A Molecular Taxonomy for Urothelial Carcinoma

Gottfrid Sjödahl1, Martin Lauss1, Kristina Lövgren1, Gunilla Chebil1, Sigurdur Gudjonsson2, Srinivas Veerla1, Oliver Patschan2, Mattias Aine1, Mårten Fernö1, Markus Ringnér1, Wiking Månsson2, Fredrik Liedberg2,4, David Lindgren1,3, and Mattias Höglund1

Abstract

**Purpose:** Even though urothelial cancer is the fourth most common tumor type among males, progress in treatment has been scarce. A problem in day-to-day clinical practice is that precise assessment of individual tumors is still fairly uncertain; consequently efforts have been undertaken to complement tumor evaluation with molecular biomarkers. An extension of this approach would be to base tumor classification primarily on molecular features. Here, we present a molecular taxonomy for urothelial carcinoma based on integrated genomics.

**Experimental Design:** We use gene expression profiles from 308 tumor cases to define five major urothelial carcinoma subtypes: urobasal A, genomically unstable, urobasal B, squamous cell carcinoma like, and an infiltrated class of tumors. Tumor subtypes were validated in three independent publically available data sets. The expression of 11 key genes was validated at the protein level by immunohistochemistry.

**Results:** The subtypes show distinct clinical outcomes and differ with respect to expression of cell-cycle genes, receptor tyrosine kinases particularly FGFR3, ERBB2, and EGFR, cytokertains, and cell adhesion genes, as well as with respect to FGFR3, PIK3CA, and TP53 mutation frequency. The molecular subtypes cut across pathologic classification, and class-defining gene signatures show coordinated expression irrespective of pathologic stage and grade, suggesting the molecular phenotypes as intrinsic properties of the tumors.

Available data indicate that susceptibility to specific drugs is more likely to be associated with the molecular stratification than with pathologic classification.

**Conclusions:** We anticipate that the molecular taxonomy will be useful in future clinical investigations.

*Clin Cancer Res; 18(12); 3377–86. ©2012 AACR.*

Introduction

Bladder cancer is the fourth most common tumor type among males. More than 90% of bladder cancers are urothelial cell carcinoma and about 5% are squamous cell carcinoma (SCC). The gender ratio of male to female is 3 to 1 and the best known environmental risk factor is smoking. Urothelial cell carcinoma patients are stratified by pathologic stage and grade; the basis of clinical decision-making. The stage classification differentiates between nonmuscle invasive (NMI; Tis, Ta, and T1) and muscle-invasive tumors (T2, T3, and T4) according to the invasion depth. Ta tumors are restricted to the urothelium; T1 tumors have invaded the lamina propria; and T2, T3, and T4 tumors have invaded the superficial muscle, perivesical fat, and surrounding organs, respectively. Tis is poorly understood and believed to be a precursor of muscle-invasive tumors. The majority of patients, 70%, initially present with NMI tumors, however, up to 70% of these develop local recurrences, and patients may have several recurrences. Roughly 25% of NMI patients progress to muscle-invasive tumors disease with a potential to develop metastasis. One problem in day-to-day clinical practice is that pathologic assessment is reported to be fairly uncertain (1–3). Accordingly, there have been efforts to complement the pathologic evaluation with biomarkers that can be judged in a more objective manner (4–6). A further extension of this approach would be to base a tumor classification system primarily on molecular features, integrating molecular data from several biologic levels. An advantage of such an approach would be that a more comprehensive description of existing tumor subtypes could be attained. One method is to apply gene expression data to stratify tumors based on molecular phenotypes. Only a limited number of high-throughput gene expression analyses of bladder cancer have, however, been carried out (7–12) and the main focus has been on the identification of gene signatures with possible prognostic
Translational Relevance
An important factor for optimal cancer treatment is correct tumor classification. In this investigation, we define five molecular subtypes of bladder cancer that show significant differences in prognosis. The suggested subtypes are defined by distinct gene expression signatures specific for cell cycle, cytokeratin, cell adhesion, receptor tyrosine kinases, and immune response. The class-defining gene signatures show coordinated expression irrespective of pathologic stage and grade, indicating the molecular subtypes as intrinsic properties of the tumors. Hence, our proposed molecular stratification adds valuable additional information to current pathologic staging and grading. A systematic analysis revealed that specific drug target profiles were associated with individual subtypes. We anticipate that the suggested molecular classification will be valuable in future evaluations of urothelial carcinoma and help to define clinicogenomic subtypes of importance for new therapeutic strategies.

