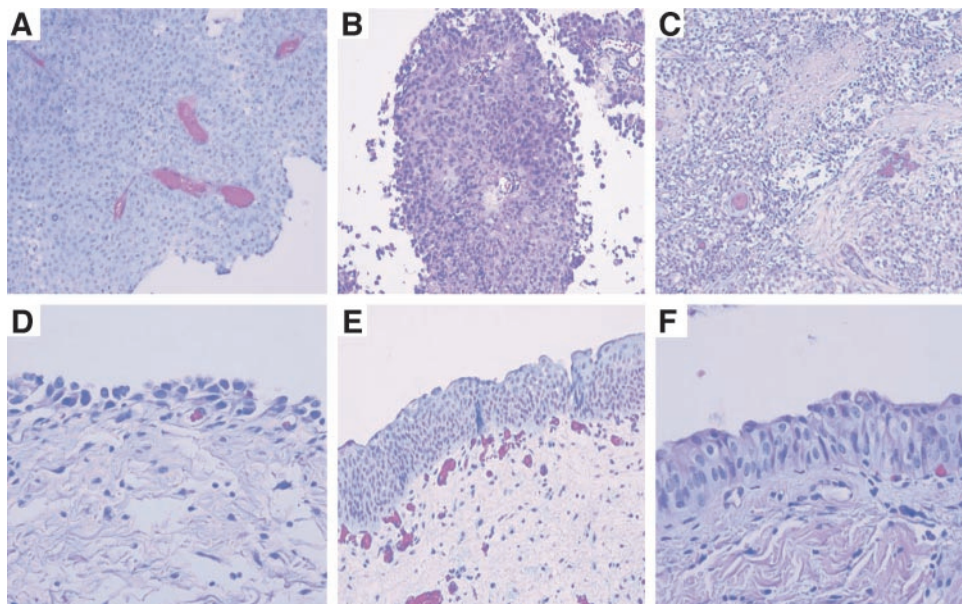
**Hierarchical Cluster Analysis.** After appropriate normalization and expression intensity calculations, we selected those genes that showed variation across the 41 TCC samples for additional analysis. The filtering produced a gene set consisting of 5491 genes (gene set 1), and two-way hierarchical cluster analysis was performed based on this gene set. The sample clustering showed a separation of the three groups of samples with only a few exceptions (Fig. 3A). The sTCC

with surrounding CIS clustered in the one main branch of the dendrogram, whereas the sTCC without CIS and the mTCC clustered in two separate sub-branches in the other main branch of the dendrogram. The only exceptions were that the mTCC samples 1044-1 and 1124-1 clustered in the CIS group and 2 TCC with CIS clustered in the sTCC without CIS group (samples 1330-1 and 956-2). The only sTCC without CIS that clustered in the CIS group was sample 1482-1. The distinct clustering of the tumor groups indicated a large difference in gene expression patterns.

Two-way hierarchical clustering (Fig. 3C) identified large clusters of genes characteristic for each tumor phenotype. Cluster 1 showed a cluster of genes down-regulated in cystectomy biopsies, sTCC with adjacent CIS, and some invasive carcinomas (Fig. 3C). No obvious functional relationship between the genes in this cluster was found (see “Materials and Methods”). Cluster 2 showed a tight cluster of genes related to immunology, and cluster 3 contained genes expressed mostly in muscle and connective tissue. Expression of genes in this cluster was observed in the normal and cystectomy samples, in a fraction of the sTCC with CIS, and in the invasive tumors. Cluster 4 contained genes up-regulated in the cystectomy biopsies, sTCC with adjacent CIS, and invasive carcinomas (Fig. 3C). This cluster included genes involved in cell cycle regulation, cell proliferation, and apoptosis. However, for most of the genes in this cluster, there is no apparent functional relationship either. Comparisons of chromosomal location

<sup>6</sup> Supplementary data is available online at <http://www.mdl.dk>.

Fig. 2. Stained tissue sections showing typical histological appearance of the different groups of samples used. A, superficial Ta tumor from a bladder with no surrounding carcinoma *in situ* (CIS; patient no. 1060-1). B, superficial Ta tumor from a bladder with surrounding CIS (patient no. 1330-1). C, muscle invasive T2+ tumor (patient no. 1015-1). D, biopsy with CIS lesion from a bladder removed by cystectomy (patient no. 1475; biopsy 16-4). E, biopsy with normal appearing urothelial cells from a bladder with CIS lesions removed by cystectomy (patient no. 1475; biopsy 16-1). F, biopsy from a normal bladder showing normal urothelium, the patient had no bladder cancer history.



of the genes in the clusters revealed no correlation between the observed gene clusters and chromosomal position of the identified genes. A positive correlation could have indicated chromosomal loss or gain or chromosomal inactivation by, for example, methylation of promoter regions.

