MOLECULAR PROFILING OF UROTHELIAL CARCINOMA

DAVID LINDGREN

DEPARTMENT OF CLINICAL GENETICS

LUND UNIVERSITY
2006
# Table of Contents

**Original Articles**

1. **Gene Expression Profiling of Urothelial Carcinomas**
   - Materials and Methods
   - Results and Discussion
   - Multifocal and Recurrent Tumors
   - Concluding Remarks

2. **Gene Expression Patterns and Tumor Histopathology**

3. **Progression of Superficial Tumors**

4. **FGFR3 and TP53 Mutated Tumors**

5. **Superficial Tumor Recurrence**

6. **Progression of Superficial Tumors**

7. **Progression of Muscle-Invasive Disease and Overall Survival**

8. **Normal and Morphologically Altered Bladder Mucosa**

9. **Multifocal and Recurrent Tumors**

10. **Concluding Remarks**

**Reviews**

1. **Molecular Alterations in Urothelial Carcinoma**
   - Recurrent Chromosomal and Genetic Changes
   - FGFR3
   - TP53

2. **Molecular Pathways in UC Development**

3. **Tumor Clonality and Tumor Spread**

4. **Gene Expression Profiling of Urothelial Carcinomas**

5. **Gene Expression Patterns and Tumor Histopathology**

6. **Prediction of Disease Recurrence and Progression**
   - Superficial Tumor Recurrence
   - Progression of Superficial Tumors
   - Progression of Muscle-Invasive Disease and Overall Survival
   - Normal and Morphologically Altered Bladder Mucosa
   - Multifocal and Recurrent Tumors
   - Concluding Remarks

**Conclusions**

**Summary in Swedish**

**Acknowledgements**

**References**
This thesis is based on the below listed original articles, referred to in the text by their Roman numerals (I-IV):


# Abbreviations

<table>
<thead>
<tr>
<th>Abbreviation</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>ACH</td>
<td>Achondroplasia</td>
</tr>
<tr>
<td>ANN</td>
<td>Artificial neural network</td>
</tr>
<tr>
<td>APM</td>
<td>Antigen presenting machinery</td>
</tr>
<tr>
<td>aRNA</td>
<td>Antisense RNA</td>
</tr>
<tr>
<td>BAC</td>
<td>Bacterial artificial chromosome</td>
</tr>
<tr>
<td>BCG</td>
<td>Bacillus Calmette-Guérin</td>
</tr>
<tr>
<td>cDNA</td>
<td>Complementary DNA</td>
</tr>
<tr>
<td>CGH</td>
<td>Comparative genomic hybridization</td>
</tr>
<tr>
<td>COA</td>
<td>Correspondence analysis</td>
</tr>
<tr>
<td>cT</td>
<td>Tumor clinical stage</td>
</tr>
<tr>
<td>CV</td>
<td>Cross validation</td>
</tr>
<tr>
<td>ECM</td>
<td>Extracellular matrix</td>
</tr>
<tr>
<td>FDR</td>
<td>False discovery rate</td>
</tr>
<tr>
<td>FGF</td>
<td>Fibroblast growth factor</td>
</tr>
<tr>
<td>FGFR</td>
<td>Fibroblast growth factor receptor</td>
</tr>
<tr>
<td>FNR</td>
<td>False negative rate</td>
</tr>
<tr>
<td>FISH</td>
<td>Fluorescence in situ hybridization</td>
</tr>
<tr>
<td>GO</td>
<td>Gene ontology</td>
</tr>
<tr>
<td>HCA</td>
<td>Hierarchical clustering analysis</td>
</tr>
<tr>
<td>HCH</td>
<td>Hypochondroplasia</td>
</tr>
<tr>
<td>IHC</td>
<td>Immunohistochemistry</td>
</tr>
<tr>
<td>k-NN</td>
<td>$k$-nearest neighbor</td>
</tr>
<tr>
<td>LOH</td>
<td>Loss of heterozygosity</td>
</tr>
<tr>
<td>MDS</td>
<td>Multidimensional scaling</td>
</tr>
<tr>
<td>MHC</td>
<td>Major histocompatibility complex</td>
</tr>
<tr>
<td>mut</td>
<td>Amino acid changing sequence mutation</td>
</tr>
<tr>
<td>mRNA</td>
<td>Messenger RNA</td>
</tr>
<tr>
<td>PAM</td>
<td>Prediction analysis of microarray</td>
</tr>
<tr>
<td>PCA</td>
<td>Principal component analysis</td>
</tr>
<tr>
<td>PCR</td>
<td>Polymerase chain reaction</td>
</tr>
<tr>
<td>SAM</td>
<td>Significance analysis of microarrays</td>
</tr>
<tr>
<td>SVM</td>
<td>Support vector machines</td>
</tr>
<tr>
<td>SNP</td>
<td>Single nucleotide polymorphism</td>
</tr>
<tr>
<td>pT</td>
<td>Tumor histopathological stage</td>
</tr>
<tr>
<td>PCA</td>
<td>Principle component analysis</td>
</tr>
<tr>
<td>QT clust</td>
<td>Quality cluster algorithm</td>
</tr>
<tr>
<td>UC</td>
<td>Urothelial carcinoma</td>
</tr>
<tr>
<td>SADDAN</td>
<td>Severe hypochondroplasia with developmental delay and acanthosis nigricans</td>
</tr>
<tr>
<td>TD</td>
<td>Thanatophoric dysplasia</td>
</tr>
<tr>
<td>wt</td>
<td>Wild type</td>
</tr>
</tbody>
</table>
Urothelial carcinoma (UC) is a heterogeneous disease ranging from low grade papillary tumors with a clinically benign disease course to poorly differentiated invasive lesions associated with an unfavorable prognosis. Although a number of recurrent genetic alterations have been found in bladder cancer, the molecular events underlying bladder cancer development and progression are still largely unknown. It is therefore of importance to further characterize UC to increase the biological understanding of the disease. Moreover, biomarkers with improved prognostic potential are needed since current clinical parameters are insufficient in predicting outcome for the individual patient.

The general aim of this thesis was to molecularly characterize UC at the transcriptional level using gene expression microarrays. This thesis is divided into four major sections. The first section includes a brief introduction to bladder cancer and an introduction to the microarray technology, which is the key methodology used. In the second part, the specific aims of the thesis, a summary of materials and methods, and the results are given with a short discussion. In the third section, two reviews on molecular changes in UC are given, where the results from the present study are included and discussed in relation to previous investigations. The fourth and final section contains the original articles (I-IV) on which this thesis is based.

Lund, August 2006
INTRODUCTION

UROTHELIAL CARCINOMA

Bladder cancer is the fifth most common malignancy in men after prostate, skin, lung, and colon cancer and show a male to female ratio of approximately 3:1. There are around 2000 new cases in Sweden each year and the disease only rarely affects individuals younger than 30 years; roughly 85% of all newly diagnosed bladder cancer patients in Sweden 2004 were 60 years or older. Bladder tumors are almost exclusively of epithelial origin, and in the industrialized countries more than 90% are of the urothelial carcinoma type (UC).

Tumor Pathology and Clinical Manifestation

Urothelial carcinomas arise in the urothelium which is the epithelial cell layer lining the urinary bladder. Pathologically, UC is grouped into different tumor stages according to the depth of tumor invasion (Figure 1) and into different grades based on the morphological differentiation. Ta tumors are exophytically growing papillary tumors which are strictly confined to the epithelial mucosa. Carcinoma in situ (Tis), also superficially confined, is a flat lesion composed of poorly differentiated cells (grade 3; G3). Tumors which have spread through the basal lamina and into the underlying connective tissue layer are termed stage 1 (T1). A subsequent invasion into, and beyond, the muscle layers is collectively termed muscle-invasion. Of these, stage 2 tumors (T2) are organ confined, i.e., have not extended further than the muscle layers, whereas the non-organ confined stages have grown through and extended into
the perivesical tissue (stage 3; T3) or into adjacent organs such as the prostate or the abdominal wall (stage 4; T4). Most T2 to T4 tumors are non-papillary growing, of high grade, and present without any history of papillary low grade tumors. Furthermore, these tumors are often associated with concomitant fields of dysplastic urothelium or Tis.

Figure 1: Histological staging of UC.

Approximately 75% of patients with UC present with a non muscle-invasive disease (stages Ta, T1, or Tis), and of these, the majority is Ta tumors of low to medium grade. A characteristic of UC is the high frequency of tumor recurrence (metachronous tumors); more than half of the patients with Ta tumors show tumor recurrences. Despite the high recurrence rate, the disease course for many of these patients is highly favorable, and only a minority develop a subsequent muscle-invasive disease. Non muscle-invasive tumors are mainly managed by transurethral resection or fulguration of visible tumor growth. However, in the presence of adverse prognostic signs, such as high grade disease, large or multiple (synchronous) tumors, or concomitant Tis, intravesical therapy, such as Bacillus Calmette-Guérin (BCG), are often administered and have been shown to significantly improve patient prognosis. Stage Ta, T1, and Tis tumors are often collectively termed as superficial tumors. However, in contrast to low and moderately differentiated Ta tumors, high grade Ta, T1, and Tis lesions are associated with a much adverse
Introduction

Hence, due to the heterogeneity in clinical outcome within this group, the term superficial should be used with care.12

Muscle invasive tumors are highly malignant and if left without aggressive therapeutic management, the majority of patients die of the disease within a few years from initial diagnosis.15 The current standard treatment for muscle invasive tumors is radical cystectomy, i.e., removal of the urinary bladder, or less commonly, radiation therapy.16 The survival after cystectomy is mainly influenced by lymph node status and histopathological tumor stage. The 10-year recurrence free survival for organ confined lymph node negative disease is 80% after cystectomy, in contrast to 30% for lymph node positive T3 and T4 tumors.17

In UC, individual patient disease course is unpredictable and conventional prognostic factors such as tumor grade/stage, size, and multiplicity, do not sufficiently predict clinical outcome.8,9 Even the clinically benign group of Ta tumors requires close follow-up with check-up cystoscopies; examinations that affect patient’s quality of life and are costly. Therefore, an improved understanding of the molecular changes that characterize UC is important. In particular, it is of importance to identify molecular markers that define subgroups of patients who may benefit from either less or more intensive treatment modalities, as well as molecular networks that can serve as targets for the development of novel drugs or treatment strategies.

