A Molecular Taxonomy for Urothelial Carcinoma.

Sjödahl, Gottfrid; Lauss, Martin; Lövgren, Kristina; Chebil, Gunilla; Gudjonsson, Sigurdur; Veerla, Srinivas; Hultman Patschan, Oliver; Aine, Mattias; Fernö, Mårten; Ringnér, Markus; Månsson, Wiking; Liedberg, Fredrik; Lindgren, David; Höglund, Mattias

Published in:
Clinical Cancer Research

DOI:
10.1158/1078-0432.CCR-12-0077-T

2012

A molecular taxonomy for urothelial carcinoma

Gottfrid Sjödahl[1], Martin Lauss[1], Kristina Lövgren[1], Gunilla Chebil[1], Sigurdur Gudjonsson[2], Srinivas Veerla[1], Oliver Patschan[2], Mattias Aine[1], Mårten Fernö[1], Markus Ringnér[1], Wiking Månsson[2], Fredrik Liedberg[2,3], David Lindgren[1,4], Mattias Höglund[1]

[1] Department of Oncology, Clinical Sciences, Lund University Hospital, Lund University, Lund, Sweden
[2] Department of Urology, Clinical Sciences, Lund University Hospital Malmö, Lund University, Malmö, Sweden
[3] Section of Urology, Växjö County Hospital, Växjö, Sweden
[4] Department of Laboratory Medicine, Center for Molecular Pathology, Malmö University Hospital, Lund University, Malmö, Sweden

Running title: Molecular subtypes of urothelial carcinoma

Key Words: Bladder cancer, gene expression, classification, gene signature, molecular subtypes,

Corresponding Author: Mattias Höglund, Department of Oncology, Clinical Sciences, Lund University Hospital, Lund University, Lund, 221 85, Sweden. Phone: +46(0)462220393; E-mail: mattias.hoglund@med.lu.se

Word Count: 4577

Total Figures: 6 Total Tables: 0

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, SCC-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*, cytokeratins, and cell adhesion genes, as well as with respect to *FGFR3, PIK3CA*, and *TP53* mutation frequency. The molecular subtypes cut across pathological classification, and class defining gene signatures show coordinated expression irrespective of pathological 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 pathological classification.

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

Statement of translational relevance

An important factor for optimal cancer treatment is correct tumor classification. In the present 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, cytokeratins, cell adhesion, receptor tyrosine kinases, and immune response. The class-defining gene signatures show coordinated expression irrespective of pathological stage and grade, indicating the molecular subtypes as intrinsic properties of the tumors. Hence, our proposed molecular stratification adds valuable additional information to current pathological 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.

Introduction

Bladder cancer is the fourth most common tumor type among males. More than 90% of bladder cancers are urothelial cell carcinoma (UC) 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. UC patients are stratified by pathological stage and grade; the basis of clinical decision-making. The stage classification differentiates between non-muscle invasive (NMI) (Tis, Ta, and T1) and muscle-invasive tumors (MI) (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 MI tumors. The majority of patients, 70%, initially present with NMI tumors, however, up to 70% of these develop local recurrences, and patients may frequently have recurrences. Roughly 25% of NMI patients progress to MI disease with a potential to develop metastasis. One problem in day-to-day clinical practice is that pathological assessment is reported to be fairly uncertain (1-3). Accordingly, there have been efforts to complement the pathological 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 biological 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 performed (7-12) and the main focus has been on the identification of gene signatures with possible prognostic values. For example Kim et al. (12) described a gene profile for progression as well as for response to BCG treatment, Sanchez-Carbayo et al. (8) and Blaveri et al. (9) reported gene profiles for tumor staging and disease-

specific survival, and Dyrskjot et al. (10) have reported profiles for staging, recurrence, and progression. Very little has been done, though, to use expression data to investigate the existence of inherent molecular subtypes that may complement current histopathological classification systems. Such a molecular classification system has been developed for, e.g., breast cancer in which four main classes of tumors have been defined; Luminal A, Luminal B, HER2-enriched, and Basal-like, which show different clinical outcomes (13, 14). In a recent study, we defined two molecular subtypes of UC governed by distinct biological processes and mutation profiles (11). In the present investigation, we have extended the molecular classification of UC in a much more comprehensive analysis of 308 tumors allowing the definition of five major molecular subtypes of UC; Urobasal A, Genomically unstable, Urobasal B, SCC-like, and a heterogeneous infiltrated class of tumors. These subtypes show distinct molecular profiles, differ in survival rates, and can be validated in publicly available data. They furthermore cut across pathological staging and grading and may thus add valuable additional molecular information to pathological classification.

Materials and Methods

Tumor Samples

Urothelial carcinomas were collected by cold-cup biopsies from the exophytic part of the bladder tumor in 308 patients undergoing transurethral resection at hospitals of the Swedish southern healthcare region. Informed consent was obtained from all patients and the study was approved by the Local Ethical Committee of Lund University. Detailed pathological and clinical data is given in Table S1, and summarized in table S2.

RNA extraction, labeling and hybridization, preprocessing of expression data

Tumor samples were thawed in RNAlater ICE (Ambion), disrupted and homogenized using TissueLyser (Qiagen) and Qiashredder (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 performed by the SCIBLU facility at Lund University (http://www.lth.se/sciblu). Preprocessing and quality control steps are described in detail in 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 (HCA) of bootstrapped datasets as described in Text S1. For validation, the same procedure was applied to data sets of Stransky et al. (7) Sanchez-Carbayo et al. (8) and Kim et al., (12) (Text S1). Quality threshold clusters (QTCs) were defined by a minimum correlation of 0.5 and a minimum of 20 genes for each cluster (15). Sample classification was performed 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 performed using R 2.9.2 (http://www.r-project.org) and TMEV (17).