To analyze additionally the impact of surrounding CIS lesions, we used the 28 superficial tumors only and created a new gene set consisting of 5252 varying genes (gene set 2). This new gene set included 4882 (93%) of the genes from gene set 1. Hierarchical cluster analysis of the tumor samples (Fig. 3B) based on the new gene set separated the samples according to the presence of CIS in the surrounding urothelium with only one exception ( $P < 0.000001$ ;  $\chi^2$ -test). Sample 1482-1 clustered in the sTCC with the CIS group; however, no CIS was detected in selected site biopsies during routine examinations of this patient. Tumor samples 1182-1 and 1093-1 did not have CIS in selected site biopsies in the same visit as the profiled tumor but showed this in later visits. However, the profile of these 2 superficial tumor samples already showed the adjacent CIS profile.

To validate the strength of the observed sample clustering, we applied different filtering criteria to gene set 1 and gene set 2 ( $SD \geq 200$ ). Minor changes were observed; however, the filtering did not introduce any overall changes to the observed clusters in Fig. 3, A and B (data not shown).

**Marker Selection.** To delineate the tumors with surrounding CIS from the tumors without surrounding CIS, we used *t* test statistics to select the 50 most up-regulated genes in each group. The relative expressions of these 100 genes together with the expressions of the genes in mTCC biopsies are shown in Fig. 4A. The CIS profile was identified in almost all of the mTCC samples. Permutation of the sample labels 500 times revealed that the 50 genes up-regulated in the CIS group are highly significantly differentially expressed and unlikely to be found by chance, because all of the markers were significant at a 5% confidence level. This means that the *t* test value for each of these genes was so high that similar high values were found in  $<5\%$  of 500 random data sets. The 50 genes up-regulated in the no-CIS group showed a poorer performance in the permutation tests, because these were not significant at a 5% confidence level (see Supplementary Data for details).<sup>6</sup> Fig. 4B shows the relative expression of the marker genes in 9 biopsies from normal bladder and 10 biopsies (5 histologically normal and 5 with CIS) from cystectomies

with CIS. The no-CIS profile was found in all of the normal samples. However, all of the histologically normal samples adjacent to the CIS lesions as well as the CIS biopsies showed the CIS profile.

To determine additionally the significance of the 100 genes selected, we performed a significance analysis of microarrays analysis of the expression data (21). When using a 5% false discovery rate in the significance analysis of microarrays analysis, we identified 62% of the genes in the no-CIS group and 100% of the genes in the CIS group.

**Construction of a Molecular CIS Classifier.** A classifier able to diagnose CIS from gene expressions in TCC or in bladder biopsies might increase the detection rate of CIS. Our first approach was to try to classify sTCC with or without CIS in the surrounding mucosa, based on tissue from the sTCC.

We build a CIS classifier as described previously (13) using cross-validation for determining the optimal number of genes for classifying CIS with the fewest errors. The best classifier performance (one error) was obtained in cross-validation loops using 25 genes (see Supplementary Data, Fig. 2);<sup>6</sup> 16 of these were included in 70% of the cross-validation loops, and these were selected to represent our final classifier for CIS diagnosis (Fig. 5A; Supplementary Data, Table 2).<sup>6</sup> Permutation analysis showed that 13 of these were significant at a 1% confidence level; the remaining 3 genes were above a 10% confidence level.

We explored additionally the strength of predicting surrounding CIS based on analysis of sTCC tissue. We built a classifier by randomly selecting half of the samples for training and using the other half for testing. Cross-validation was used again in the training of this classifier for optimization of the gene set for classifying independent samples. Cross-validation with 15 genes showed a good performance (see Supplementary Data, Fig. 3),<sup>6</sup> and 7 of these genes were included in 70% of the class-validation loops. These 7 genes classified the samples in the test set with one error only, sample 1482-1 ( $\chi^2$ -test;  $P < 0.002$ ). Only 2 of the genes were included also in the 16-gene classifier, which is understandable considering the many tests performed and the limitations in sample size. This classification performance is notable considering the few samples used for training the classifier.

The gene expression profiles of the 16-classifier genes in the 9 normal biopsies from patients without a bladder cancer history and in