Genetic Changes in Urothelial Carcinomas

Numerous molecular studies of UC have been performed over the years, identifying several recurrent genetic changes.18,19 For example, molecular and conventional cytogenetic studies have revealed loss of chromosome 9 as the most characteristic karyotypic change in UC. Moreover, oncogenic activation of FGFR3 or RAS gene family members, or inactivation of known tumor
suppressor genes such as $TP53$ and $CDKN2A$ are frequently seen. A more
detailed summary of these characteristic alterations together with the findings
obtained in the present study will be presented in the review-section of this
thesis.

**GENE EXPRESSION PROFILING**

**Methodological Principles**

Large scale gene expression analyses have during recent years become an
appreciated method for exploring molecular characteristics of a wide variety of
biological states and conditions. Gene expression profiling, a term collectively
used for techniques which simultaneously measures transcript levels of a large
number of genes, was first described in 1991 by Lennon and Lehrach\(^{20}\) who let
radiation-labeled targets hybridize to filter-arrays spotted with cDNA clones.
These array-based techniques provided a huge advantage compared to
previously used approaches, such as Northern blot\(^{21}\) and PCR-based methods,
which only could monitor the expression of single genes. In 1995, Schena *et al.*\(^{22}\) successfully showed that the technique could be significantly improved by
printing cDNA clones on glass slides and by hybridizing fluorescently labeled
cDNA targets. Shortly thereafter, the first gene expression study using this
technique on a human malignancy was published, studying the expression levels
of 1000 genes in a melanoma cell line.\(^{23}\) Currently, a variety of different settings
are available for array-based expression profiling providing a probe density per
array of hundreds of thousands of elements. Hence, gene expression profiling
may be performed on a large scale and, at least theoretically, cover the complete
transcriptome.

Gene expression analysis using microarray technology relies on the
nature of single stranded DNA to base pair with its complementary sequence.\(^{24}\)
Introduction

In practice, a large number of probes, each representing a part of a gene, are immobilized on chemically modified glass slides using robotic arrayers. The probes used are typically cDNA clones of 500-3000 base pairs or *in vitro*-synthesized oligomers of 70-80 bases. Alternatively, the commercially available Affymetrix GeneChips rely on fabrication of small oligonucleotide sequences directly on a chip, providing a much higher density than for printed glass arrays.

Microarray experiments using cDNA or oligonucleotide probes printed on glass slides basically use the same methodological principle. In short, differentially labeled targets, derived from RNA samples from *e.g.*, tumor and a common reference sample, are hybridized to a glass slide containing the probes. Most commonly, the labeling reaction is performed by reverse transcription of the RNA samples with the simultaneous incorporation of fluorescent dyes. Typically, the tumor sample is labeled with Cy3 and the reference sample is labeled with Cy5. The differentially labeled cDNA targets, are purified, pooled, and subsequently incubated on the probe containing microarray slide. Hybridized slides are thereafter washed in solutions of increasing stringency to promote disassociation of mismatched target-probe heteroduplexes. Subsequently, target hybridization signals are quantified in a high-resolution laser scanner by excitation of the fluorescent dyes and detection of the emitted light. The scanner images are thereafter analyzed using image analysis software. The competitive hybridization between the tumor and reference targets will reflect the abundance of the specific targets and thus, for each probe present on the array, a relative expression ratio between the tumor and reference sample is obtained.

A common problem with microarray analysis is that large amounts of RNA are needed for efficient hybridization. Many biological samples are small and therefore it may be difficult to obtain enough RNA for analysis. To overcome this problem, RNA amplification has become a widely accepted
method to obtain sufficient material for hybridization. Most commonly, this is performed by utilizing the T7 RNA polymerase promoter for \textit{in vitro} transcription of antisense RNA (aRNA).\textsuperscript{30,31}

**Analysis of Microarray Data**

After collection of primary data, the expression values must be normalized to adjust for methodological biases, \textit{e.g.}, for variation in the quantity of initial RNA input and for differences in the detection efficiencies of the fluorescent dyes.\textsuperscript{32,33} In its simplest way, this may be performed by using total intensity normalization, which rescales the two channels so that the total quantity of hybridization for each channel is the same.\textsuperscript{32} However, since the measured intensities often are non-linear, regression methods such as the LOWESS are more appropriate.\textsuperscript{32-34} Following normalization, the expression ratio, typically the $\log_2$ of the tumor intensity divided by the reference sample, is calculated. Hence, a gene twice as abundant in the tumor sample as compared to the reference sample will have an expression value of 1, whereas a gene half as abundant will have an expression value of -1.

Each microarray hybridization generates gene expression values for thousands of genes, and a multitude of statistical analysis methods for interpretation and comparisons of data have been described. Basically, two different main approaches exist, \textit{unsupervised} and \textit{supervised} analyses.\textsuperscript{32} Unsupervised methods involve exploration of the gene expression profiles without the use of any \textit{a priori} knowledge and are thus suitable for \textit{i.e.}, the identification of subgroups based on their gene expression, or covarying genes of possible biological relevance. The perhaps most commonly used method is hierarchical clustering analysis (HCA) in which a hierarchical dendogram is constructed, reflecting the similarities in expression profiles between samples and/or genes.\textsuperscript{35} The HCA is often presented as a so called heatmap; a chart in
which genes, color coded according to their relative expression, are ordered horizontally and samples are ordered vertically. Apart from HCA, grouping of genes expressed in a similar fashion across experiments may be performed using other algorithms, such as $k$-means clustering\textsuperscript{36} or the quality cluster algorithm (QT clust).\textsuperscript{37} Multidimensional scaling (MDS),\textsuperscript{38} principal component analysis (PCA),\textsuperscript{39} and correspondence analysis (COA)\textsuperscript{40} are other ways to explore and visualize global gene expression similarities between samples. In these methods, the expression data are reduced into two- or three-dimensional subspaces with as little loss of information as possible.

Supervised methods aim to identify gene signatures specific for predefined groups, e.g., distinct histological subtypes or groups with different clinical outcome. Such class comparisons are often performed on a gene-by-gene basis using standard statistical methods such as student’s t-test and Mann Whitney U-test, or related methods such as significance analysis of microarrays (SAM),\textsuperscript{41} specifically developed for the application of microarray data. However, since the number of genes tested greatly exceed the number of samples, it is vital to take the multiple statistical testing problem into consideration; a good compromise between the false discovery rate (FDR) and the false negative rate (FNR) is needed, i.e., it is important to keep the number of falsely identified genes to a minimum while not discarding too many true positive genes.\textsuperscript{42} More refined statistical strategies are often applied for the identification of gene signatures that may be used for classification of unknown samples into predefined groups. Common discriminatory techniques are support vector machines (SVM),\textsuperscript{43} artificial neural networks (ANN),\textsuperscript{44} or prediction analysis of microarray (PAM).\textsuperscript{45}

An important aspect of microarray analysis is the biologic interpretation of the data. Often, relatively large gene lists of coexpressed genes or gene signatures specific for tumor subtypes are identified. Because of confounding factors, such as the multiple testing problem or methodological biases, each
gene list is likely to include false positive genes, and therefore biologic interpretations relying on single genes have to be made with caution. Furthermore, for a large gene list it is easy to focus only on well-known genes and thereby miss important but less well characterized genes. One way of exploring the biologic properties of an isolated gene cluster is to use annotation systems in which genes have been systematically ordered into groups based on their specific characteristics, e.g., cellular function or location. Such annotations are provided by, for example, the Gene Ontology (GO) consortium (http://www.geneontology.org/) or the KEGG database (http://www.genome.jp/kegg/). One advantage of grouping genes into classes with biological relevance is the possibility of calculating a statistical measure of enrichment within a given list of genes, thereby providing a highly useful way of biologic interpretation.
THE PRESENT STUDY

In this section, the major aims of this thesis are presented followed by a summary of the patient material and the methods used. Thereafter, a short summary and discussion of the obtained results are given.