Tissue microarrays and immunohistochemistry

Tissue microarrays (duplicate 1.0 mm punches) were constructed for 275 cases and stained with 13 antibodies; ACTA2 (mouse 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), EGFR (mouse mAb 3C6 Ventana), ERBB2 (rabbit mAb 4B5 Ventana), FGFR3 (rabbit mAb C51F2 Cell Signaling), KRT5 (rabbit mAb EP1601Y Labvision), KRT6 (rabbit mAb EPR1603Y Nordic Biosite), KRT14 (mouse mAb LL002 Labvision), KRT20 (mouse mAb K_s20.8 Dako), TP63 (mouse mAb 4A4 Imgenex). As negative controls, the primary antibodies were omitted for each staining.

Results

Defining UC 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 two-group split approach (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 seven tumor clusters (Figure 1). We then performed an ANOVA based on 13 953 genes and used the seven 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 three external data sets using an identical unsupervised bootstrap analysis and organization into molecular subtypes (Text S1, Figure S1). We then derived Nearest Centroid Classification (NCC) classifiers using two 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% (Text S1). Based on the above results we conclude that UC may be robustly classified into at least seven distinct gene expression clusters.

Biological characterization of UC molecular subtypes

To reveal biological themes specific for the tumor clusters we examined groups of genes with coordinated expression as well as genes selected based on their biological 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 Figure 2. More detailed biological interpretations of the data are provided in Supplementary Text S2. These analyses resulted in the definition of five 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 non-tumor cells (MS2b1) (Figure 2). Importantly, these molecular subtypes show distinct survival patterns where Urobasal A shows good prognosis, Genomically unstable and the infiltrated group intermediate, and the Urobasal B and the SCC-like the worst prognosis, also in grade 3 tumors only (Figure 3).

UC molecular subtypes show different immune and wound healing gene signatures

Genes with coordinated expression across the samples were identified using quality threshold clustering (Figure 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 (Figure 2, 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 (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 four 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 (Text S2). The presence of T cells and myofibroblasts was validated by immunohistochemistry using antibodies for CD3 and ACTA2 (Text S2) and motivate the characterization of the MS2b1 as *Infiltrated.*

UC molecular subtypes show different cell cycle gene signatures

A cell cycle gene cluster (Figure S2; QTC3), showed high expression in three of the MS2 subtypes and moderate in one, MS2b2.1. 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 performed 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 two distinct gene expression patterns with one group of early cell cycle genes, e.g., *CCND1*, predominantly expressed in the MS1 tumors, and one group of late cell cycle genes, e.g., *CCNE*, *CCNA*, and *CCNB* expressed in MS2a and MS2b2.2 tumors (Figure 2, Text S2). CCND1, CCNE1, and CCNB1 protein expression was validated by immunohistochemistry (Figure 4). In addition to *CCND1*, MS1 cases showed expression of three *ID* gene family members and of *RBL2* (Figure 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 suggests that cell cycle activity in MS1 tumors is primarily engaged in releasing the cells from G0 to G1 i.e., associated with *CCND1* expression, whereas the pattern in MS2a and MS2b2.2

indicates that these tumors have evaded the cell cycle restriction point and are associated with *CCNE* expression.

UC molecular subtypes show different cytokeratin signatures

Keratins of simple epithelial cells, *KRT8/KRT18* and *KRT7/KRT19*, were predominantly expressed in MS1 and MS2a (Figure 2, Text S2). *KRT20*, specifically expressed in differentiated umbrella cells, was expressed in a sub population of the MS1 tumors and in the MS2a tumors, but not in the other subtypes (Figure 2, Figure 4B). The association between *KRT20* expression and urothelial cell differentiation was confirmed by the finding that the *KRT20* expressing subtypes also expressed several uroplakin genes (Figure 2, Text S2). *KRT5*, *KRT13*, *KRT15*, and *KRT17*, basal/intermediate cytokeratins, showed high expression in MS1 as well as in MS2b2 cases. IHC data showed that MS1 tumors maintain degrees of urothelial stratification and express the basal KRT5 in cells in contact with the basal membrane, similar to what is seen in the normal urothelium (Figure 4). Hence, as MS1 is characterized by a close to normal urobasal cytokeratin expression a better designation of this subtype would be *Urobasal A.* The MS2b2.2 tumors differed drastically from the other subtypes by showing elevated expression of *KRT6A*, *KRT6B*, *KRT6C*, *KRT14*, and *KRT16* (5-31 fold), indicating a keratinized/squamous phenotype (Figure 2, Figure 4B), further supported by expression of several keratinization-associated genes e.g., *SPRR2D*, *DSG3*, *LOX*, and *SERPINA1*. We scored the tumors with bladder cancer specific SCC signatures (9), which highlighted MS2b2.2 as SCC-like (Text S2). Furthermore, after pathological re-evaluation signs of squamous differentiation were seen in 18 of 29 MS2b2.2 cases. Hence, it is motivated to call this subtype *SCC-like*. Intriguingly, the gender proportion in MS2b2.2, with equal number of females and males, was significantly different compared with what is normally seen in UC (Text S2).