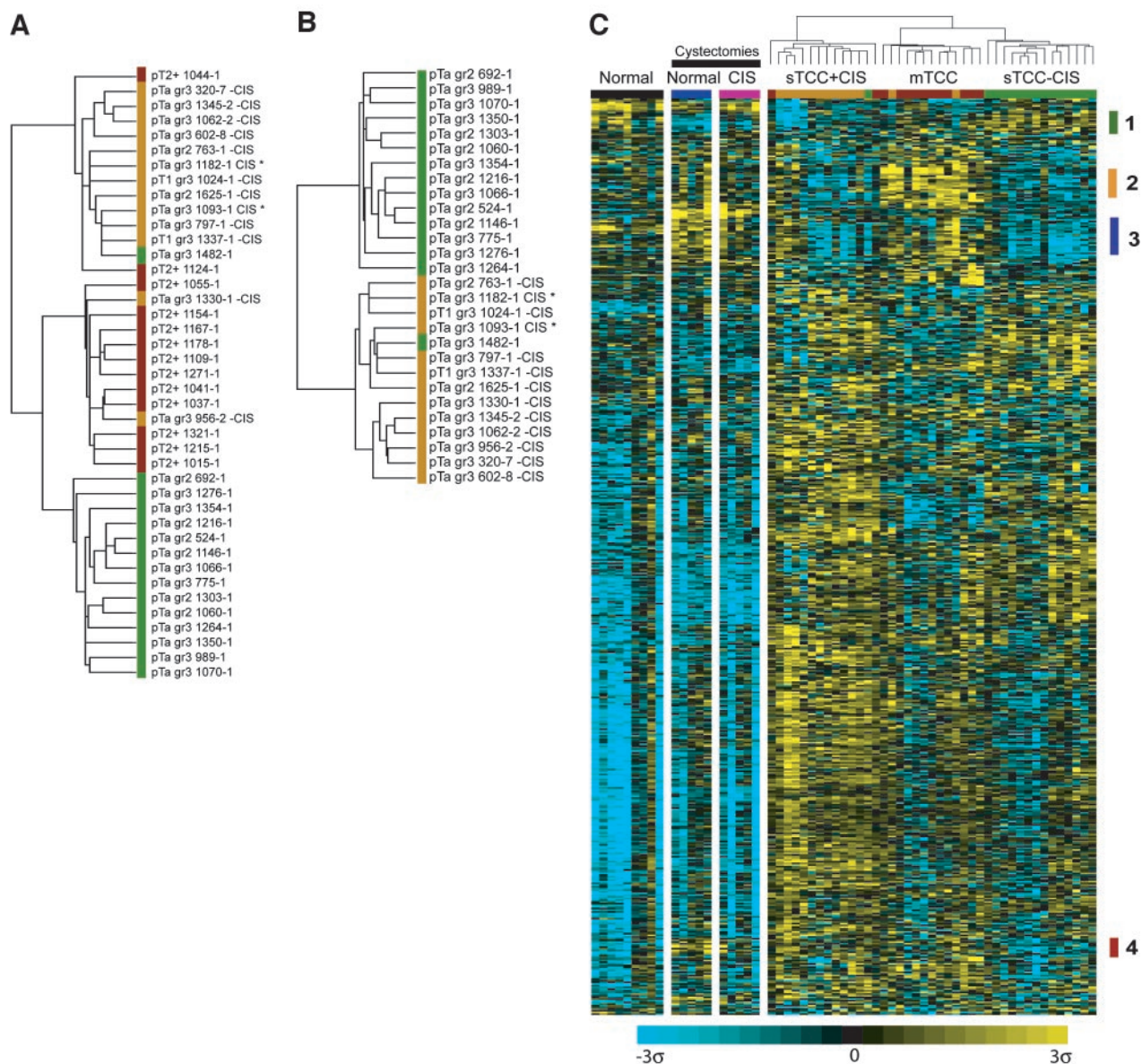


Fig. 3. Hierarchical cluster analysis of the gene expression in 41 transitional cell carcinoma (TCC), 9 normal samples, and 10 samples from cystectomy specimens with carcinoma *in situ* (CIS) lesions. *A*, cluster dendrogram of all 41 of the TCC tissues based on the expression of 5491 genes. \* denote samples from patients that show CIS at subsequent visits to the clinic. *B*, cluster dendrogram of all of the superficial transitional cell carcinoma (sTCC) tissues based on the expression of 5252 genes. *C*, two-way cluster analysis diagram of the 41 TCC tissues together with gene expressions in the normal and cystectomy samples (left columns). Each row represents a gene, and each column represents a tissue sample. Yellow indicates up-regulation compared with the median expression (black) of the gene, and blue indicates down-regulation compared with the median expression. The color saturation indicates degree of gene regulation in terms of SDs from the median expression as illustrated by the colorbar. The sidebars to the right of the diagram represent gene clusters 1–4; enlarged views of cluster 1 and 4 are shown (right) with all of the gene symbols listed.

the 10 biopsies (CIS and histologically normal samples) from cystectomy specimens are shown in Fig. 5B. The samples were separated based on the normalized gene expression values using hierarchical cluster analysis. The clustering separated the samples from cystectomies (both histologically normal biopsies and CIS biopsies) from the normal samples from patients with no bladder cancer history with only a few exceptions. Eight of the 10 biopsies from cystectomies were found in the one main branch of the dendrogram, and 8 of the 9 normal biopsies were found on the other main branch ( $\chi^2$ -test;  $P < 0.002$ ). The 16-gene CIS classifier was not able to recognize the CIS signature in these biopsies, probably because of the large difference in sample composition between the sTCC samples (papillomas) used for training the classifier and the cystectomy and normal biopsies (flat lesions), which contain large amounts of connective tissue and muscle cells.