SPECIFIC AIMS

The general aim of this thesis was to explore transcriptional events associated with bladder cancer development and progression using gene expression microarrays. More specifically, the aims were:

- to establish gene expression profiles of a large series of bladder tumors, representing all histological stages and to investigate possible transcriptional events associated with tumor histopathology and morphology (ARTICLES I and II),

- to correlate expression data with molecular alterations that previously have been shown to be frequent in UC, such as \( \text{FGFR3} \) and \( \text{TP53} \) mutation status, and allelic loss on chromosome 9. With this data, the aim was to establish a more refined view of the molecular alterations characteristic for UC (ARTICLES I and II),

- to identify gene signatures associated with tumor recurrence in superficial tumors (ARTICLE I) and the presence of lymph node metastasis in patients submitted to radical cystectomy (ARTICLE II),
- to gain an insight into the development of syn- and metachronous UC using a combined analysis of gene expression patterns, genomic copy number profiles, LOH analysis for chromosome 9, and \textit{FGFR3/TP53} mutation status (ARTICLE III),

- to perform a high-resolution screening for target oncogenes in UC using genomic copy-number profiles overlaid with expression data (ARTICLE IV).
MATERIALS AND METHODS

Patients and Tissues
The present study comprises a total of 190 urothelial tumor samples collected from 163 patients undergoing transurethral resection at the University Hospital of Lund, Sweden. The studies were approved by the Research Ethics Committee of Lund University and informed consent was obtained in all cases. Samples were collected by cold-cup biopsies from the exophytic part of the bladder tumor. Histopathological staging and grading were reviewed according to the 2002 TNM\textsuperscript{46} and 1999 WHO\textsuperscript{47} classification systems by one single pathologist.

Gene Expression Profiling
Two different microarray platforms were used in the four articles included in this thesis. In ARTICLE I, a cDNA array spotted with 25,648 cDNA clones corresponding to 11,592 unique genes was used. In ARTICLES II and IV the expression analyses were performed on oligonucleotide arrays printed with 36,288 70-mers from the OPERON v 3.0 set corresponding to 18,466 unique genes. ARTICLE III included samples hybridized to both platforms. All arrays were obtained from the Swegene DNA microarray resource centre at Lund University.

Unsupervised Analyses of Gene Expression Patterns
Tumor-tumor associations based on gene expression profiles were visualized by HCA (ARTICLE I-III), MDS (ARTICLE I), and COA (ARTICLE II). HCA is perhaps the most commonly used method to visualize microarray data. It is, however, important to note that different distance measures and clustering methods, as well as filtering and normalization methods will influence the
clusters formed. Furthermore, clustering will be performed irrespectively if natural clusters exist or not, and the validity of the formed clusters is not easily determined. It is therefore mandatory to evaluate the basis of the formed clusters, such as possible biologic differences. In ARTICLE I-III, HCA was performed using 1-Pearson correlation and Wards’ algorithm or complete linkage for cluster formation. COA (ARTICLE II) is analogous to other projection methods such as PCA and MDS and enables projection of the gene expression matrix into a low dimensional subspace that accounts for the main variance in the data. It is based on Chi-square distances of discrete values and the representation obtained is similar to representations obtained by PCA. The advantage of COA is, however, that it enables projection of both genes and tumors within the same space. Hence, genes and tumors that co-localize to the same space show positive associations.

The QT clust algorithm was used to identify clusters of co-expressed genes (ARTICLE I and II). QT clust works by forming a candidate cluster of the first gene and grouping genes with the highest correlation iteratively in a way that minimizes the cluster diameter (d; defined as 1-Pearson correlation), until no further genes may be added without exceeding a predetermined d-value. This procedure is performed with all genes in the data set as a seed. The largest cluster is then retrieved and the procedure repeated excluding the genes selected for the preceding cluster. This makes sure that the largest and most coherent clusters of genes are formed. For this algorithm, we used a d-value of 0.35 and cluster size restricted to at least 20/25 reporters; the d-values were adjusted empirically to result in a reasonable amount of formed clusters showing high correlation.
Supervised Analyses

In ARTICLES I, II, and IV, discriminatory comparisons between different group assignments were performed using the SAM algorithm. To evaluate the predictive strength of the gene signature associated with lymph node status (ARTICLE IV), the score for each tumor was bias corrected using leave-one-out cross validation (CV). Optimally, a training set is used for defining the class-predicting expression signature, whose predictive strength is subsequently determined in an independent sample test set. However, in microarray experiments the number of samples in each group is often too small for obtaining training and test sets. In these situations, CV approaches may be more suitable. In these methods, one sample is left out when constructing the gene list to be used for scoring. The predictor score for the withdrawn sample is thereafter calculated and the sample is classified. This procedure is repeated until all samples have been classified. In the present investigation we used gene lists of fixed length (50 genes) in all CV-loops, since this number of genes corresponded to a good compromise between FDR and FNR.

Biological Interpretation

To investigate possible biological properties of the gene clusters identified in ARTICLES I and II, the EASE software was applied. EASE enables identification of significantly enriched GO categories in a given list of genes. Step-down Bonferroni multiplicity-corrected p-values <0.05, calculated using EASE statistics, were considered significant.

Array-based CGH

Array based comparative genomic hybridization (CGH) uses the same microarray-based principles as gene expression profiling. Typically, large genomic bacterial artificial chromosome (BAC) clones are printed on glass slides, and genomic DNA from a test sample and a reference sample with
normal karyotype are differentially labeled with fluorescent dyes and subsequently hybridized to the slide. The large number of clones possible to print on one array slide provides array CGH with a much better resolution than conventional CGH. In ARTICLE IV, whole-genome tiling resolution 32k BAC-arrays (Swegene DNA microarray resource centre at Lund University) were used to investigate genomic alterations in 38 UCs.

**Mutation Screening**

Mutation screening of *FGFR3* and *TP53* was carried out with direct sequencing using standard procedures. All *FGFR3* regions in which mutations have been reported for UC (exons 7, 10, and 15)\(^{18}\) were sequenced. For *TP53*, exons 4-9 were sequenced; a region covering the DNA-binding domain of *TP53* and in which the vast majority of mutations have been identified.\(^{51,52}\) For samples included in the cDNA microarray experiments (ARTICLE I, and III), mutation analyses were carried out using cDNA from amplified RNA as template: identified mutations were sequence verified in samples from which genomic DNA was available. Samples hybridized to the oligonucleotide array (ARTICLES II-IV) were sequenced with genomic DNA as template and mutations were verified via reverse sequencing.

**LOH Analysis**

LOH analysis was performed on tumors from which both tumor DNA and matching blood samples were available. A total of 17 highly polymorphic microsatellite markers, evenly distributed over both arms of chromosome 9, were selected (ARTICLES I-III). In ARTICLE I, two markers close to *TP53* at chromosome 17 were also analyzed. Allelic imbalance was defined as a more than 50% reduction in allele signal intensity and LOH was defined as the presence of at least two consecutive markers showing allelic imbalance.
RESULTS AND DISCUSSION

ARTICLE I and ARTICLE II

For ARTICLES I and II, the experimental settings and aims were largely the same. In both studies, gene expression profiling using microarray was used to explore the transcriptional diversity in UC. In addition, we investigated possible associations between the expression profiles and tumor stage/grade, or mutation status of \( \text{FGFR3}/\text{TP53} \) and LOH for chromosome 9. In ARTICLE I, non muscle-invasive tumors (Ta and T1) were analyzed whereas ARTICLE II focused on T1-T4 tumors. Below, a summary of the combined findings in these two studies is presented.

The mutation screening of \( \text{FGFR3} \) and \( \text{TP53} \) was performed on 150 out of 163 tumors (Table 1). As expected, \( \text{FGFR3} \) mutations were found to be inversely associated and \( \text{TP53} \) mutations to be positively associated with tumor stage/grade.\(^{53-61} \) Also, in line with previous investigations,\(^{61,62} \) \( \text{FGFR3} \) mutations were highly correlated with \( \text{FGFR3} \) overexpression; in fact \( \text{FGFR3} \) expression was the most discriminatory transcriptional alteration when comparing \( \text{FGFR3}^{\text{mut}} \) and \( \text{FGFR3}^{\text{wt}} \) cases.

<table>
<thead>
<tr>
<th>Mutation</th>
<th>Ta</th>
<th>T1</th>
<th>T2-T4</th>
<th>G1</th>
<th>G2</th>
<th>G3</th>
</tr>
</thead>
<tbody>
<tr>
<td>( \text{FGFR3}^{\text{mut}} )</td>
<td>68%</td>
<td>34%</td>
<td>19%</td>
<td>80%</td>
<td>59%</td>
<td>20%</td>
</tr>
<tr>
<td>( \text{TP53}^{\text{mut}} )</td>
<td>8%</td>
<td>37%</td>
<td>40%</td>
<td>9%</td>
<td>12%</td>
<td>43%</td>
</tr>
</tbody>
</table>

Table 1. Mutation frequencies of \( \text{FGFR3} \) and \( \text{TP53} \) in UCs from 150 individual patients. Cases are stratified into tumor stage and grade, respectively.

When investigating tumor mutation status in relation to expression profiles it was observed that the mutational patterns of \( \text{FGFR3} \) and \( \text{TP53} \) were strongly associated with global gene expression. The MDS of Ta/T1 tumors (ARTICLE I) showed a highly polarized pattern where \( \text{FGFR3}^{\text{mut}} \) tumors clustered together and were separated from \( \text{TP53}^{\text{mut}} \) cases (Figure 2a). This
pattern was also evident in the COA representation of T1-T4 tumors (ARTICLE II); tumors with $FGFR3^{\text{mut}}$ clustered separately from tumors with TP53 impairment (either by $TP53^{\text{mut}}$ or $MDM2$ overexpression) in the COA representation, seemingly independent of histological and morphological status (Figure 2b). Moreover, both the MDS and the COA representations indicated that tumors with concomitant $FGFR3$ and $TP53$ mutations ($FGFR3^{\text{mut}}/TP53^{\text{mut}}$) were more similar to $FGFR3^{\text{mut}}/TP53^{\text{wt}}$ than to $FGFR3^{\text{wt}}/TP53^{\text{mut}}$ tumors. The difference in gene expression profiles was largely dependent on a gene signature specific for cell-cycle related transcripts that showed increased expression in tumors with TP53 impairment. The fact that $FGFR3^{\text{mut}}/TP53^{\text{mut}}$ tumors showed lower expression of this signature as compared to $FGFR3^{\text{wt}}/TP53^{\text{mut}}$ tumors further strengthens the hypothesis that $FGFR3^{\text{mut}}/TP53^{\text{mut}}$ cases are genetically progressed $FGFR3^{\text{mut}}$ tumors rather than $TP53^{\text{mut}}$ tumors with subsequent accumulation of $FGFR3^{\text{mut}}$.