UC molecular subtypes show different mutation and FGFR3 gene expression signature profiles

FGFR3 mutation analysis revealed a distinct difference in mutation frequencies between Urobasal A (MS1) and MS2a cases (55% vs. 7%, $p<0.0001$, Chi-2) (Figure 2, Text S2). Urobasal A tumors also showed a higher *PIK3CA* mutation frequency compared to MS2a (25% vs. 8%, $p<0.002$, Chi-2), whereas the frequency of *TP53* mutations was significantly higher in MS2a (48% vs. 11%, $p<0.0001$, Chi-2). This identifies the Urobasal A as *FGFR3* and *PIK3CA* mutated, and MS2a as *TP53* mutated. The strong association between TP53 and the MS2a tumor cluster, and the fact that these tumors show grossly rearranged genomes (11) prompted us the rename this group to *Genomically unstable. FGFR3* and *PIK3CA* mutation frequencies in MS2b2.1 did not differ from the Urobasal A subtype ($p>0.25$, Chi-2), and *TP53* mutation frequencies in MS2b2.1 did not differ from frequencies in the Genomically unstable (MS2a) subtype ($p>0.6$, Chi-2). Furthermore, the *FGFR3* associated gene signature showed high expression in both the Urobasal A and the MS2b2.1 subtypes, but low in Genomically unstable and SCC-like tumors (Figure 2, Text S2). Apart from *FGFR3*, this signature included *TP63*, making high *CCND1*, *FGFR3*, and *TP63* expression a common denominator of Urobasal A and MS2b2.1 tumors (Figure 2, Figure 4). Furthermore, 10 out of 20 MS2b2.1 cases were muscle invasive, compared to 8 out of 130 Urobasal A tumors, indicating MS2b2.1 as a high risk variant of the Urobasal A. Apart from *TP53* mutations this subtype also show signs of a keratinized/squamous phenotype, but less pronounced than seen in the SCC-like. Immunohistochemistry showed that this feature is not caused by tumor heterogeneity as cells showing aberrant KRT5, KRT6, and KRT14 expression also show FGFR3, CCND1, and TP63 expression (Figure 4). A better designation of this tumor cluster would thus be *Urobasal B*, indicating the molecular link between this group and the Urobasal

A tumors (Figure 2). *FGFR3* mutated cases were also detected among the Genomically unstable and the SCC-like cases, albeit at lower frequencies. Importantly, whereas the Urobasal B cases maintain expression of the *FGFR3* gene signature, *FGFR3* mutated Genomically unstable and SCC-like cases show a considerable drop in *FGFR3* gene signature expression (Figure 2, Text S2).

UC molecular subtypes show different cell adhesion gene signatures

Several cell-adhesion genes showed significant differential expression across the molecular subtypes (Figure 2, Figure S3). Urobasal A and Genomically unstable tumors expressed tight junction associated genes, e.g. 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 show 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 indicate that Urobasal A tumors maintain epithelial cell-cell and cell-matrix contacts, 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.

UC molecular subtypes are independent of pathological stratification

The defined molecular subtypes do not overlap with pathological stratification (Figure 5). Even though Ta tumors are dominated by the Urobasal A subtype, T1 tumors are composed of Urobasal A and Genomically unstable cases, and MI 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 (Figure 5B). Finally, when limiting the analysis to non-muscle invasive high grade tumors (T1G3) it may be concluded that these are very heterogeneous at the molecular level (Figure 5C). Hence, the molecular subtypes differentiate the tumors within each pathological entity further and add additional information for tumor classification. We grouped each molecular subtype into pathological stage (Figure 6A) and grade (Figure 6B), and then estimated mean expression levels for subtype-defining gene signatures for each class. This demonstrated that the early cell cycle gene signature, specific for Urobasal A, was expressed irrespective of pathological stage and grade, and that the late cell cycle signature, specific for Genomically unstable tumors, also was expressed independently of stage and grade. Similarly, key receptor tyrosine kinases *FGFR3* and *ERBB2* showed subtype-specific expression, independent of pathological stage and grade (Figure 6, 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 (Figure 4B). Hence, the molecular phenotype is stable across pathological stage and grade, emphasizing the molecular subtypes as intrinsic and divergent properties of tumors within the same pathological classification group.

UC 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 (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 Figure 2 we show a heat map of a selected number of drug target genes (see Text S2 for all target genes). Importantly, gene expression of potential drug targets was associated with molecular subtype rather than with pathological stratification, as exemplified by the targets for tipifarnib and valrubicin, both tested in clinical trials for UC. (Figure 6) (19, 20).