We conclude that it is possible to detect a CIS expression signature in sTCCs and histologically normal mucosa from bladders having CIS lesions. Thus, analyzing the sTCC papilloma alone may provide information that makes the sampling of random biopsies unnecessary.

## DISCUSSION

Our present study demonstrates that sTCC with surrounding CIS have a notably different expression profile than sTCC without CIS. One of the most striking findings was that some of the gene expression patterns found in sTCC with surrounding CIS were similar to those found in CIS biopsies and in histologically normal samples adjacent to CIS lesions. These findings suggest that a CIS signature is present in general in the urothelium of these patients no matter whether the urothelial cells are organized as tumors, as flat lesions, or

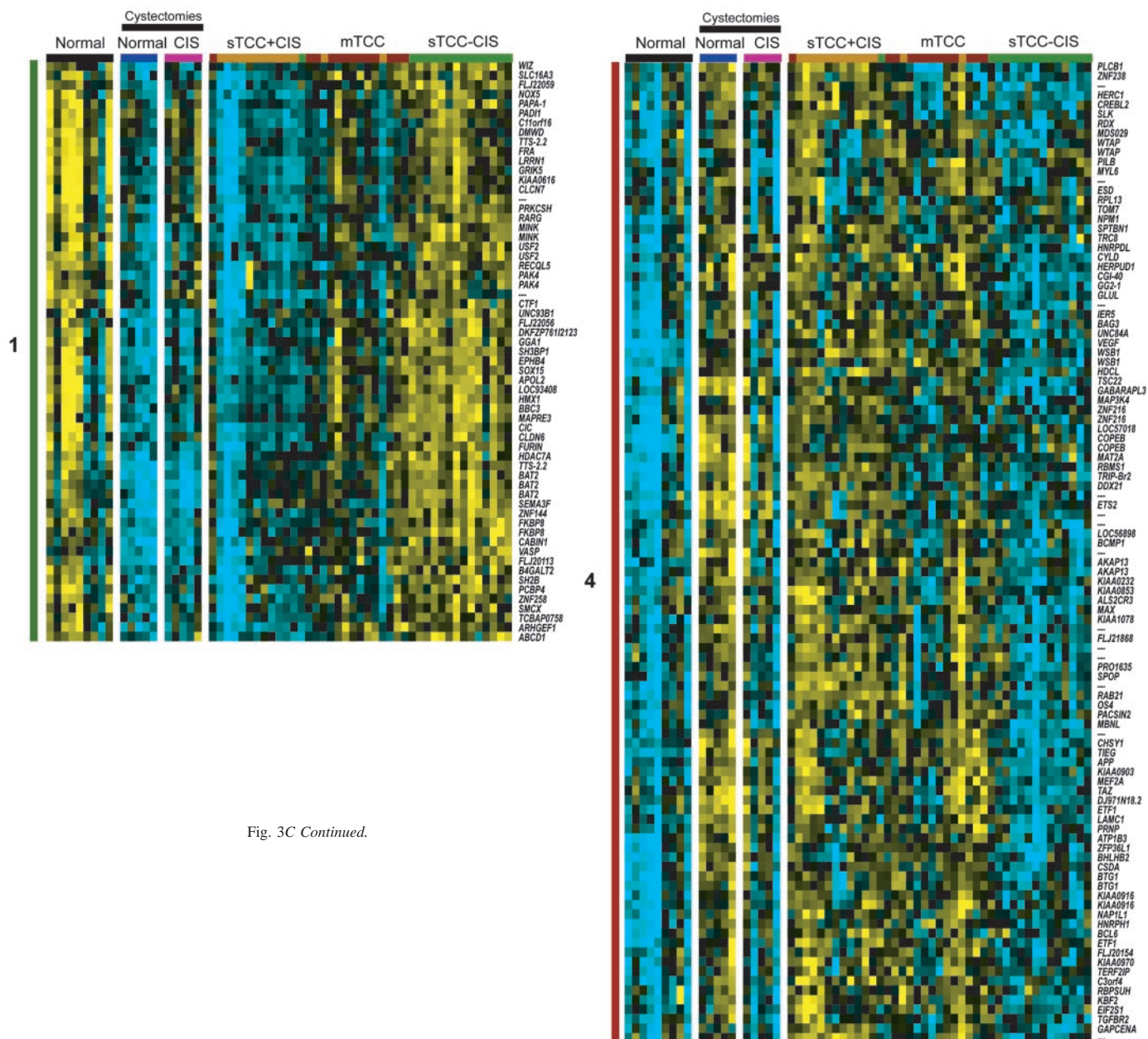


Fig. 3C Continued.