RNA extraction, labeling and hybridization, and preprocessing of expression data
Tumor samples were thawed in RNAlater ICE (Ambion), disrupted and homogenized using TissueLyser (Qiagen) and QiAshedder (Qiagen), and RNA extracted using AllPrep or RNeasy kits (Qiagen). RNA quality was assessed on Agilent 2100 Bioanalyzer (Agilent). Labeling and hybridization to Direct Hyb HT-12 V3 BeadArrays (Illumina) were carried out by the SCIBLU facility at Lund University (http://www.lth.se/sciblu). Preprocessing and quality control steps are described in detail in Supplementary Text S1. A 50% intensity filter followed by merging of probes for the same gene resulted in 13,953 genes used for supervised analyses. For unsupervised analysis, a further 50% variance filter was applied. Raw and processed data, together with sample annotations, are deposited at the Gene Expression Omnibus (GSE32894).

Statistical analyses
Molecular subtypes were identified through a step-wise procedure using hierarchical clustering analyses of bootstrapped data sets as described in Supplementary Text S1. For validation, the same procedure was applied to data sets of Stransky and colleagues (7), Sanchez-Carbayo and colleagues (8), and Kim and colleagues (12; Supplementary Text S1). Quality threshold clusters (QTC) were defined by a minimum correlation of 0.5 and a minimum of 20 genes for each cluster (15). Sample classification was carried out in a leave-one-out cross-validation loop using ANOVA or t test as feature selection method, and nearest centroid classification (NCC) as classification algorithm (16). When applied to independent data, the classifier was built using all 308 cases. Statistical analyses were conducted with R 2.9.2 (http://www.r-project.org) and Multi Experiment Viewer v. 4.7.4 (http://www.tm4.org) (17).

Tissue microarrays and immunohistochemistry
Tissue microarrays (TMA, duplicate 1.0-mm punches) were constructed for 275 cases and stained with 13 antibodies; ACTA2 [mouse monoclonal antibody (mAb) 1A4; Dako], CCNB1 (rabbit mAb Y106; Nordic Biosite), CCND1 (mouse mAb SP4; Dako), CCNE1 (mouse mAb 13A3; Leica Microsystems), CD3 (mouse mAb F7.2.38; Dako), EGF receptor (mouse mAb 3C6; Ventana), ERBB2 (rabbit mAb 4B5; Ventana), FGFR3 (rabbit mAb C51F2; Cell Signaling), KRT5 (rabbit mAb EP1601Y; Lab Vision), KRT6 (rabbit mAb EPR1603Y; Nordic Biosite), KRT14 (mouse mAb LL002; Lab Vision), KRT20 (mouse mAb K20.8; Dako), TP63 (mouse mAb 4A4; Imgenex). As negative controls, the primary antibodies were omitted for each staining.
Results

Defining urothelial cell carcinoma molecular subtypes

Hierarchical cluster analysis of the 308 samples indicated the presence of several tumor clusters. To firmly establish these results, we used a successive 2 group split approach (Supplementary Text S1). The first split grouped the tumors into MS1 and MS2 subtypes described by us previously (11). MS1 and MS2 tumors were then treated individually to establish further divisions, each division being subjected to several quality tests. This procedure was repeated resulting in a total of 7 tumor clusters (Fig. 1). We then carried out an ANOVA based on 13,953 genes and used the 7 clusters as grouping variable; a total of 8,377 genes showed a significant association with molecular subtype. This shows that a large proportion of the assayed genes are in fact associated with the identified tumor clusters. The overall structure of the tumor classification was corroborated in 3 external data sets using an identical unsupervised bootstrap analysis and organization into molecular subtypes (Supplementary Text S1; Supplementary Fig. S1). We then derived NCC classifiers using 2 different feature selection methods, resulting in a classification accuracy of 91% in both cases (LOOCV), and when applied to the independent Kim data an accuracy of 81% (Supplementary Text S1). On the basis of the above results, we conclude that urothelial cell carcinoma may be robustly classified into at least 7 distinct gene expression clusters.