Discussion

We performed an extensive gene expression study of UC 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 pathological 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 three previously published bladder cancer data sets, and could thus validate the overall structure of the tumor subtypes in independent data. We investigated the biological significance of the subtypes by identifying co-expressed genes as well as the expression patterns of selected genes. Using this approach we identified five major subtypes of UC with distinct biological and clinical properties; Urobasal A, Genomically unstable, Urobasal B, SCC-like, and Infiltrated. The Infiltrated subtype showed a very strong immunological and ECM signal, indicating the presence of immunological 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 demonstrated 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 urothelium 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 supra basal 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, i.e., 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 non-muscle invasive and of low pathological grade. The low pathological 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 UC described by Puzio-Kuter et al. (25) (Text S2). Several genes previously associated with tumor progression, recurrence, or positive cytology, were found to be up-regulated within the Genomically unstable group, e.g., *KPNA2*, (26), *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 e.g., *CCNA*, *CCNB*, and *CDC20*. Hence, Genomically unstable tumors may have created a short-circuit that evades the cell cycle restriction point. In contrast to 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 et al. (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 were of high grade and had lost expression of most cell adhesion genes, except those normally associated with the apical tight junctions. This makes high pathological 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 UC (10, 32-34) we applied the bladder SCC gene signature of Blaveri et al. (8) to our data, which

underscored this conclusion. This finding was validated by pathological 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 suggests this subtype as an evolved/progressed version of Urobasal A. Importantly, tumor cells expressing SCC-associated cytokeratins also express FGFR3, CCND1, and TP63, typical for the Urobasal A tumors, thus excluding tumor heterogeneity, i.e., 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 two 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 over-ridden 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 pathological stratification. Cases classified as Genomically unstable included tumors with pathological

stages Ta, T1, as well as MI, and reversely, T1G3 tumors contained representatives from at least four of the five subtypes. In particular, several of the class-defining gene signatures showed coordinated expression irrespective of pathological 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 pathological staging and grading. Particularly, we expect that molecular phenotype will have a greater influence on tumor behavior and treatment response to, e.g., chemo treatment, than pathologic stratification.

Acknowledgements

We thank Pontus Eriksson for developing scripts for TMA image handling.

Grant Support

The Swedish Cancer Society, the Swedish Research Council, the Lund University Hospital, The Crafoord foundation, and the Gunnar Nilsson Cancer foundation. The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

References

1. Ooms EC, Anderson WA, Alons CL, Boon ME, Veldhuizen RW. Analysis of the performance of pathologists in the grading of bladder tumors. Hum Pathol 1983;14:140-43.

2. van Rhijn BW, van der Kwast TH, Kakiashvili DM, Fleshner NE, van der Aa MN, Alkhateeb S, et al. Pathological stage review is indicated in primary pT1 bladder cancer. BJU Int 2010;106:206-11.

3. May M, Brookman-Amissah S, Roigas J, Hartmann A, Störkel S, Kristiansen G, et al. Prognostic accuracy of individual uropathologists in noninvasive urinary bladder carcinoma: a multicentre study comparing the 1973 and 2004 World Health Organisation classifications. Eur Urol 2010;57:850-58.

4. van Rhijn BW, Vis AN, van der Kwast TH, Kirkels WJ, Radvanyi F, Ooms EC, et al. Molecular grading of urothelial cell carcinoma with fibroblast growth factor receptor 3 and MIB-1 is superior to pathologic grade for the prediction of clinical outcome. J Clin Oncol 2003;21:1912-21.

5. van Oers JM, Wild PJ, Burger M, Denzinger S, Stoehr R, Rosskopf, et al. FGFR3 mutations and a normal CK20 staining pattern define low-grade noninvasive urothelial bladder tumours. Eur Urol 2007;52:760-8.

6. Forster JA, Paul AB, Harnden P, Knowles MA. Expression of NRG1 and its receptors in human bladder cancer. Br J Cancer 2011;104:1135-43.

7. Stransky N, Vallot C, Reyal F, Bernard-Pierrot I, de Medina SG, Segraves R, et al. Regional copy number-independent deregulation of transcription in cancer. Nat Genet 2006;38:1386-96.

8. Sanchez-Carbayo M, Socci ND, Lozano J, Saint F, Cordon-Cardo C. Defining molecular profiles of poor outcome in patients with invasive bladder cancer using oligonucleotide microarrays. J Clin Oncol 2006;24:778-89.

9. Blaveri E, Simko JP, Korkola JE, Brewer JL, Baehner F, Mehta K, et al. Bladder cancer outcome and subtype classification by gene expression. Clin Cancer Res 2005;11:4044-55.

10. Dyrskjøt L, Thykjaer T, Kruhøffer M, Jensen JL, Marcussen N, Hamilton-Dutoit S, et al. Identifying distinct classes of bladder carcinoma using microarrays. Nat Genet. 2003;33:90-96.

11. Lindgren D, Frigyesi A, Gudjonsson S, Sjödahl G, Hallden C, Chebil G, et al. Combined gene expression and genomic profiling define two intrinsic molecular subtypes of urothelial carcinoma and gene signatures for molecular grading and outcome. Cancer Res 2010;70:3463-72.

12. Kim WJ, Kim EJ, Kim SK, Kim YJ, Ha YS, Jeong P, et al. Predictive value of progression-related gene classifier in primary non-muscle invasive bladder cancer. Mol Cancer 2010;9:3.

13. Sørlie T, Perou CM, Tibshirani R, Aas T, Geisler S, Johnsen H, et al. Gene expression patterns of breast carcinomas distinguish tumor subclasses with clinical implications. Proc Natl Acad Sci U S A. 2001;98:10869-74.

14. Sotiriou C, Pusztai L. Gene-expression signatures in breast cancer. N Engl J Med. 2009;19:790-800.

15. Heyer LJ, Kruglyak S, Yooseph S. Exploring expression data: identification and analysis of coexpressed genes. Genome Res 1999;9:1106-15.