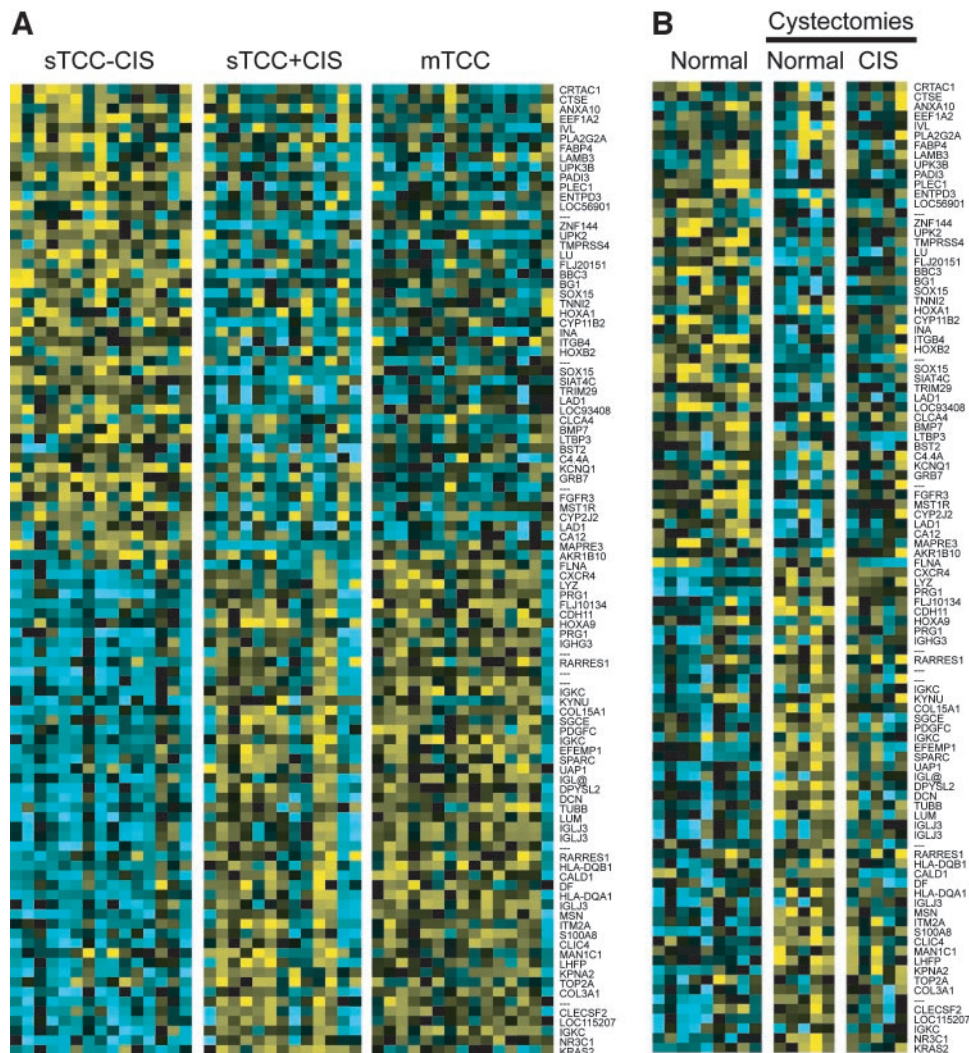
as histologically normal-appearing urothelium. This could be explained partly by the development of both TCC and CIS from the same precursor cells (16).

The observation that histologically normal areas in the bladder adjacent to CIS lesions harbored the same gene expression patterns as observed in the CIS lesions is remarkable. These findings are supported by observations made by Hartmann *et al.* (22), who found genetic alterations on chromosome 9 in normal samples adjacent to superficial papillary tumors. In addition, Steidl *et al.* (23) found chromosome copy number changes in normal biopsies adjacent to superficial tumors. Consequently, our and other studies suggest that apparently normal mucosa adjacent to neoplastic lesions may harbor some of the genetic changes associated with neoplasias. The mechanism for such alteration is not understood. A hypothesis could be that alterations in the underlying connective tissue are driving the expression changes in the cells surrounding the CIS lesion, due to tumor host communication during tumor development and progression (24). These changes may not be restricted to appear solely below the

carcinoma but are likely to span a larger area. Such a phenomenon may also help explain the frequently observed multifocality and recurrence patterns of mono- or oligoclonal bladder tumors. However, the tumors are separated both in time and location in the bladder, and, thus, an unknown effect seems to modify the tumor formation.

Delineation of the 100 best markers to separate TCC with CIS from TCC without CIS identified several interesting genes. Among the genes with higher expression in TCC without adjacent CIS were the *LAMB3* and *ITGB4* genes, which encode proteins involved in cell adhesion (25). This group also contained the *UPK2* gene, which encodes the bladder-specific uroplakin 2 protein. Uroplakin 2 is a marker of advanced-stage urothelial differentiation (26). The *FABP4* gene also showed elevated expression in the non-CIS group compared with the group with adjacent CIS. Fatty acid binding protein has been shown to be down-regulated in invasive bladder tumors (27). Lastly, the group contains the *FGFR3* gene, which has been shown to be mutated in sTCC with low recurrence rates (28). In the group of genes showing increased expression in the TCC with adjacent CIS, we found