Biologic characterization of urothelial cell carcinoma molecular subtypes

To reveal biologic themes specific for the tumor clusters, we examined groups of genes with coordinated expression as well as genes selected based on their biologic functions. We also investigated subtype-specific distribution of mutations in the FGFR3, PIK3CA, and TP53 genes. Below we summarize key differences between the subtypes exemplified in Fig. 2. More detailed biologic interpretations of the data are provided in Supplementary Text S2. These analyses resulted in the definition of 5 major molecular subtypes of urothelial cancer; urobasal A (MS1 subdivided into MS1a and MS1b), genomically unstable (MS2a subdivided into MS2a1 and MS2a2), urobasal B (MS2b2.1), SCC-like (MS2b2.2), and one highly infiltrated by nontumor cells (MS2b1; Fig. 2). Importantly, these molecular subtypes showed distinct survival patterns in which urobasal A showed good prognosis, genomically unstable and the infiltrated group intermediate, and the urobasal B and the SCC-like the worst prognosis (Fig. 3A). The differential behavior of the suggested molecular subtypes was also evident in high-grade (G3) tumors, a class of tumors strongly associated with adverse prognosis (Fig. 3B).

Urothelial cell carcinoma molecular subtypes show different immune and wound healing gene signatures

Genes with coordinated expression across the samples were identified using QTC (Supplementary Fig. S2; QTC1–18). Two QTC gene clusters had a prominent activated T-cell theme including key genes ranging from T-cell stimulatory chemokines, T-cell receptor complex genes, as well as signal transducers and effector genes of cytotoxic T cells (Fig. 2; Supplementary Text S2). An additional immune-related cluster contained several genes involved in chemotaxis of the neutrophil/monocyte lineage, indicating the presence of myeloid cells. A fourth gene cluster significantly enriched for ECM genes was highly expressed in MS2b1 and included several genes for collagens, proteoglycans, and basal lamina components (Supplementary Text S2). This signature also included a number of genes known to be specifically expressed in myofibroblasts, notably ACTA2, VIM, and PDPN. The expression pattern of these 4 QTC signatures indicates that the gene expression profile of particularly MS2b1 is heavily compromised by tumor-infiltrating cells, such as T lymphocytes and myofibroblasts, and possibly also by endothelial cells (Supplementary Text S2). The presence of T cells and myofibroblasts was validated by immunohistochemistry (IHC) using antibodies for CD3 and ACTA2 (Supplementary Text S2) and motivated the characterization of the MS2b1 as infiltrated.

Urothelial cell carcinoma molecular subtypes show different cell-cycle gene signatures

A cell-cycle gene cluster (Supplementary Fig. S2; QTC3), showed high expression in 3 of the MS2 subtypes and...
moderate in one, MS2b1.1. The majority of the QTC3 genes was associated with the S, G2, and M phases and did not include genes typical for the G1 phase. We therefore carried out a supervised selection of genes with key regulatory functions of the cell cycle and selected the top ANOVA genes ($P < 10^{-10}$). The resulting 46 genes formed 2 distinct gene expression patterns with one group of early cell-cycle genes, for example, CCND1, predominantly expressed in the MS1 tumors, and one group of late cell-cycle genes, for example, CCNE, CCNA, and CCNB expressed in MS2a and MS2b2.2 tumors (Fig. 2, Supplementary Text S2). CCND1, CCNE1, and CCNB1 protein expression was validated by IHC (Fig. 4). In addition to CCND1, MS1 cases showed expression of 3 ID gene family members and of RBL2 (Fig. 2). The association of MS2a and MS2b2.2 tumors with late cell-cycle activity was underlined by the high expression of the CCNB activators CDC25A, CDC25B, and CDC25C, as well as of genes related to chromosome segregation and cell division, such as BUB1, CDC20, and CENP genes. Taken together, the expression pattern of the cell-cycle genes suggested that cell-cycle activity in MS1 tumors is primarily engaged in releasing the cells from G0 to G1, that is, associated with CCND1 expression, whereas the pattern in MS2a and MS2b2.2 indicated that these tumors have evaded the cell-cycle restriction point and are associated with CCNE expression.