Figure 2. Expression patterns of UC are highly associated with $FGFR3$ and $TP53$ mutation status. a) Unsupervised MDS representation of Ta and T1 tumors based on the 6749 genes, and b) unsupervised COA representation of TaG3 and T1-T4 tumors based on the 3334 genes. Cases are color-coded according to $FGFR3$ mutation status and TP53 impairment.
In line with previously published expression studies of UC, we found a general association between tumor stage/grade and the global gene expression patterns. This was illustrated by HCA and MDS in ARTICLE I, and by HCA and COA in ARTICLE II. Thus, the cosegregation of tumors of similar grades and stages in unsupervised analyses indicate that large gene signatures correlate with tumor histopathology. Specifically, overexpression of genes related to extracellular matrix (ECM) and immunological transcripts correlated with tumor invasiveness. Conversely, low grade Ta tumors revealed low expression of cell-cycle genes and high expression of ribosomal transcripts.

Apart from a general downregulation of a large number of genes on chromosome 9 (ARTICLE I), we did not find any specific transcriptional event associated with LOH for chromosome 9. However, we noted that a subset of low grade Ta tumors, which all showed FGFR3 mutation and retention of chromosome 9, were highly similar in gene expression (ARTICLE I). This group of tumors also showed the lowest expression of cell-cycle genes and hence, this may indicate that the loss of chromosome 9 is associated with tumor progression rather than neoplastic initiation.

Finally, discriminatory analyses were applied to explore transcriptional events associated with clinical variables. In ARTICLE I, a gene signature associated with short recurrence-free follow-up for superficial tumors was described. The gene list was highly enriched for genes associated with cell-adhesion. In ARTICLE II, discriminatory analyses identified a gene signature whose high expression predicted absence of lymph node metastasis in patients submitted to radical cystectomy. In fact, this gene signature was superior to clinical stage in predicting lymph node metastasis. Furthermore, the discriminatory genes showed an impressive predictive power for overall survival when combined with histopathological tumor stage. It is noteworthy that a large proportion of these genes were associated with the antigen presenting machinery (APM). We therefore hypothesize that the low expression
of APM-related transcripts in tumors with metastasis may reflect an immune escape phenotype,\textsuperscript{72,73} which is not able to trigger immune responses vital for the suppression of tumor spread.

**ARTICLE III**

To explore the genetic relation between syn- or metachronous tumor development, 49 tumors from 22 patients were studied with gene expression profiling. The majority of tumors were also investigated using CGH, LOH analysis of chromosome 9, and mutation analysis of the\textit{FGFR3} and\textit{TP53} genes. Although CGH, LOH, and mutation analyses indicated a clonal relationship between most syn- and metachronous tumors from individual patients, the combined analyses pointed to a complex genetic progression model; no simple karyotypic relationship was detected between initial and recurring or synchronous bladder tumors. Thus, the present findings corroborate previous results that the chronology of tumor presentation is not reflected in the genetic progression of the tumors.\textsuperscript{74} In contrast, expression profiling revealed a strong similarity between syn- and metachronous tumors; samples from the same patient almost exclusively clustered close to each other in unsupervised HCA when compared with a large reference set of tumors of similar stages and grades. A more detailed investigation revealed that the strong expressional correlation between syn- and metachronous tumors was close to the correlation seen between normal urothelial samples from different individuals, implying that the small differences observed are largely caused by experimental variation. The only exception was tumors that progressed from Ta/T1 to T2, possibly reflecting the shift in tumor microenvironment connected with muscle invasion.

The observation that tumors with more complex genomes frequently appear earlier than clonally related, but less evolved, tumors, suggest that genetically altered cells, with clonally related but differently evolved genomes
The Present Study

may co-exist in the urothelium, and that cells in these fields independently may produce overt tumors.\textsuperscript{75} Indeed, studies have shown that hyperplastic and dysplastic intraurothelial lesions in patients with UC commonly share genetic alterations with adjacent tumors.\textsuperscript{76-81} In addition, it has been observed that genomic imbalances and gene mutations are also present in histologically normal-looking mucosa surrounding the tumor.\textsuperscript{76-85} Given the remarkably stable expression profiles of syn- and metachronous tumors within individual patients, the formation of a fixed expression profile may be considered as a possible primary event in the development of UC.

**ARTICLE IV**

Genomic copy-number profiles were obtained for 38 UCs using array-based CGH. With the aim to identify oncogenes of importance for UC progression, commonly amplified regions specific for high grade and invasive tumors were delineated and compared with gene expression data collected using microarray analysis and real time PCR. The most frequent amplification was found in 35\% of high grade cases and localized to 6p22. The high resolution provided by the 32k BAC-arrays, revealed a minimal overlap of 1 Mb that was present in 90-100\% of cases; a region containing \emph{OACT1}, \emph{E2F3}, \emph{CDKAL1}, and \emph{SOX4} as the only mapped genes. Using cases with no copy-number gain of this region as reference, it was further shown that primarily \emph{CDKAL1} and \emph{E2F3}, and to a somewhat lesser extent \emph{SOX4}, showed a strong association between amplification and gene overexpression. We conclude that \emph{CDKAL1}, \emph{E2F3}, and \emph{SOX4} may be considered as possible oncogenes in UC with possible importance for an aggressive phenotype.

Two additional amplicons were specific for advanced tumors, at 8q22 (4.5Mb) and at 2p55 (1Mb), respectively. Intriguingly, these two regions occurred in a mutually exclusive pattern and contained several related genes and were therefore seemingly paralogous. Further analyses revealed that
amplification of \textit{YWHAZ} (14-3-3-zeta) at 8q22 and \textit{YWHAQ} (14-3-3-theta) at 2p55 both correlated with gene expression. Thus, the results highlight gene members of the 14-3-3-family to be of possible for tumor progression in UC.

Homozygous deletions were detected at nine separate chromosomal locations. Deletion of the \textit{CDKN2A}/\textit{CDKN2B} region at 9p21 was most frequent (32%), covering all stages and grades. Other known tumor suppressor genes found homozygously deleted were \textit{PTEN} at 10q23 and \textit{RB1} at 13q14, deleted in one case each. Noteworthy, four of the six remaining homozygous deletions were found on chromosome 9, indicating that chromosome 9 is particularly susceptible for homozygous deletions. A homozygous deletion at 9q21 was observed in two cases. We therefore screened a cohort of 48 UCs for sequence mutations in a gene within this region, \textit{OSTF1}, but did not find any missense mutations.
Numerous studies of molecular changes in UC have been performed over the years. In line with other neoplasms, investigations have revealed a variety of recurrent genetic changes. Below, results obtained in the present investigation will be discussed in relation to previously published studies. The first section encompasses frequent alterations associated with UC, as well as recent advances in the understanding of tumor evolution and clonality. The second section gives a comprehensive review of gene expression profiling studies performed on bladder cancers.

MOLECULAR ALTERATIONS IN UROTHELIAL CARCINOMA

Recurrent Chromosomal and Genetic Changes

A recent review of 188 published karyotypes of urothelial tumors concluded that, cytogenetically, chromosome 9 was the most frequently involved chromosome, being numerically or structurally aberrant in 45% of the investigated cases. Most often, monosomy 9 or loss of whole chromosome 9 arms is seen. Molecular cytogenetic studies using CGH corroborate the frequent loss of chromosome 9 material in UC, and conclude that TaG1 and TaG2 tumors show few additional genomic changes. In contrast, cytogenetic and CGH studies have shown that T1-T4 and high grade tumors display a multitude of chromosome aberrations. Apart from chromosome 9, recurrent structural changes in advanced tumors involve gain of 1q, 3q, 5p, 8q, 17q, and 20q, and loss of 5q, 6q, 8p, 11p, and 11q. It has also been shown that the number of chromosomal aberrations significantly correlate with future stage
progression in T1 tumors.\textsuperscript{88} Thus, a complex karyotype may be considered an adverse sign in UC.