Fig. 4. Delineation of the 100 best markers that separate superficial transitional cell carcinoma (sTCC) without carcinoma *in situ* (CIS) from sTCC with concomitant CIS. A, the 50 best up-regulated marker genes in sTCC without CIS are shown at top left, and the 50 best up-regulated marker genes in transitional cell carcinoma with CIS are shown at bottom right. The expressions of the 100 marker genes in muscle invasive carcinoma samples are shown at right. The gene symbols are listed to the right of the diagram. B, expression profiles of the 100 marker genes in 9 biopsies from normal bladder (left column), 5 histologically normal samples adjacent to CIS lesions (middle column), and 5 biopsies with CIS lesions detected (right column). The expression data for each gene were median-centered and normalized before visualization.

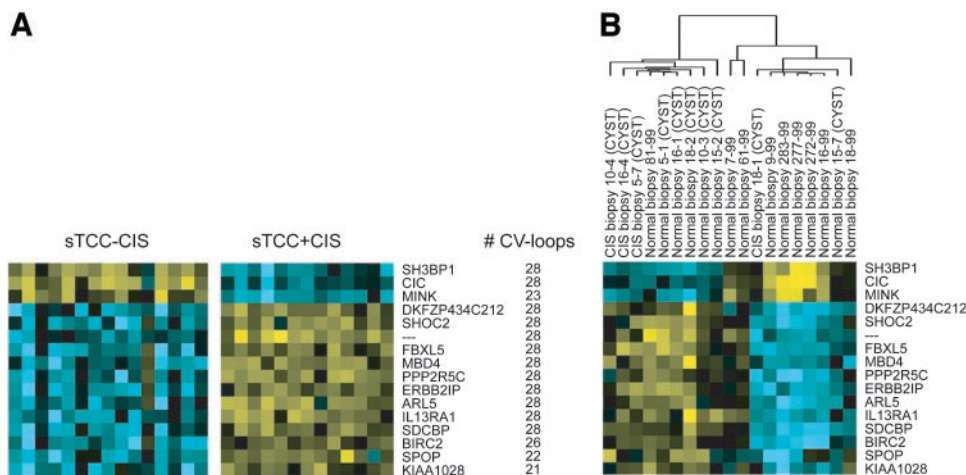


several genes encoding proteins being part of connective tissue and the immune response.

It is believed that muscle invasive TCCs arise from different pathways: (a) a pathway from a stepwise progression of transitional cell papillomas; or (b) a pathway from CIS precursor lesions. Genetic analysis (loss of heterozygosity studies and mutation analysis studies of *TP53*) has identified the CIS lesions to be the most

common precursor of invasive carcinomas, because the CIS lesions harbor many of the chromosomal abnormalities and mutations frequently observed in invasive carcinomas (4, 5). There is no direct evidence that invasive tumors arise solely from CIS lesions, although it has been demonstrated, using a transgenic mouse model expressing the SV40 oncogene, that only CIS and invasive tumors develop in the mouse bladder depending on the gene dose (29). In

Fig. 5. Expression profiles of the 16 genes in the carcinoma *in situ* (CIS) classifier. A, the expression of the 16-classifier genes in superficial transitional cell carcinoma (sTCC) with no surrounding CIS (left) and superficial transitional cell carcinoma with surrounding CIS (right). The gene symbols of the classifier genes are listed together with the number of the times used in cross-validation loops. B, the expression of the 16-classifier genes in normal samples, histologically normal samples adjacent to CIS lesions, and biopsies with CIS lesions. The top dendrogram shows the sample clustering from hierarchical cluster analysis based on the 16-classifier genes. The genes appear in the same order as in A.



our present study, we did not see a large resemblance in gene expression patterns between TCC with adjacent CIS and mTCC, which might otherwise be expected from the similar genetic alterations observed for the two tumor types. We observed that the TCC with adjacent CIS harbored their own characteristic expression profiles distinct from invasive carcinomas and TCC without adjacent CIS. In our previous expression profiling study of bladder carcinoma, we did find a close relationship between invasive carcinomas and some (four) papillary tumors with synchronous CIS lesions (13). These tumors showed up-regulation of genes involved in matrix remodeling, angiogenesis, and the immune response. However, this was not a general finding for all of the tumors with adjacent CIS studied in this article. We did identify clusters of genes with expression similarities between the TCC with CIS and the invasive tumors (cluster 1 and 4); however, a systematic functional relationship for most of the genes was not found. Among the genes with elevated expression in cystectomy biopsies (with or without CIS) and in the TCC with CIS and the invasive tumors (cluster 4) were a few genes that highlight the aggressive nature of CIS. We found the *GG2-1* gene, which may encode an antiapoptotic protein (inferred from structural domains), and the *VEGF* gene, which encodes a protein that is a prognostic marker for stage progression and recurrence in bladder cancer (30, 31) and is involved in angiogenesis (32). Furthermore, the cluster included the *TGFBR2* gene, which has been shown to be mutated in colon cancer cell lines with high rates of microsatellite instability. It is believed that the mutations make cancer cells able to escape the transforming growth factor  $\beta$ -mediated growth control (33). Interestingly, the *TIEG* gene (transforming growth factor  $\beta$ -inducible early gene) encoding a transcription factor that plays an important role in the transforming growth factor  $\beta$  signaling pathway (34) is also found in this cluster. Most other genes identified in these clusters remain interesting candidates for additional studies.