Urothelial cell carcinoma molecular subtypes show different cytokeratin signatures

Keratins of simple epithelial cells, KRT8/KRT18 and KRT7/KRT19, were predominantly expressed in MS1 and MS2a (Fig. 2, Supplementary Text S2). KRT20, specifically
Urothelial cell carcinoma molecular subtypes show different cell adhesion gene signatures

Several cell-adhesion genes showed significant differential expression across the molecular subtypes (Fig. 2, Supplementary Text S2).
Supplementary Fig. S3). Urobasal A and genomically unstable tumors expressed tight junction associated genes, for example, claudins, albeit with slightly different profiles. The SCC-like tumors, as well as the urobasal B showed low claudin expression, except for CLDN1. These subtypes also showed a similar adherence junction profile with CDH3 (P-cadherin) as the most prominent component. Desmosome-related genes cadherins, desmogleins, and desmocollins showed low expression in genomically unstable and high expression in urobasal A, urobasal B, and in SCC-like tumors. A similar pattern was seen for the gap junction, hemidesmosome genes and integrins. These results indicated that urobasal A tumors maintain epithelial cell–cell and cell–matrix contacts and that genomically unstable tumors have progressively more disrupted cell adhesion structures further away from the apical side of normal epithelial cells. The opposite is seen for SCC-like and urobasal B tumors that have lost expression of the majority of tight junction genes but maintain expression of genes associated with basolateral cell adhesion.

Urothelial cell carcinoma molecular subtypes are independent of pathologic stratification

The defined molecular subtypes do not overlap with pathologic stratification (Fig. 5). Even though Ta tumors are dominated by the urobasal A subtype, T1 tumors are composed of urobasal A and genomically unstable cases, and muscle-invasive tumors cases may be of any subtype. Low-grade tumors, G1 and G2, are predominantly of the urobasal A subtype, whereas G3 tumors may be of any subtype (Fig. 5B). Finally, when limiting the analysis to nonmuscle invasive high-grade tumors (T1G3), it may be concluded that these are very heterogeneous at the molecular level (Fig. 5C). Hence, the molecular subtypes differentiate the tumors within each pathologic entity further and add additional information for tumor classification. We
grouped each molecular subtype into pathologic stage (Fig. 6A) and grade (Fig. 6B) and then estimated mean expression levels for subtype-defining gene signatures for each class. This showed that the early cell-cycle gene signature, specific for urobasal A, was expressed irrespective of pathologic stage and grade, and that the late cell-cycle signature, specific for genomically unstable tumors, also was expressed independent of stage and grade. Similarly, key receptor tyrosine kinases FGFR3 and ERBB2 showed subtype-specific expression, independent of pathologic stage and grade (Fig. 6, Supplementary Text S2); FGFR3 in the urobasal A and B, ERBB2 in the genomically unstable subtype. Subtype-specific receptor expression was validated at the protein level by IHC (Fig. 4B). Hence, the molecular phenotype is stable across pathologic stage and grade, emphasizing the molecular subtypes as intrinsic and divergent properties of tumors within the same pathologic classification group.

**Urothelial cell carcinoma molecular subtypes differ in expression of possible drug targets**

We downloaded potential drug targets from the Drugbank database (18) with a described or potential use in cancer (Supplementary Text S2). Of 60 genes that were targetable, 39 were expressed in a subtype-specific pattern (ANOVA, Bonferroni corrected $P < 0.05$). We then searched The Cochrane Central Register of Controlled Trials for compounds in clinical trial for use in cancer patients. The obtained list was refined to include only drugs with described gene targets. This resulted in 46 compound–target pairs of which 37 showed subtype-specific expression in the current data set. In Fig. 2 we showed a heat map of a selected number of drug target genes (see Supplementary Text S2 for all target genes). Importantly, gene expression of potential drug targets was associated with molecular subtype rather than with pathologic stratification, as exemplified by the targets for tipifarnib and valrubicin, both tested in clinical trials for urothelial cell carcinoma. (Fig. 6; refs. 19, 20).