The frequent loss of chromosome 9 material is also reflected by allelic loss, as shown by LOH studies.\textsuperscript{92-100} Much effort has been directed to the identification of tumor suppressor genes on chromosome 9. However, it has been difficult to isolate small regions of common deletion. To this date, four candidate regions have been proposed. In the present study, array CGH analysis of 38 UCs revealed homozygous deletion of the \textit{CDKN2A} loci at 9p21, coding for the well characterized cell-cycle inhibitors p16 and p14\textsuperscript{ARF}, in 32% of cases.\textsuperscript{94} Similar high frequencies of homozygous deletion has been reported previously.\textsuperscript{101,102} Thus, \textit{CDKN2A} seem to be an important tumor suppressor loci that contribute to UC development. The other three regions involve 9q22.3, \textsuperscript{92,95,99,100,103} 9q32, \textsuperscript{95,96,104} and 9q34.\textsuperscript{92,98} However, the importance of the candidate tumor suppressor genes within these regions (\textit{PTC}, \textit{DBCCR1}, and \textit{TSC1}, respectively) is at present not clear. Homozygous deletion of these genes are infrequent,\textsuperscript{97,105,94} and mutation rates are low; \textit{TSC1} and \textit{PTC} both show missense sequence mutation in less than 5% of the cases,\textsuperscript{106-108} and no mutation in the \textit{DBCCR1} coding region has been found.\textsuperscript{97,104} Nevertheless, studies using cell lines have shown that \textit{DBCCR1} is commonly silenced by hypermethylation\textsuperscript{104} and that exogenous expression of this protein suppresses proliferation.\textsuperscript{109} Furthermore, \textit{ptc}\textsuperscript{+/−} transgenic mice showed earlier onset of neoplastic bladder transformation than wild type mice after exposure to chemical carcinogens, suggesting that \textit{PTC} haploinsufficiency may be of importance for UC development.\textsuperscript{110}

Apart from \textit{CDKN2A} on 9p and the putative tumor suppressor genes on 9q, several important regulatory genes are recurrently found inactivated in UC. For example, mutations or homozygous deletion in the \textit{PTEN} gene has been identified in 23% of late-stage cases,\textsuperscript{111} and for the \textit{RB1} gene, LOH and aberrant
protein expression is frequent in muscle-invasive tumors. Aberrant RB1 expression has also been shown to correlate with poor prognosis.

A number of recurrently amplified chromosomal regions have been observed in UC and UC-derived cell lines. CGH studies followed by fine-mapping by semiquantitative PCR or fluorescence in situ hybridization (FISH) have delineated two common amplicons at 8q22 and 6p22, respectively. Subsequent array CGH investigations have verified and further delineated the boundaries of these amplicons. Within the 6p22 region the CDKAL1, E2F3, and SOX4 have been described as the target oncogenes. In the present study, we observed that the 8q22 amplicon was mutually exclusive to another amplicon at 2p55. Interestingly, these regions contained two related genes, YWHAZ (14-3-3-zeta) and YWHAQ (14-3-3-theta), respectively, which both showed high correlation between overexpression and amplification. We therefore suggest that gene members of the 14-3-3-family may be oncogenes important for high grade disease. Amplification or overexpression of other genes have been reported, e.g., ERBB2, TOP2A, CCND1, MDM2, and CMYC. Moreover, using direct sequencing, point mutations causing oncogenic activation of RAS gene family members (HRAS, KRAS2, and NRAS) has been found in 13% of UCs. However, the two most commonly mutated genes in UC are the oncogenic activation of the FGFR3 and inactivation of the TP53 genes, respectively. Below, a detailed description of these genes is given.

FGFR3

In 1999, Cappellen and coworkers found that a significant proportion of UCs display specific missense mutations in the fibroblast growth factor 3 gene (FGFR3). Numerous studies subsequently confirmed their findings and FGFR3 mutations are now recognized as one of the most characteristic genetic alterations seen in UC.
FGFR3 belongs to a family of at least four transmembrane receptor tyrosine kinases, FGFR1-FGFR4. The fibroblast growth factors (FGFRs) are activated by the fibroblast growth factors (FGFs), which in concert with heparin or heparan sulfate proteoglycan induce receptor dimerization, and autophosphorylation of tyrosine residues in the cytoplasmic domain of the receptor molecule. At least 23 different FGFs have been identified. Furthermore, different splice variants for each receptor have been found, each with different affinity for the various FGF ligands. The FGFR3 gene contains 19 exons of which exons 2-18 encode the FGFR3 protein product. Two major tissue specific splice variants have been reported; the IIIb and the IIIC variant, of which IIIb is preferentially expressed by epithelial cells. The specific details of FGFR3 signaling are however not yet fully resolved. Studies have shown that SH2 mediated activation of the RAS-ERK and the PI3K-AKT signaling pathways may be important. Also, it has been suggested that cell-specific differences in FGFR3 signaling may depend upon differences in STAT activation.

Intriguingly, mutations in FGFR3 have previously been found to be associated with several germline human skeletal disorders such as thanatophoric dysplasia type I and II (TDI and TDII), achondroplasia, hypochondroplasia, and severe hypochondroplasia with developmental delay and acanthosis nigricans (SADDAN). The severity of these syndromes varies according to the specific FGFR3 mutation and in vitro studies have shown that the identified mutations lead to constitutive activation of the receptor. In UC, the R248C, S249C, G372C, and Y375C mutations constitute approximately 95% of all identified mutations (Figure 3). All of these involve amino acid changes from a non-cystein to a cystein, creating ligand independent homodimerization, though de novo disulphide linkage. Furthermore, they are associated with TD, possibly reflecting that these mutations have the greatest effect on the protein. Indeed, studies have shown that the most profound FGFR3 activation
is seen for the mutations causing the lethal TD disorder. Clearly, activating mutations in the FGFR3 gene cause reduction of chondrocyte proliferation. Conversely, cell lines transfected with a S249C mutant FGFR3 IIIb have been demonstrated to induce tumor formation when xenografted into nude mice. It is therefore interesting that identical activating mutations of FGFR3 may either suppress or promote proliferation, implying that the effects of FGFR3 signaling must be highly cell-type specific.

Figure 3: Schematic view of the FGFR3 gene. The FGFR3 protein consists of three extracellular immunoglobulin-like domains (IgI-IgIII), a transmembrane domain (TM) and a split intracellular tyrosine-kinase domain (TKI and TKII). Somatic mutations identified in UC are given in relation to the gene structure. The distribution of FGFR3 mutations is obtained from a total of 849 cases; 784 previously published cases as reviewed by Knowles, and 65 cases with FGFR3 mutations identified in the present study. For each mutation, the germline associated form of skeletal dysplasia is also indicated: thanatophoric dysplasia type I and II (TDI and TDII), achondroplasia (ACH), hypochondroplasia (HCH), Crouzon syndrome and acanthosis nigricans (CSAN), and severe hypochondroplasia with developmental delay and acanthosis nigricans (SADDAN).
Apart from a low proportion of cervical carcinomas\textsuperscript{129,149,150} and multiple myelomas,\textsuperscript{151} FGFR3 mutations are not commonly found in other malignancies.\textsuperscript{150,152} Recently, however, FGFR3 mutations have been identified in 39\% of seborrheic keratoses,\textsuperscript{153} 85\% of adenoid seborrheic keratoses,\textsuperscript{154} and 33\% of nonorganoid, nonepidermolytic epidermal nevi,\textsuperscript{155} which are benign skin lesions with papillary growth. These skin lesions are related with another skin lesion, acanthosis nigricans, which is found in patients with some of the FGFR3-associated germline skeletal disorders.\textsuperscript{141} It has also been demonstrated that targeted expression of an activated \textit{fgfr3} mutant to the epidermis of mice, induce the development of benign epidermal tumors.\textsuperscript{153}

In UC, a striking association between \textit{FGFR3} mutation and tumor stage and grade has consistently been observed.\textsuperscript{53-61,111} In a large study comprising 260 primary UCs, \textit{FGFR3} mutations were found in 77\% of Ta, but only in 31\% of T1 and in 15\% of T2-T4 tumors.\textsuperscript{56} A similar distribution was also observed within the present study (Table 1). Even within the group of stage T1 tumors differences are seen with respect to \textit{FGFR3} mutation status; T1 tumors with microinvasion show a significantly higher mutation rate than extensively invading T1 tumors.\textsuperscript{131} Furthermore, mutations have been identified in 75\% of urothelial papillomas, a benign urothelial lesion.\textsuperscript{55} This suggests that the \textit{FGFR3} mutation contribute to papillary tumor formation of urothelial and epidermal lesions, but probably does not alone contribute to a malignant phenotype.

The possible importance of FGFR3 signaling in UC raises the possibility of using therapeutic drugs directed at FGFR3 or downstream mediators of signaling such as RAS. For example, CHIR-258, a FGFR3 inhibitor has been shown to inhibit growth of multiple myeloma cell lines with activated FGFR3, both \textit{in vitro} and \textit{in vivo} mouse models, results favoring future clinical studies.\textsuperscript{156}
TP53

TP53 is one of the most vital genes in a cell’s protection from DNA damage and aberrant growth signals. Impairment of TP53 function is seen in most human malignancies, commonly by DNA point mutations or deletions of the coding sequence, but also by alterations of regulatory proteins such as MDM2 or p14ARF, or by inactivation through viral infection. In the present study, we observed TP53 mutations in 9% of G1, 12% of G2, and 43% of G3 tumors (Table 1), frequencies similar to what has previously been reported. Hence, a correlation between TP53 mutation and tumor stage and grade is apparent; mutations are much more common in high grade and invasive tumors. Most analyses of TP53 alterations have however been performed by investigating aberrant protein expression in tissue sections using immunohistochemistry (IHC). In these studies, a similar association with increasing tumor stage and grade has been observed, and a good correlation between TP53 sequence mutations and TP53 immunoreactivity has been reported. Even though TP53 alterations are highly associated with adverse tumor phenotypes, published reports are contradictory regarding TP53 as an independent marker for disease outcome. A newly published meta-analysis, evaluating the results from 117 studies comprising more than 10,000 patients, concluded that TP53 changes are only weakly predictive of recurrence, progression, and mortality in bladder cancer.