Because inflammation is a common finding in CIS mucosa, and an inflammation cluster was detected in some of the sTCC+CIS samples, it could be argued that inflammation was driving the clustering of sTCC with or without surrounding CIS. However, the inflammatory genes were only expressed in a subset of the sTCC+CIS samples, and we did not find an overrepresentation of immunology-related genes among the 100 marker genes or the 16-classifier genes. One could argue, as seen from a classification point of view, that it would not matter whether the genes that work well for classification were involved in immunology or other functions. A classifier should classify correctly, and the function of the genes working well in a classifier may be of less importance.

We found that it is possible to classify tumor samples according to the presence or absence of concomitant CIS using relatively few genes. This may lead to a better diagnosis of CIS and, as shown in this work, the expression profiles of these genes were found in the actual CIS lesions as well as in histologically normal biopsies next to the CIS lesions. The fact that CIS may be diagnosed from histologically normal areas in the bladder by gene expression profiling may increase the CIS detection rate and, as a result, lead to better optimized treatment regimens. Because TCC with surrounding CIS is relatively rare, we were not able to include more samples for validation of the CIS classifier. Consequently, the classifier for classification of CIS was validated by the "leave one out" cross-validation methodology; this procedure indicates the robustness of the classifier, but validation using independent samples is a better test of the classifier performance, because the samples in the training set may be biased because of, for example, the selection process. On the basis of this, we used an approach in which we randomly divided the samples into two groups,

one for training and one for testing. The classification performance using this approach was good despite the few samples involved in the training procedure, and we would expect a successful validation of the 16-gene classifier on independent samples.

Some of the biopsies used for expression profiling in this study were taken from cystectomy specimens, and the histopathological diagnosis of the profiled sample was determined from two adjacent biopsies. This procedure involved some uncertainty, which may explain the finding that a few of the biopsies with CIS showed the no-CIS profile. Laser-assisted microdissection of the CIS lesions and adjacent normal mucosa samples may partly eliminate this problem in the future if the RNA quality can be kept at a high level during the procedure. Such technology may be useful also for analyzing the origin of the gene expressions identified in this work and, in this way, delineate the signaling between the carcinoma cells and the underlying connective tissue.

In conclusion, we have detected a CIS signature that is reflected not only in CIS biopsies but also in TCCs and histologically normal mucosa from bladders containing CIS. We have constructed a 16-gene molecular classifier for identification of the CIS gene expression signature. This signature could be useful in the follow-up of bladder cancer patients.

## ACKNOWLEDGMENTS

We thank the technicians Hanne Steen, Bente Pytlick, Birgitte Stougaard, and Bente Devantié for excellent assistance and the staff at the Departments of Urology, Clinical Biochemistry, and Pathology at Aarhus University Hospital.