16. Dabney AR. Classification of microarrays to nearest centroids. Bioinformatics 2005;21:4148–54.

17. Saeed AI, Sharov V, White J, Li J, Liang W, Bhagabati N, et al. TM4: a free, open-source system for microarray data management and analysis. Biotechniques 2003;34:374-78.

18. Knox C, Law V, Jewison T, Liu P, Ly S, Frolkis A et al. DrugBank 3.0: a comprehensive resource for 'omics' research on drugs. Nucleic Acids Res 2011;39:D1035-41.

19. Rosenberg JE, Von der Maase H, Seigne JD, Mardiak J, Vaughn DJ, Moore M, et al. A phase II trial of R115777, an oral farnesyl transferase inhibitor, in patients with advanced urothelial tract transitional cell carcinoma. Cancer 2005;130:2035-41.

20. Steinberg G, Bahnson R, Brosman S, Middleton R, Wajsman Z, Wehle M, et al. Efficacy and safety of Valrubicin for the treatment of bacillus Calmette-Guerin refractory carcinoma in situ of the bladder. J Urol 2000;163:761-67.

21. Castillo-Martin M, Domingo-Domenech J, Karni-Schmidt O, Matos T, Cordon-Cardo C. Molecular pathways of urothelial development and bladder tumorigenesis. Urol Oncol 2010;28:401-08.

22. Karni-Schmidt O, Castillo-Martin M, HuaiShen T, Gladoun N, Domingo-Domenech J, Sanchez-Carbayo M, et al. Distinct expression profiles of p63 variants during urothelial development and bladder cancer progression. Am J Pathol 2011;178:1350-60.

23. Sayan AE, D'Angelo B, Sayan BS, Tucci P, Cimini A, Cerù MP et al. p73 and p63 regulate the expression of fibroblast growth factor receptor 3. Biochem Biophys Res Commun 2010;394:824-28.

24. Zebedee Z, Hara E. Id proteins in cell cycle control and cellular senescence. Oncogene 2001;20:8317-25.

25. Puzio-Kuter AM, Castillo-Martin M, Kinkade CW, Wang X, Shen TH, Matos T, et al. Inactivation of p53 and Pten promotes invasive bladder cancer. Genes Dev 15;23:675-680.

26. Jensen JB, Munksgaard PP, Sørensen CM, Fristrup N, Birkenkamp-Demtroder K, Ulhøi BP, et al. High Expression of Karyopherin-α2 Defines Poor Prognosis in Non-Muscle-Invasive Bladder Cancer and in Patients with Invasive Bladder Cancer Undergoing Radical Cystectomy. Eur Urol 2011;59:841-48.

27. Yim MS, Ha YS, Kim IY, Yun SJ, Choi YH, Kim WJ. HMOX1 is an important prognostic indicator of nonmuscle invasive bladder cancer recurrence and progression. J Urol 2011;185:701-05.

28. Staack A, Koenig F, Daniltchenko D, Hauptmann S, Loening SA, Schnorr D, et al. Cathepsins B, H, and L activities in urine of patients with transitional cell carcinoma of the bladder. Urology 2002;59:308-12.

29. Harnden P, Mahmood N, Southgate J. Expression of cytokeratin 20 redefines urothelial papillomas of the bladder. Lancet 1999;20:974-77.

30. Alsheikh A, Mohamedali Z, Jones E, Masterson J, Gilks CB. Comparison of the WHO/ISUP classification and cytokeratin 20 expression in predicting the behavior of low-grade papillary urothelial tumors. World/Health Organization/Internattional Society of Urologic Pathology. Mod Pathol 2001;14:267-72.

31. He X, Marchionni L, Hansel DE, Yu W, Sood A, Yang J, et al. Differentiation of a highly tumorigenic basal cell compartment in urothelial carcinoma. Stem Cells 2009;27:1487-95.

32. Chu PG, Lyda MH, Weiss LM. Cytokeratin 14 expression in epithelial neoplasms: a survey of 435 cases with emphasis on its value in differentiating squamous cell carcinomas from other epithelial tumours. Histopathology 2001;39:9-16.

33. Harnden P, Southgate J. Cytokeratin 14 as a marker of squamous differentiation in transitional cell carcinomas. J Clin Pathol 1997;50:1032-33.

34. Tungekar MF, Gatter KC, Al-Adnani MS. Immunohistochemistry of cytokeratin proteins in squamous and transitional cell lesions of the urinary tract. J Clin Pathol 1988;41:1288-96.

Figure Legends

Figure 1. UC tumor clusters.

(A) Top, hierarchical tree indicating the successive divisions producing the seven tumor clusters. Individual tumor clusters indicated by MS designations and color bars. Bottom, heat map of the top 500 ANOVA genes.

Figure 2. Excerpts of identified gene expression profiles.

Representative genes from described gene signatures. Red, high expression; green, low expression; black; mutation; white, wild-type; grey, no mutation data, NMI, number of non muscle invasive cases; MI, number of muscle invasive cases. The cell adhesion genes shown are structural components of the different cell adhesion complexes indicated in parentheses. TJ, tight junction; AJ, adherence junction; Des, desmosome; GJ, gap junction; HD, hemidesmosome; ITG, epithelial integrins.

Figure 3. Kaplan-Meier analyses.

(A) Including all stages and grades. (B) Including G3 tumors only. P-values according to log rank test. Only cases that received standard treatment were included.

Figure 4. IHC validation.