**Discussion**

We carried out an extensive gene expression study of urothelial cell carcinoma particularly aimed at defining molecular subtypes of bladder cancer with the belief that such subtypes may be objectively assessed, biologically relevant, and function as a complement to the current pathologic classification. In a first step, we used robust statistical methods to arrive at well-separated groups of
tumors. Independently, we applied the same strategy to 3 previously published bladder cancer data sets and could thus validate the overall structure of the tumor subtypes in independent data. We investigated the biologic significance of the subtypes by identifying coexpressed genes as well as the expression patterns of selected genes. Using this approach, we identified 5 major subtypes of urothelial cell carcinoma with distinct biologic and clinical properties; urobasal A, genomically unstable, urobasal B, SCC-like, and infiltrated. The infiltrated subtype showed a very strong immunologic and ECM signal, indicating the presence of immunologic and myofibroblast cells. This subtype most likely represents a heterogeneous class of tumors as IHC revealed the presence of tumors with typical genomically unstable, urobasal B, and SCC-like protein expression patterns within this group.

Urobasal A tumors were characterized by elevated expression of FGFR3, CCND1, and TP63, as well as KRT5 gene expression in cells at the tumor–stoma interface. In addition, urobasal A tumors showed very good prognosis. The importance of FGFR3 was shown by frequent FGFR3 mutations, high FGFR3 expression, and strong expression of the FGFR3 gene signature. The FGFR3 gene signature includes TP63, a member of the TP53 family of transcriptional regulators, with a basal/intermediate expression in the normal urothelium (21) and crucial for normal urothelial differentiation (22). TP63 may have a direct influence on FGFR3 expression as the FGFR3 gene has TP63-responsive promoter elements and is activated by TP63 (23). A further characteristic of urobasal A was the expression of CCND1, RBL2, and the ID genes. CCND1 is expressed in the basal and suprabasal cell layers of the normal urothelium. ID2 is known to interact with RBL2 and may influence the activity of the RBL2-E2F4/F5 complexes that inhibit cell growth in the G0 phase (24). Hence, the urobasal A tumors show activity of cell-cycle genes operating before the cell-cycle restriction point, indicating a phenotype reminiscent of undifferentiated urothelial cells, that is, basal or intermediate. This observation was underscored by the finding that urobasal A tumors expressed KRT5, KRT13, KRT15, and KRT17, with the same cellular patterns as is seen in normal urothelium. The majority of the urobasal A tumors were nonmuscle invasive and of low pathologic grade. The low pathologic grade is in line with the finding that these tumors, in contrast to the genomically unstable tumors, have retained expression of most cell adhesion genes important for the epithelial architecture of the cell layers. The genomically unstable subtype was characterized by frequent TP53 mutations, CCNE and ERBB2 expression, and low cytokeratin expression. Genomically unstable cases represent a high-risk group as close to 40% were muscle invasive. This subtype also showed low PTEN expression and thus coincides with the high-risk urothelial cell carcinoma described by Puzio-Kuter and colleagues (ref. 25; Supplementary Text S2). Several genes previously associated with tumor progression, recurrence, or positive cytology, were found to be upregulated within the genomically unstable group, for example, KPNA2, HMOX1 (27), and CTSL1 and CTSL2 (28). It would thus be expected that a large fraction of the genes associated with this subtype would show prognostic values with similar magnitudes to the reported ones. A major difference between the urobasal A and the genomically unstable subtype was that the latter showed increased activity of late G1 phase, CCNE, and late cell-cycle genes, for example, CCNA, CCNB, and CDC20. Hence, genomically unstable tumors may have created a short circuit that evades the cell-cycle restriction point. In contrast, the urobasal A tumors, genomically unstable tumors did not show expression of the basal/intermediate cytokeratins, but rather of KRT20, associated with umbrella cells (29, 30). At first hand this may seem contradictory. However, He and colleagues (31) have shown that the basal phenotype, defined as KRT17-positive and KRT20-negative cells, is only maintained in the tumor–stoma interface, and when tumor cells lose stromal contact, parts of the normal differentiation program is activated, including KRT20 expression. Our data indicate that a similar effect is seen for several of the uroplakin genes, also expressed in the umbrella cells of normal urothelium. The majority (>70%) of the genomically unstable tumors was of high grade and had lost expression of most cell adhesion genes, except those normally associated with the apical tight junctions. This makes high pathologic grade a significant feature of the genomically unstable group of tumors.