## REFERENCES

- Pisani P, Parkin DM, Bray F, Ferlay J. Estimates of the worldwide mortality from 25 cancers in 1990. *Int J Cancer* 1999;83:18–29.
- Wolf H, Kakizoe T, Smith PH, et al. Bladder tumors. Treated natural history. *Prog Clin Biol Res* 1986;221:223–55.
- Althausen AF, Prout GR Jr, Daly JJ. Non-invasive papillary carcinoma of the bladder associated with carcinoma in situ. *J Urol* 1976;116:575–80.
- Spruck CH III, Ohneseit PF, Gonzalez-Zulueta M. Two molecular pathways to transitional cell carcinoma of the bladder. *Cancer Res* 1994;54:784–8.
- Rosin MP, Cairns P, Epstein JI, Schoenberg MP, Sidransky D. Partial allelotyping of carcinoma in situ of the human bladder. *Cancer Res* 1995;55:5213–6.
- Anderstrom C, Johansson S, Nilsson S. The significance of lamina propria invasion on the prognosis of patients with bladder tumors. *J Urol* 1980;124:23–6.
- Cummings KB. Carcinoma of the bladder: predictors. *Cancer (Phila)* 1980;45:1849–55.
- Cheng L, Chevillet JC, Neumann RM, et al. Survival of patients with carcinoma in situ of the urinary bladder. *Cancer (Phila)* 1999;85:2469–74.
- Kriegmair M, Baumgartner R, Lumper W, Waidelich R, Hofstetter A. Early clinical experience with 5-aminolevulinic acid for the photodynamic therapy of superficial bladder cancer. *Br J Urol* 1996;77:667–71.
- van 't Veer LJ, Dai H, van de Vijver MJ, et al. Gene expression profiling predicts clinical outcome of breast cancer. *Nature (Lond)* 2002;415:530–6.
- van de Vijver MJ, He YD, van't Veer LJ, et al. A gene-expression signature as a predictor of survival in breast cancer. *N Engl J Med* 2002;347:1999–2009.
- Beer DG, Kardias SL, Huang CC, et al. Gene-expression profiles predict survival of patients with lung adenocarcinoma. *Nat Med* 2002;8:816–24.
- Dyrskjot L, Thykjaer T, Kruhoffer M, et al. Identifying distinct classes of bladder carcinoma using microarrays. *Nat Genet* 2003;33:90–6.
- Golub TR, Slonim DK, Tamayo P, et al. Molecular classification of cancer: class discovery and class prediction by gene expression monitoring. *Science (Wash DC)* 1999;286:531–7.
- Rosenwald A, Wright G, Chan WC, et al. The use of molecular profiling to predict survival after chemotherapy for diffuse large-B-cell lymphoma. *N Engl J Med* 2002;346:1937–47.
- Sidransky D, Frost P, Von Eschenbach A, et al. Clonal origin bladder cancer. *N Engl J Med* 1992;326:737–40.
- Irizarry RA, Hobbs B, Collin F, et al. Exploration, normalization, and summaries of high density oligonucleotide array probe level data. *Biostatistics* 2003;4:249–64.
- Eisen MB, Spellman PT, Brown PO, Botstein D. Cluster analysis and display of genome-wide expression patterns. *Proc Natl Acad Sci USA* 1998;95:14863–8.
- Tamayo P, Slonim D, Mesirov J, et al. Interpreting patterns of gene expression with self-organizing maps: methods and application to hematopoietic differentiation. *Proc Natl Acad Sci USA* 1999;96:2907–12.



20. Hosack DA, Dennis G Jr, Sherman BT, Lane HC, Lempicki RA. Identifying biological themes within lists of genes with EASE. *Genome Biol* 2003;4:R70.
21. Tusher VG, Tibshirani R, Chu G. Significance analysis of microarrays applied to the ionizing radiation response. *Proc Natl Acad Sci USA* 2001;98:5116–21.
22. Hartmann A, Moser K, Kriegmair M, et al. Frequent genetic alterations in simple urothelial hyperplasias of the bladder in patients with papillary urothelial carcinoma. *Am J Pathol* 1999;154:721–7.
23. Steidl C, Simon R, Burger H, et al. Patterns of chromosomal aberrations in urinary bladder tumours and adjacent urothelium. *J Pathol* 2002;198:115–20.
24. Liotta LA, Kohn EC. The microenvironment of the tumour-host interface. *Nature (Lond)* 2001;411:375–9.
25. Wilson CB, Leopard J, Cheresh DA, Nakamura RM. Extracellular matrix and integrin composition of the normal bladder wall. *World J Urol* 1996;14(Suppl 1):S30–7.
26. Wu XR, Manabe M, Yu J, Sun TT. Large scale purification and immunolocalization of bovine uroplakins I, II, and III. Molecular markers of urothelial differentiation. *J Biol Chem* 1990;265:19170–9.
27. Celis JE, Ostergaard M, Basse B, et al. Loss of adipocyte-type fatty acid binding protein and other protein biomarkers is associated with progression of human bladder transitional cell carcinomas. *Cancer Res* 1996;56:4782–90.
28. van Rhijn BW, Lurkin I, Radvanyi F, et al. The fibroblast growth factor receptor 3 (FGFR3) mutation is a strong indicator of superficial bladder cancer with low recurrence rate. *Cancer Res* 2001;61:1265–8.
29. Zhang ZT, Pak J, Shapiro E, Sun TT, Wu XR. Urothelium-specific expression of an oncogene in transgenic mice induced the formation of carcinoma in situ and invasive transitional cell carcinoma. *Cancer Res* 1999;59:3512–7.
30. Crew JP, O'Brien T, Bradburn M, et al. Vascular endothelial growth factor is a predictor of relapse and stage progression in superficial bladder cancer. *Cancer Res* 1997;57:5281–5.
31. Crew JP, O'Brien T, Bicknell R, et al. Urinary vascular endothelial growth factor and its correlation with bladder cancer recurrence rates. *J Urol* 1999;161:799–804.
32. Ferrara N, Gerber HP, LeCouter J. The biology of VEGF and its receptors. *Nat Med* 2003;9:669–76.
33. Markowitz S, Wang J, Myeroff L, et al. Inactivation of the type II TGF-beta receptor in colon cancer cells with microsatellite instability. *Science (Wash DC)* 1995;268:1336–8.
34. Hefferan TE, Reinholz GG, Rickard DJ, et al. Overexpression of a nuclear protein, TIEG, mimics transforming growth factor-beta action in human osteoblast cells. *J Biol Chem* 2000;275:20255–9.