(A) Protein expression of subtype-specific markers. For Urobasal B the same case is shown for both FGFR3 and KRT5 staining. (B) TMA cores stained with 11 subtype-specific markers. Representative cases are shown for each of the original seven tumor clusters.

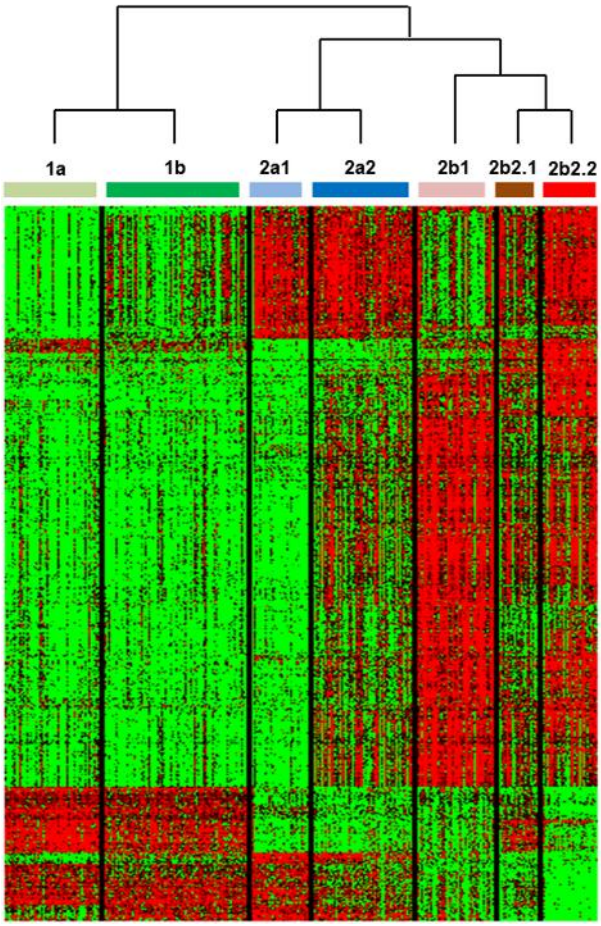
Figure 5. Molecular subtype and pathological stage/grade.

(A) Composition of Ta, T1, and MI tumors with respect to molecular subtype. (B) Composition of G1, G2, and G3 tumors with respect to molecular subtypes. (C) Composition of NMI tumors of different stage/grade with respect to molecular subtype.

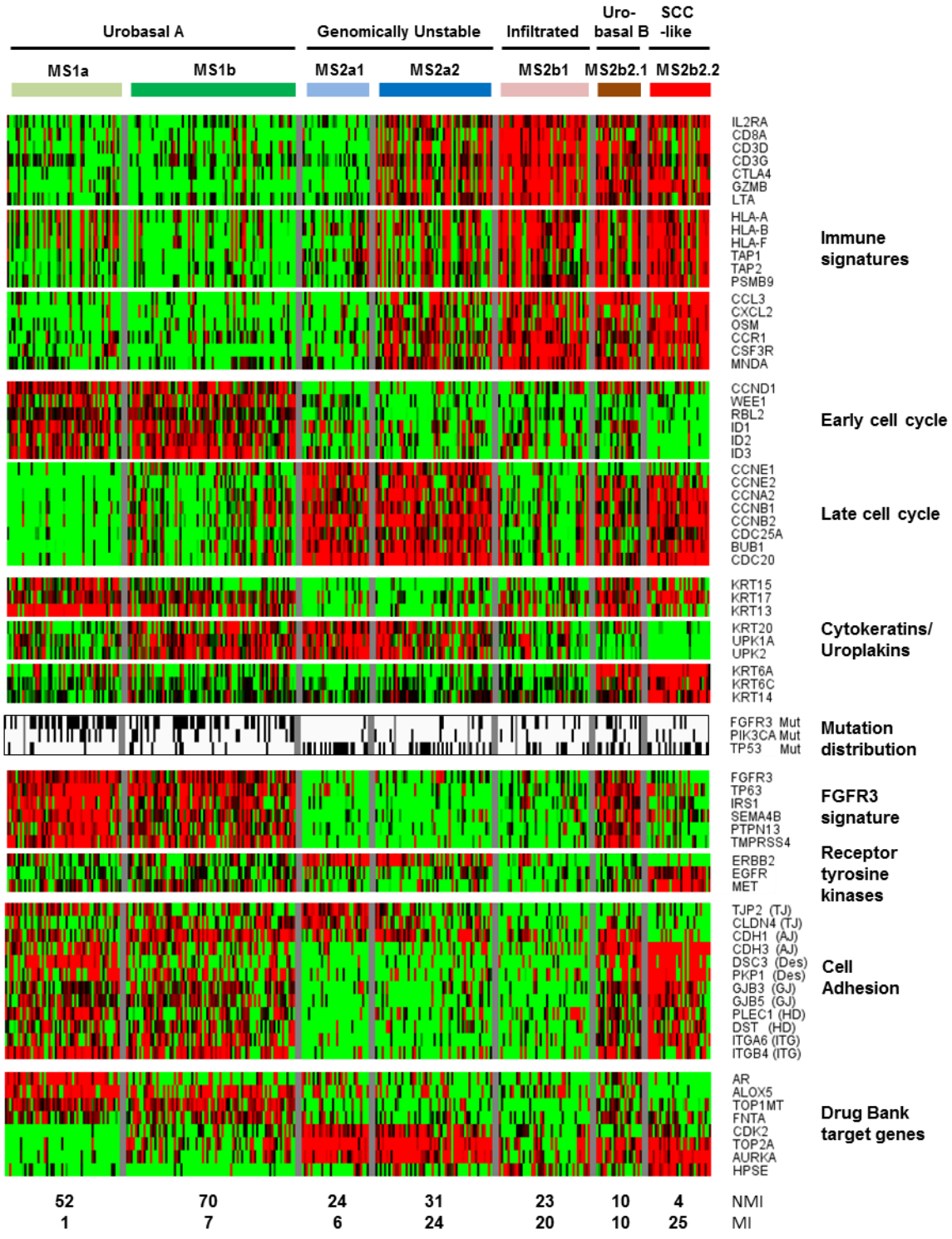
Figure 6. Molecular signatures are independent of tumor stage and grade.