TP53 inactivation may also be accomplished by overexpression of the MDM2 protein, which acts by promoting proteosomal degradation of TP53. Amplification or overexpression of MDM2 is not uncommonly seen in high grade tumors. Interestingly, by using a combination of markers, such as TP53 and MDM2, or TP53, MDM2, and CDKN1A (p21), independent predictive information for poor survival in UC has been observed.
Molecular Pathways in UC Development

Based on histologic mapping data, Koss\textsuperscript{165} suggested the existence of two different developmental pathways in UC; one primarily giving rise to papillary low grade tumors and one nonpapillary in which dysplasia and Tis subsequently develops into muscle-invasive and metastasizing tumors. Subsequent molecular analyses have further strengthened this theory. For example, several studies have found that chromosome 9 alterations present in hyperplasias also seen in concomitant or recurrent papillary tumors\textsuperscript{79,166,167} In contrast, Tis and dysplasia regularly display changes intimately associated with high grade and invasive tumors\textsuperscript{76-78,168-171}. Thus, both histologic and molecular evidence suggest that at least two different developmental pathways exist in UC. During recent years, much attention has been given to the \textit{FGFR3} and \textit{TP53} genes, and it has been suggested that alterations of these genes may be key events in two such pathways\textsuperscript{18,19,56,60}.

For \textit{FGFR3}, the extremely high mutation frequency in Ta tumors, contrasted by a low mutation frequency in invasive tumors and Tis\textsuperscript{54,56,61}, undoubtedly suggests that oncogenic signaling by \textit{FGFR3} is of vital importance for the development of non-invasive papillary tumors. A recent study performed by Jebar \textit{et al.} showed that activating mutations of \textit{RAS} gene family members, which are possible downstream mediators for FGFR3 induced activation of the \textit{RAS-ERK} signaling pathway, are found mutually exclusive to \textit{FGFR3} mutations\textsuperscript{128}. Further evidence for the \textit{RAS-ERK} pathway to be of crucial importance in development of papillary tumors was presented by Zhang \textit{et al.}\textsuperscript{172}. The authors introduced urothelium-specific expression of mutant Ha-ras in transgenic mice and observed that, \textit{in vivo}, activated Ha-ras induce urothelial tumor formation. Interestingly, the mice only developed hyperplasia and papillary low grade tumors. Thus, increased signaling via the Ras-ERK pathway, by either \textit{FGFR3} or \textit{HRAS} mutations, may be a key event in the development of hyperplasia and papillary tumors in UC.
In contrast to papillary tumors, impairment of the TP53 pathway, by TP53 mutations or MDM2 overexpression, is frequently observed in Tis lesions and in the more advanced tumor stages. Further evidence that TP53 inactivation may trigger development of Tis and subsequent progression into invasive tumors was reported by Zhang et al. The authors found that transgenic mice with urothelial-directed expression of SV40, which specifically inhibits TP53 and RB1, developed Tis and invasive tumors, but not hyperplasia and low grade papillary tumors.

Thus, oncogenic activation of FGFR3/HRAS and impairment of TP53 may be key genetic events in two alternative pathways of UC. The former promotes formation of papillary low grade tumors, possibly preceded by hyperplasia, whereas the latter pathway often seems to involve TP53 and lead to Tis and subsequent invasive tumors. In the present study, we could show that the difference between FGFR3- and TP53-mutated tumors is also reflected by large differences in gene expression profiles. In addition, we observed that tumors with TP53 mutation and MDM2 overexpression have highly similar global gene expression patterns, further suggesting that alterations of MDM2 may be a common and complementary mechanism for TP53 pathway inactivation. It must, however, be noted that, albeit with low frequency, FGFR3 mutated tumors sometimes do progress into high grade and invasive lesions, and when corrected for tumor stage/grade FGFR3 mutation does not seem to add any prognostic information. Nevertheless, the high grade and invasive FGFR3 mutated tumors often show TP53 impairment, implying that FGFR3 mutated tumors may acquire subsequent TP53 mutations and the two pathways may thus ultimately converge upon muscle-invasion.
Tumor Clonality and Tumor Spread

As previously noted, a characteristic feature of bladder cancer is the high occurrence of multiple (synchronous) and recurrent (metachronous) tumors. To explain the origin of these events, two major hypotheses have been put forth. The first hypothesis suggests tumor spread from a primary tumor by either intraepithelial migration or intraluminal seeding, thereby assuming a monoclonal spread of tumor cells. The second theory assumes field cancerization of large urothelial areas caused by exposure of carcinogens, and that this leads to the development of independent tumors. Thus, in the field cancerization model, tumors of different clonal origin will develop in the bladder.

To address these hypotheses, the genetic relationship between syn- and metachronous tumors has been investigated in a large number of molecular studies. For example, X-chromosome inactivation studies have been performed, as well as cytogenetic analyses, LOH, FISH, CGH, genome-wide single nucleotide polymorphism (SNP) arrays, and mutation analyses. Based on the accumulated data, it may be concluded that the vast majority of syn- and metachronous tumors shows a clonal relationship. It is however important to note that discrepancies in the definition of monoclonality and oligoclonality are found between the above-mentioned studies. Monoclonality may be used to describe clonal relation to a common progenitor cell. In this scenario, synchronous or metachronous tumors may harbor different additional changes acquired during a subsequent genetic divergence. Alternatively, tumors may be described as oligoclonal if such differences are found. Accordingly, if markers for genetic changes responsible for early tumor development are omitted from the analysis (e.g., only late occurring genetic changes are studied), no genetic relationship will be seen and the tumors may be classified as oligoclonal even though they still may be descendants of a single progenitor clone. Nevertheless, irrespective of what
terminology is used, the results point to that the leading mechanism behind tumor spread is clonal and that tumors originate from common progenitor cells. This process, however, often seems to be associated with the accumulation of sub-populations carrying additional and different genetic changes.

Intriguingly, during recent years histologic and genetic mapping studies have shown that large and continuous areas of malignant and pre-malignant pre-neoplastic lesions as well as morphologically normal-looking mucosa share genetic alterations. For example, identical TP53 mutations, are not only found within synchronous tumors at different locations, but also in pre-cancerous lesions and normal-looking urothelium separating the different tumor foci. Moreover, it has been shown that, even though clonally related, later appearing tumors often show less complex patterns of changes than earlier occurring tumors, i.e., the chronology of tumor recurrence does not adhere to the genetic evolution, a finding also corroborated within the present study.

Thus, the reported data strongly suggests that metachronous tumors rather originate from a pool of progenitor cells, contained within the urothelium, and that clonally related tumors, but with different secondary changes, may develop independently.

In light of this, an alternative theory, the field-first-tumor-later model, was recently presented. This proposes that, rather than an intraurothelial migration from a primary tumor, an intraepithelial spread of nonmalignant but genetically modified cells occur. The subsequent accumulation of genetic changes in these clones ultimately may produce overt tumors at different locations within the bladder. In this scenario, several different but clonally related subclones may coexist within the bladder simultaneously.
GENE EXPRESSION PROFILING OF UROTHELIAL CARCINOMAS

In 2001, the first study of gene expression patterns in UC using microarray was published by Tykjaer et al.\textsuperscript{70} who subjected 19 tumor biopsies and pooled single cell suspensions to microarray analysis on a 6.5k Affymetrix array. Although a rather small set of samples, the authors observed that expression of a large amount of genes correlated with tumor histopathology. A number of subsequent studies, using different array platforms and independent cohorts of samples, corroborated that much of the transcriptional variation observed in UC associates with tumor stage and grade, as reflected by a recurrent cosegregation of similar tumor stages and grades in unsupervised analyses.\textsuperscript{63-71,II}

Gene Expression Patterns and Tumor Histopathology

Dyrskjøt et al.\textsuperscript{65} showed that it is possible to molecularly predict UC histopathology based on gene expression profiles using supervised learning algorithms. Utilizing a 7k Affymetrix array, a 32-gene classifier was built that provided correct classification in 84\% of Ta, 50\% of T1, and 74\% of T2 tumors in a test set of 68 tumors.\textsuperscript{65} Similar classification results have later been obtained in independent studies using 10k cDNA arrays,\textsuperscript{63} and 22k Affymetrix arrays.\textsuperscript{68} Although the individual gene lists used for classification differ significantly between studies, Blaveri et al.\textsuperscript{63} showed that a high prediction accuracy could be maintained both when the classifier previously described by Dyrskjøt et al.\textsuperscript{65} was applied on their own data, and vice versa. Taking into consideration that a large number of genes have shown differential expression between superficial and invasive tumors, many genes may behave similarly for classification purposes and differences between individual classifiers may therefore be expected.\textsuperscript{63}
A common finding is that covarying genes with related biologic function are found to associate with specific histologic subtypes. For example, superficial tumors, especially low grade Ta tumors, frequently show increased expression of ribosomal genes, possible reflecting a high level of protein synthesis.\textsuperscript{70,118,11} Invasive tumors on the other hand, frequently show coordinated overexpression of genes with immunological function, and genes with structural and remodeling related to the ECM, \textit{e.g.}, collagens, metalloproteases, cathepsins, and fibronectin.\textsuperscript{63-67,70,185,11} Immunological transcripts have also been shown to correlate with Tis.\textsuperscript{64,186} As expression of ECM-related transcripts is not characteristic for epithelial cells, several authors have raised concerns that the high expression of ECM-related transcripts may merely reflect contamination of the analyzed sample with non-neoplastic stromal cells. Nevertheless, this group of genes has been found to be associated with invasive tumors in most investigations, despite tumor specimens with very low amount of contaminating non-neoplastic cells.\textsuperscript{185,186} For example, upregulation of ECM-related genes, \textit{e.g.}, MMP7 and \textit{COL3A2}, has been observed and subsequent protein expression analyses using IHC on tissue sections confirmed MMP7 expression in tumor cells whereas \textit{COL3A2} was only expressed in the surrounding stromal tissue.\textsuperscript{185} Moreover, expression of \textit{CTGF}, overexpressed in samples from high stage tumors,\textsuperscript{63,11} has been shown to promote angiogenesis and tumorigenesis in prostate cancer models when overexpressed in adjacent stromal cells.\textsuperscript{187} Hence, albeit some genes may be expressed by other cells than tumor cells, the identification of such transcripts in biopsies may still be important for understanding the biology of aggressive tumor growth, as well as for tumor classification purposes.