The SCC-like subtype was characterized by high expression of basal keratins normally not expressed in the urothelium, KRT4, KRT6A, KRT6B, KRT6C, KRT14, and KRT16, as well as by bad prognosis. As these keratins have been associated with squamous differentiation of urothelial cell carcinoma (10, 32–34), we applied the bladder SCC gene signature of Blaveri and colleagues (8) to our data, which underscored this conclusion. This finding was validated by pathologic reevaluation by which the majority of the cases showed signs of squamous cell differentiation. Furthermore, this group showed a different proportion of female/male patients compared with the remaining cases, reminiscent of the 1:1 proportions seen in patients diagnosed with bladder SCC, suggesting that females are more likely to develop urothelial carcinomas with a keratinized/squamous phenotype associated with an adverse prognosis.

The urobasal B tumors showed several similarities to the urobasal A tumors, such as a high FGFR3 mutation frequency, elevated FGFR3, CCND1, TP63 levels, and expression of the FGFR3 gene signature. This group, however, showed frequent TP53 mutations and expression of several keratins specific for the SCC-like subtype. In addition, 50% of the cases were muscle invasive; including 5 of 9 FGFR3 mutated cases. Altogether, our data suggest this subtype as an evolved/progressed version of urobasal A. Importantly, tumor cells expressing SCC-associated cytokeratin also express FGFR3, CCND1, and TP63, typical for the urobasal A tumors, thus excluding tumor heterogeneity, that is, that...
an additional cell population show this phenotype. Apart from the urobasal A and urobasal B tumors, FGFR3 mutations were also present in the genomically unstable and SCC-like tumors. However, whereas the urobasal B cases maintained expression of the FGFR3 gene signature, this signature was lost in the 2 other subtypes. This may indicate that if a FGFR3 mutated urobasal A tumor evolves to a genomically unstable or SCC-like phenotype, dependence on FGFR3 activity is overridden by other changes and that the presence of FGFR3 mutations is a sign of the tumor history only.

An important aspect of the suggested classification is the independence from pathologic stratification. Cases classified as genomically unstable included tumors with pathologic stages Ta, T1, as well as muscle-invasive tumors, and reversely, T1G3 tumors contained representatives from at least 4 of the 5 subtypes. In particular, several of the class-defining gene signatures showed coordinated expression irrespective of pathologic stage and grade, indicating the molecular subtypes as intrinsic properties of the tumors. Importantly, the subtypes showed different outcome also when looking at high-grade tumors separately. Hence, our proposed molecular stratification adds valuable additional information to current pathologic staging and grading. Particularly, we expect that molecular phenotype will have a greater influence on tumor behavior and treatment response to, for example, chemotreatment, than pathologic stratification.

Disclosure of Potential Conflicts of Interest

The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript. No potential conflicts of interest were disclosed.

Acknowledgments

The authors thank Pontus Eriksson for developing scripts for TMA image handling.

Authors’ Contributions

Conception and design: G. Sjödahl, W. Månnson, M. Högland
Acquisition of data: G. Sjödahl, G. Chebil, S. Gudjonsson, O. Patschan, M. Fernö, W. Månnson, F. Lindberg, D. Lindgren
Analysis and interpretation of data: G. Sjödahl, M. Lauss, S. Veerla, M. Aine, M. Ringner, M. Högland

Writing, review, and/or revision of the manuscript: G. Sjödahl, M. Lauss, S. Gudjonsson, O. Patschan, M. Fernö, M. Ringner, F. Lindberg, D. Lindgren, M. Högland

Administrative, technical, or material support: G. Sjödahl, K. Lövgren, S. Veerla, O. Patschan, M. Fernö, W. Månnson, D. Lindgren

Study supervision: M. Högland

Grant Support

This work was supported by the Swedish Cancer Society, the Swedish Research Council, the Lund University Hospital, The Crafoord Foundation, and the Gunnar Nilsson Cancer Foundation.

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.

Received January 23, 2012; revised March 19, 2012; accepted April 22, 2012; published OnlineFirst May 2, 2012.

References