Heat maps showing mean expression levels of gene signatures/individual genes in each of the five molecular subtypes stratified by tumor stage; Ta, T1, and MI (A), and by tumor grade (B). Expression of target genes for tipifarnib (FNTA) and valrubicin (TOP2A) is shown. Red, high expression; green, low expression; gray fields, insufficient data.

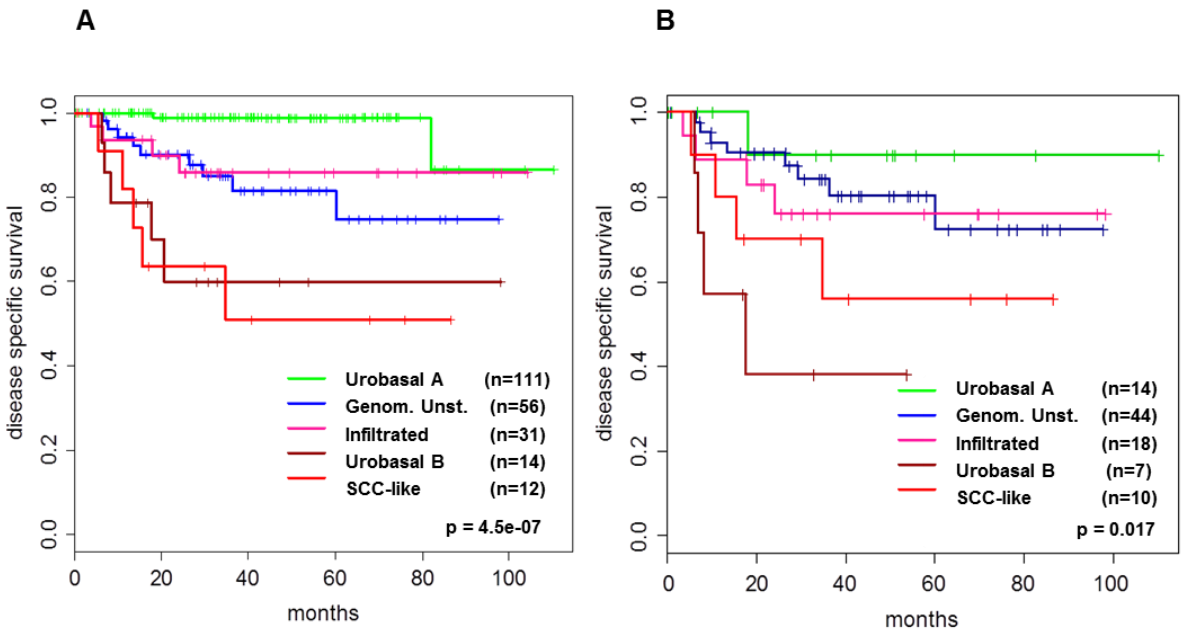
Sjödahl et al. Fig 1.



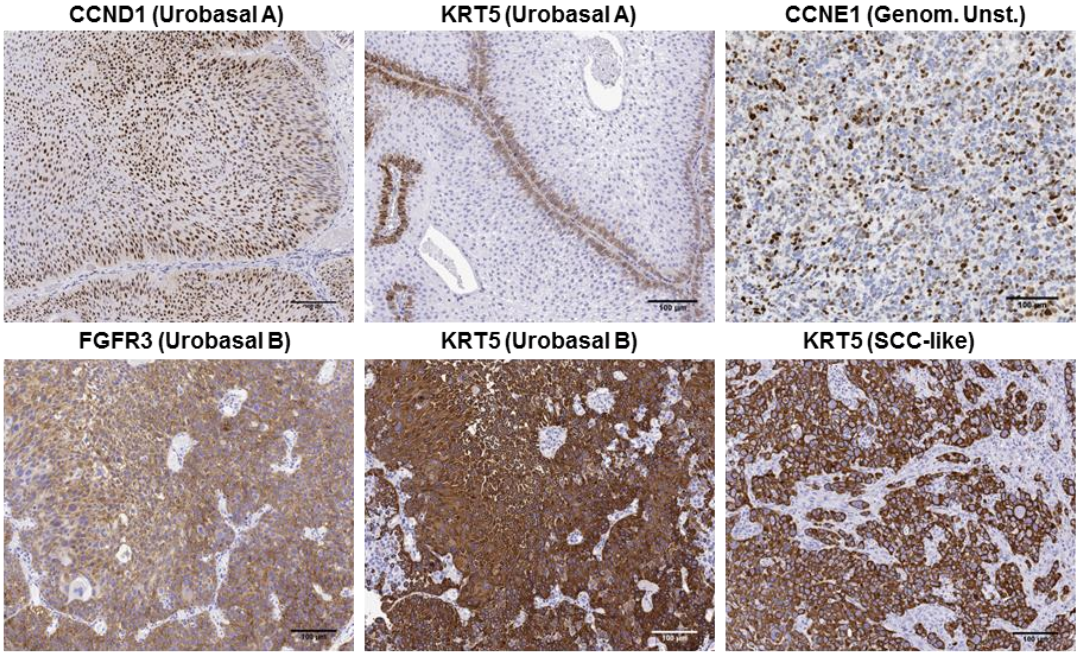
Sjödahl et al. Fig 2.