It has also been observed that the expression of genes promoting cell-cycle progression (\textit{e.g.}, \textit{TOP2A}, \textit{CDC2}, \textit{CCNA2}, \textit{CDC6}, and \textit{PCNA}) show positive association with tumor morphology,\textsuperscript{63,65,71,11} possibly reflecting the high
proliferative phenotype of low differentiated tumors. Interestingly, this gene cluster corresponds to a specific gene signature common for high grade tumors identified in a meta-analysis of expression data from several different human malignancies, further substantiating the importance of these genes for neoplastic progression.

**Gene Expression in FGFR3 and TP53 Mutated Tumors**

Our own studies, as well as previous reports, have revealed that FGFR3 is highly expressed in Ta tumors and that FGFR3 gene expression is highly associated with the presence of FGFR3 mutation. Overexpression of FGFR3 has also been observed to associate with a papillary growth pattern. As previously discussed, much evidence exist that FGFR3 and TP53 may be key players in two separate pathways of UC development. HCA data have indicated that FGFR3 mutated tumors cosegregate in unsupervised cluster analyses. In line with this, we found that the global gene expression patterns are highly associated with FGFR3 and TP53 mutation status in early-stage tumors. A similar pattern was seen in the subsequent study of T1-T4 tumors. In this study we also noted that cases with overexpression of MDM2, an alternative way for TP53 impairment, clustered with TP53 mutated cases. Moreover, we observed that high expression of cell-cycle transcripts was a characteristic feature of TP53mut as compared with FGFR3mut cases. Notably, tumors with concomitant FGFR3 and TP53 mutations showed high expression of FGFR3 but an intermediate expression of cell-cycle genes, suggesting that FGFR3mut tumors may genetically progress and accumulate TP53 mutations with an associated increase in expression of proliferative transcripts.
Prediction of Disease Recurrence and Progression

Superficial Tumor Recurrence

Genes associated with superficial tumor recurrence in patients with Ta tumors has been addressed in two studies. Dyrskjøt et al. built a 39-gene classifier that by CV showed a predictive accuracy of 75%. Also, in the present study, we identified genes associated with short-recurrence free follow up. Even though the gene signatures presented in the two studies do not overlap on a gene-by-gene basis, both gene lists are enriched for transcripts with cell-adhesion related functions, suggesting that this particular group of genes may be of importance for development of superficial recurrences.

Progression of Superficial Tumors

Efforts have also been made to identify transcripts which may be used for prediction of tumor recurrence and progression. In a detailed investigation specifically designed to predict superficial tumor progression, a number of tumor suppressor genes and genes involved in apoptotic cell death pathways (e.g., SERPINB5, FAT, BIRC4, and BIRC6) were found with higher expression in non-progressing samples. Interestingly, the FGFR3 gene was also highly expressed in non-progressing samples. Examples of genes found upregulated in progressing cases were the cell cycle related genes CDC25B, CDC20, and MCM7, and the apoptosis inhibitor BIRC5 (survivin). However, when testing the predictor on an independent test set, a relatively low specificity was observed. The authors pointed out that this may be caused by successful treatment and limited follow-up data. This was supported by a more detailed investigation of the material which revealed that most patients with misclassified samples showed adverse clinopathologic signs during follow-up, thus substantiating the possible biologic and clinical relevance of the identified
gene signature.\textsuperscript{189} Stage progression of Ta tumors has also been addressed by Wild \textit{et al.},\textsuperscript{186} who obtained 79\% correct classification using a CV approach. For example, these investigators found \textit{FABP4} and \textit{CTSE} to be upregulated in Ta tumors from patients that showed subsequent progression. A following IHC analysis for CTSE protein expression on tissue arrays supported the association between increased CTSE expression and progression of Ta tumors but did not provide independent prognostic information for overall survival.

\textbf{Progression of Muscle-Invasive Disease and Overall Survival}

For patients with muscle invasive tumors, three studies have specifically searched for gene signatures associated with tumor metastasis or cancer-associated death.\textsuperscript{63,68,II} Identification of a poor-survival molecular signature could be of great importance to select patients that may benefit from more aggressive therapeutical interventions. Blaveri \textit{et al.}\textsuperscript{63} stratified their muscle-invasive samples in two extreme groups; one with poor prognosis (death occurring in <18 months, n=27) and one with superior prognosis (survival >18 months, n=13). They applied a PAM classifier that correctly classified 78\% of the samples using a predictor composed of 24 genes. Later, Sanchez-Carbayo \textit{et al.}\textsuperscript{68} used two different approaches to identify markers for poor prognosis. A leave-one-out SVM algorithm was used with overall survival as end-point and predicted 74\%. Of these, \textit{HCLS1}, \textit{ANK3}, \textit{BIRC3}, \textit{CD54}, and \textit{TP53AP1} were studied in detail with log-rank and Kaplan-Meier tests, and all were significantly associated with poor patient outcome. Using an alternative approach, they further identified a list of 174 probes which expression associated with poor survival and lymph node metastasis. One of the top ranking genes from the latter analysis, \textit{SNCA}, was validated using IHC on tissue microarrays and was confirmed to be significantly associated with tumor stage and overall survival. In \textsc{Article II}, we investigated a homogeneous cohort of cases treated with radical cystectomy and identified a gene signature which
could predict the presence of lymph node metastasis with high specificity and sensitivity. Furthermore, the gene signature was superior to clinical stage in pre-operative multivariate models for prediction of nodal status and survival, and showed an impressive predictive power for overall survival when combined with histopathological tumor stage. This was the first report that showed a significant effect of the identified gene signature after adjustment for known prognostic parameters in multivariate analyses. The identified gene signature contained transcription factors such as \textit{STAT1} and \textit{IRF1}, and contained transcripts for several structural components in the MCH class I antigen presentation machinery. We therefore hypothesize that lymph node negative tumors may be more susceptible to tumor suppression mediated by the immune system.

None of the above mentioned studies applied an independent set of samples to test the performance of the gene expression predictors, instead cross-validation was used. Hence future follow-up studies are therefore needed to confirm the suggested clinical value. Furthermore, no overlap on a gene-to-gene basis is observed between the gene signatures described. Explanation for this could be differences in sample selection. For example, Sanchez-Carbayo \textit{et al.}\textsuperscript{68} do not report treatment for the included patients, and since patients with muscle-invasive disease not treated with cystectomy show far higher mortality than untreated patients, this could be a confounding factor. Also, multivariate analyses correcting for known prognostic markers should be applied, this was only performed in one study.\textsuperscript{11} Nevertheless, three independent investigations have identified expression patterns associated with survival and nodal status, raising hope for the possibility to, through molecular approaches, select patients who may benefit from more aggressive therapeutical interventions.
Normal and Morphologically Altered Bladder Mucosa

Detailed investigations comparing tumor and normal urothelial tissue from healthy individuals have been performed in two studies. A general increase in transcripts related to protein folding, mRNA splicing, energy pathways, glucose catabolism, and lipid metabolism was observed in tumor samples compared to normal tissue, suggesting that the transformation event is accompanied by an increased need of energy and mRNA splicing capability. Moreover, both studies found KRT7, KRT8, and SDC1 to be among the most discriminatory genes, raising the possibility to use these genes as potential markers for noninvasive tumors. A subsequent screening of KRT7 expression in urine sediments by Western blot did however not provide sufficient sensitivity, encouraging further studies using more sensitive methods.

Studies have also included normal-looking urothelium derived from bladder cancer patients. Interestingly, these studies have reported transcriptional alterations, characteristic for high grade tumors, to be present in the normal-looking urothelial samples. Specifically, it was observed that the gene signature predicting poor-outcome also was present in normal-looking urothelium samples taken from the same bladder. Similarly, morphologically normal urothelial samples from cystectomy patients has been shown to carry an expression signature specific for Tis samples. Thus, expression analyses corroborate other molecular studies showing that genetic alterations are present already in normal-looking urothelium from patients with bladder cancer.

Multifocal and Recurrent Tumors

It has been observed that syn- and metachronous tumors, respectively, often cluster together, despite differences in stage and grade, suggesting a higher similarity in global gene expression than to be expected. These findings
support the common clonal relationship between tumor recurrences as seen in many previous studies using other molecular techniques. The close genetic relationship between syn- and metachronous tumors was also confirmed within the present study. In fact, we could show that the correlation in expression profiles between such samples were remarkably high and close to what is seen when comparing normal samples from different individuals. Furthermore, the high similarity was observed even among cases with large karyotypic differences as detected by LOH and CGH analyses. This indicates that the establishment of a stable expression profile may be considered as an early event in UC.

**Concluding Remarks**

In conclusion, expression profiling using microarray has proved to be valuable technique for exploration of the molecular genetic events that characterize UC. Specific gene signatures, whose expression are related to histologic, morphologic, molecular, and clinical outcome groups, have been reported, thus offering new biologic insights into the pathogenesis of the disease. Furthermore, the findings may not only be important for classification and prediction purposes, but may also be essential for identification of molecular pathways to which novel treatment strategies may be directed.
CONCLUSIONS

In the present study, we used gene expression profiling, mutation analyses of *FGFR3* and *TP53*, LOH analysis, as well as conventional- and array CGH analyses to characterize a large cohort of UCs. The main findings may be summarized as follows:

- gene signatures with specific biological functions correlate with, e.g., high expression of ECM and immunologic genes in invasive tumors, overexpression of cell-cycle genes in *TP53* mutated tumors, overexpression of ribosomal transcripts in low grade tumors,

- mutations in *FGFR3* and *TP53* are correlated with large transcriptional differences and may thus be indicative for two different molecular pathways of UC,

- high expression of MHC-class I related genes in patients with muscle-invasive tumors provide prognostic information for lymph-node status and overall survival,

- the expression profile is remarkably stable in meta- and synchronous Ta or T1 tumors from the same patient. The formation of a fixed expression profile may be considered as a primary event in the development of UC,

- *E2F3*, *CDKAL1*, and *SOX4* in 6p22, and *YWHA* (14-3-3) gene family members in 8q22 and 2p25, respectively, are major target oncogenes important for high grade disease.
SUMMARY IN SWEDISH

Varje år nydiagnostiseras cirka 2 000 personer med blåscancer i Sverige. Blåscancer drabbar främst män och är för närvarande den femte vanligaste cancerformen bland män efter prostata-, hud-, lung- och koloncancer. Risken att insjukna i blåscancer ökar med ålder och merparten av nydiagnostiserade patienter är över 60 år. Blåscancer är en heterogen sjukdom med avseende på sjukdomsförlopp. Fler än hälften (ca 75 %) av alla patienter har ytliga tumörer, dvs tumörer som endast växer i urinblåsans ytskikt. Av dessa ytliga tumörer uppvisar de flesta (ca 60 %) låg grad av cellförändring och även om dessa patienter ofta får återfall så är tumörutvecklingen sällan aggressiv och överlevnaden är över 95 % efter 5 års uppföljning. Resterande del av ytliga tumörer uppvisar högre grad av cellförändring och i denna patientgrupp är risken hög att tumören skall växa in i, eller igenom, blåsmuskulaturen (muskelinvasiv sjukdom). Ungefär en fjärdedel av alla blåscancerpatienter har redan utvecklat muskelinvasiv tumör vid diagnostillfället. För denna grupp är prognosen dålig och ungefär hälften avlider av tumörmetastaser, ofta redan inom två år.

Cancer uppstår och utvecklas som en följd av successiva förändringar (mutationer) i en cells arvsmassa (gener). I denna flerstegsprocess krävs att flera av cellens kontrollmekanismer sätts ur spel vilket slutligen leder till okontrollerad cellväxt – en tumör bildas. Trots att många studier har genomförts i syfte att kartlägga de genetiska förändringar som orsakar blåscancer så är kunskapen fortfarande relativt begränsad. Att vidare utforska varför tumörerna uppstår och varför vissa tumörer utvecklas snabbare och mer aggressivt än andra är därför av stor betydelse.

Vi har i våra studier använt oss av så kallad microarray-teknologi. Microarray är en teknik som studerar uttrycksnivåerna av tusentals gener för
varje tumörprov. Man kan således med denna teknik få en övergripande bild av vilka gensystem som är påverkade i blåscancer. Vår övergripande målsättning var att med denna metod utforska de genetiska mekanismer som ligger bakom blåscancers uppkomst samt vilka förändringar som sker vid tumörutveckling från ytlig till muskelinvasiv cancer. Identifering av sådana förändringar kan både ge förklaring till det spridda sjukdomsförloppet som observeras hos patienter och dessutom ge möjlighet till ny diagnostik, d v s en möjlighet att identifiera de patienter som har större risk för aggressiv sjukdom.


Som tidigare nämnts så är tumöråterfall vanligt bland blåscancerpatienter. I Artikel III studerade vi därför förhållandet mellan återfallstumörer inom samma patient. Våra resultat gav ytterligare stöd till tidigare framförd teorier om hur en tumör sprids och ger upphov till återfall. Studien visade bland annat att alla tumörer hos en och samma patient uppvisar snarliga genuttrycksförändringar. Detta tyder på att tumörerna är besläktade
med varandra och att de har utvecklas från en och samma genetiska förändrade ursprungscell.

Ett vanligt sätt att aktivera gener som bidrar till tumörutveckling är kopietalsförökning. Denna relativa förökning av en gens kopietal kan leda till ökat genuttryck och således ökad genaktivitet. I Artikel IV studerade vi sådana förändringar i tumörer från 38 patienter med blåcancer. Vi kunde visa att kopietalsökning av gener som CDKAL1, E2F3, SOX4, YWHAQ och YWHAZ, leder till förhöjt genuttryck och därför kan vara av betydelse för utveckling av aggressiv cancer.

Sammanfattningsvis så har våra studier av blåscancer bidragit med värdefull information om vilka genetiska system som är förändrade i blåscancer och hur blåscancer utvecklas. Vi har dessutom kunnat koppla ett antal förändringar till specifika undergrupper av tumörer, information som kan vara av betydelse för framtida molekylär diagnostik eller behandlingsformer.
ACKNOWLEDGEMENTS

Many people have contributed to this work and supported me during my years at the Department of Clinical Genetics. I would like to express my sincere gratitude to the following:

Mattias Höglund, my supervisor. Thank you for accepting me as your student, for being so supportive, enthusiastic and generous with your time, and for all stimulating discussions we have had.

Professor Felix Mitelman for giving me the opportunity to conduct my Ph.D. studies at the Department of Clinical Genetics, and making this laboratory into such a stimulating environment.

The group of dedicated researchers in Lund and Helsingborg who have made these studies of bladder cancer possible: The excellent urologists Wiking Månsson, Fredrik Liedberg, and Sigurður Gudjónsson, from the department of Urology; Gunilla Chebil and Lena-Maria Lundberg at the departments of Pathology in Helsingborg and Lund, respectively; Imad Fadl-Elmula who introduced bladder cancer research to our department; and Mårten Fernö, Kristina Lövberg, and Pär-Ola Bendahl from the department of Oncology. Thank you for all constructive and enjoyable meetings we have had during these years. The time and effort you all have put into the different studies has truly been a great source of inspiration to me.

My friends and fellow Ph.D. students at the department of Clinical Genetics: Aikaterini, Anna A, Anna D, Emely, Björn, Charlotte, David, Helena, Henrik, Josef, Karin, Karoline, Malin, Markus, Petra, Srinivas, Therese, Tord, and Ylva who really have made these years into something extra! In particular, I would like to thank: Anna Andersson, my true microarray-buddy, with whom I have shared countless hours of hybridization experiments. Thank you for all cheerful talks we have had (whether about microarrays, statistical methods that we just can’t understand, or other, not-so-scientific topics); Markus Heidenblad, for great collaborations and fun moments (as well as on the golf course); Srinivas Veerla, for your amazing kindness and for aiding me with your programming expertise; and Kajsa Paulsson, for all enjoyable discussions, and for taking the time to constructively comment my works.

Carin Lassen and Margareth Isaksson, for always finding time to help me out in the lab.
Acknowledgements

The senior research staff at the Department of Clinical Genetics: Bertil, Catarina, Fredrik, Kristina, Karin, Ludmilla, Mia, Nils, Nina, Ioannis, Samuel, Thoas, Ulf, and Yuesheng. In particular I would like to thank Thoas Fioretos for all interesting microarray-related meetings and discussions, and for the appreciated comments on my works.

The rest of the personnel at the Department of Clinical Genetics for making this place such a nice place to work at!

Sonja Aits, Christer Halldén, and Christina Lind-Halldén for great collaboration with the sequencing and LOH analyses

Professor Åke Borg and the people at the Department of Oncology and the Swegene DNA microarray resource centre. I would like to thank Carina, Göran, Johan V, Johan S, Lao, and Sofia for making the tedious work of clone preparation and microarray testing into a truly enjoyable experience.

To all the co-authors for the stimulating collaborations.

My family, Margareta, Göran, and Hanna for the tremendous support you given me over the years. I appreciate it more than you ever could imagine.

Frida, Erika, and Tarja. Thank you for giving me so many fun moments at Fäladen.

Finally, a big up to my all my friends representing Gbg and Malmö

The work presented herein was supported by funds from: the Swedish Cancer Society; the Swedish Research Council; the Finnish Cancer Society; the Gunnar, Arvid and Elisabeth Nilsson Foundation; the Crafoord Foundation; the John and Augusta Persson Foundation; the IngaBritt and Arne Lundberg Foundation; the Maud and Birger Gustavsson Foundation; the Petrus and Augusta Hedlund Foundation; and the Sigrid Jusélius Foundation. The microarray facility was supported by the Knut and Alice Wallenberg Foundation via the Swegene Program.
REFERENCES

References

References


References


References


188. Rhodes DR, Yu J, Shanker K, et al. Large-scale meta-analysis of cancer microarray data identifies common transcriptional profiles of neoplastic