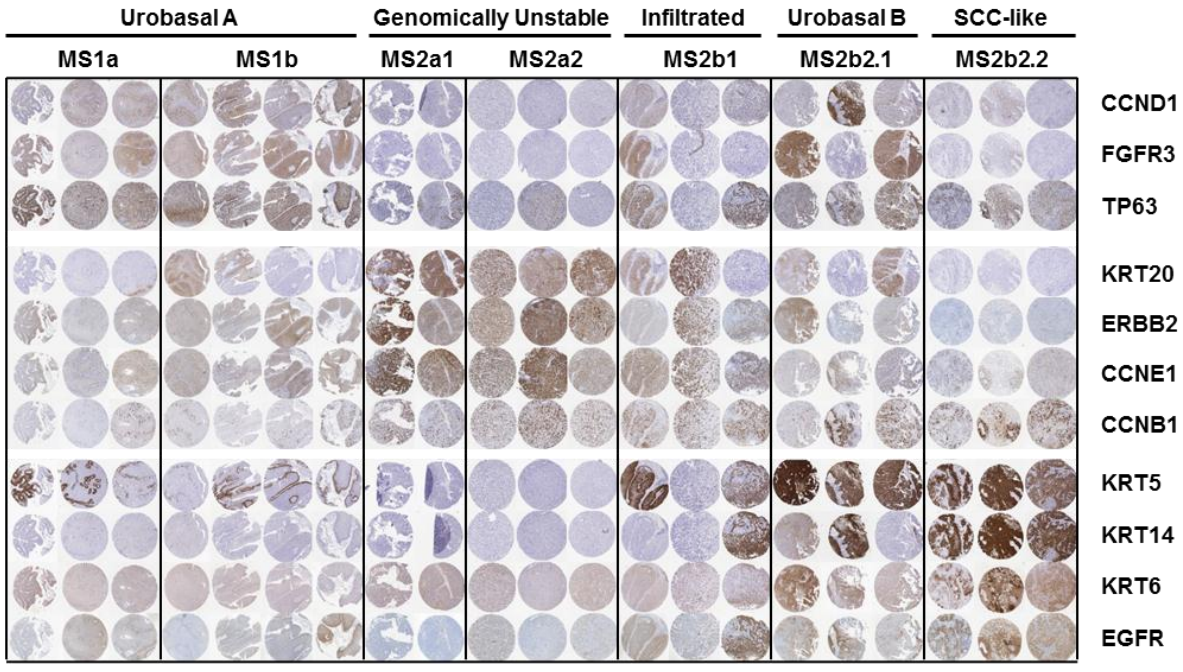
Sjödahl et al. Fig 3.



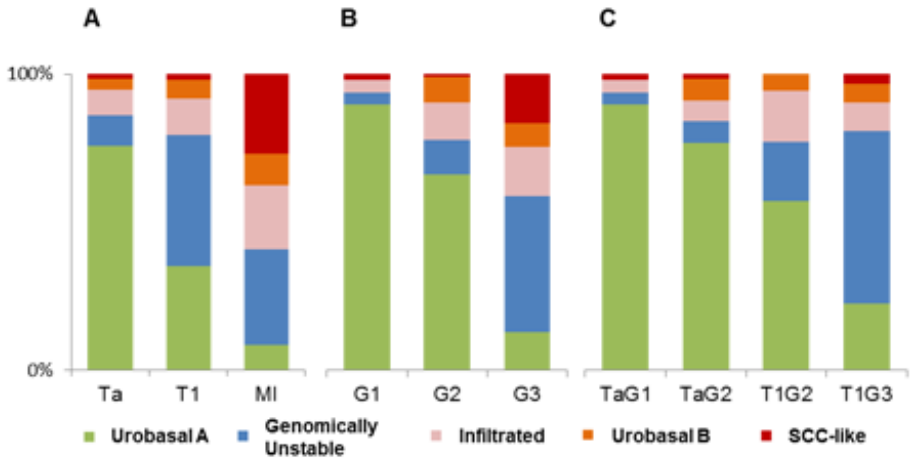
A


CCND1 (Urobasal A)
KRT5 (Urobasal A)
CCNE1 (Genom. Unst.)
FGFR3 (Urobasal B)
KRT5 (Urobasal B)
KRT5 (SCC-like)

B


Urobasal A
Genomically Unstable
Infiltrated
Urobasal B
SCC-like
MS1a
MS1b
MS2a1
MS2a2
MS2b1
MS2b2.1
MS2b2.2
CCND1
FGFR3
TP63
KRT20
ERBB2
CCNE1
CCNB1
KRT5
KRT14
KRT6
EGFR

Sjödahl et al. Fig 5.



Sjödahl et al. Fig 6.