On Molecular Classification of Bladder Cancer – Out of One, Many

Mattias Aine¹, Pontus Eriksson¹, Fredrik Liedberg², Mattias Höglund¹, Gottfrid Sjödahl²

¹ Division of Oncology and Pathology, Department of Clinical Sciences Lund, Lund University, Lund, Sweden.
² Division of Urological Research, Department of Translational Medicine, Lund University, Malmö, Sweden.

There has been a recent burst of interest in urothelial bladder cancer (UBC), much of it attributable to the landmark TCGA publication on UBC [1] as well as promising new avenues for therapeutic intervention [2-3]. Optimal utilization of traditional and novel treatment modalities likely require molecular phenotyping on the patient and tumor level. Working toward this goal, we and others have devised molecular stratification systems for UBC based on tumor mRNA expression patterns [4-6]. A recent comprehensive review has provided an excellent summary of the current state of knowledge of UBC biology [7]. Until recently however, the existing classification systems for UBC had not been simultaneously evaluated in independent data. We therefore set out to characterize the interrelations between the Lund (Urobasal A/B, Genomically Unstable, Infiltrated and SCC-like) [4], UNC (Basal and Luminal) [5] and MDA (Basal, Luminal and TP53-like) [6] classification systems in a cohort of 234 muscle invasive (MI) tumors from TCGA, including tumors classified in the landmark publication (TCGA I, II, III and IV) [1]. Our work expressly focuses on the biology underlying UBC molecular subtypes mainly by analyzing gene expression (GEX) data, but also specific mutation and copy-number changes [8]. This comment is aimed at briefly summarizing the findings of this work and to highlight recent advances in our understanding of the molecular determinants of UBC molecular subtypes [9-10] with potential implications for treatment stratification.

In Aine et al. [8] we performed unsupervised consensus clustering of the independent TCGA data to derive a two and three group clustering solution as these have been proposed as relevant stratification levels for MI UBC [5-6]. We applied the existing classifiers on this data set, and made use of gene signatures for UBC-relevant biological processes, as well as copy-number and mutation data for genes with established roles in UBC pathogenesis, e.g. RB1 mutations and chromosome 9 loss. Substantial heterogeneity remained after splitting the data into two or three groups, indicating the presence of more than three molecular subtypes of MI UBC. Further clustering resulted in six tumor clusters that could be meaningfully understood through the classifier results, biological signatures, and genomic events. Theoretically there is no definite limit to the number of subgroups that can be derived as ever larger data sets are produced. In this work we arrived at six clusters due to the size of the investigated data set (234 tumors) and our ability to reliably tie biology to the derived subgroups.

Another aim was to bring clarity to the interrelations of the UNC, MDA, TCGA, and Lund systems and highlight the molecular basis for classification (Figure 1A and B). Overall we observed a broad agreement between the classification systems for the subset of tumors that have been characterized as SCC-like/Basal/TCGA III, but also found a subset of tumors classified as
UNC/MDA Basal with differing classifications according to Lund and TCGA (TCGA IV/Infiltated). Biological themes related to this group included high levels of immune and stromal gene expression combined with absent expression of both urothelial and squamous differentiation genes. These observations are in agreement with another recent analysis using the original TCGA cohort [11] and lead us to tentatively label the group "Infiltated Mesenchymal". We also found that the Lund classification system splits Luminal/TCGA I tumors into the Urobasal A and Genomically Unstable classes respectively. These major subtypes of non-MI (NMI) disease that represent opposite poles on a low-grade to high-grade spectrum [12] showed phenotypic convergence in the uniformly high-grade MI setting where a different set of gene signatures drive subtype separation. A subgroup that corresponded to TCGA II tumors displayed expression patterns that did not conform to the dichotomous UNC paradigm but included most tumors classified as MDA TP53-like/Lund Infiltated. Unlike the "Infiltated Mesenchymal" subgroup these tumors expressed variable levels of urothelial or squamous differentiation markers in addition to immune signatures and were therefore labeled “Infiltated Epithelial”. Finally we characterize a proliferative subgroup negative for most “traditional” UBC markers but with frequent expression of genes involved in small-cell and neuroendocrine differentiation, suggesting the existence of a subgroup enriched for these histological variants.

Differential expression of immune- and stroma related genes in MI UBC may have important effects on prognosis and treatment [3], but also affect subgroup identification by hierarchical clustering. Increasing invasion depth (T-stage) and a SCC-like tumor cell phenotype are both associated with more pronounced immune/stromal signals. For some highly affected cases, the tumor-cell phenotype can be almost completely masked. It should thus be acknowledged that subgroups formed by highly infiltrated cases are composite- and not intrinsic subtypes. For the “Infiltated Epithelial" subgroup, the underlying epithelial character can still be observed, but it is not clear whether the “Infiltated Mesenchymal" identity is due to tumor cells having undergone extensive EMT or a dominance of non-tumor cell signals.

The different nomenclatures and classification methods probably add confusion to an already complex subject matter. Although the picture that emerges is complex, we believe it is rightfully so. Molecular subtypes of MI UBC do show a remarkable heterogeneity compared to other tumor types. That said, some recent results can facilitate our understanding of the subtypes. Specifically, we show that the different classification methods largely identify overlapping groups but at different levels [8]. In another recent study we set out to systematically define the transcription factor landscape underlying UBC molecular subtypes [9] and provide important links between molecular subtypes and transcription factors (e.g. PPAR-γ, RXR-α, FOXM1, FOXA1, GATA3 and STAT3, Figure 1B) that shape the tumors’ expression profiles [6,9-10]. Specifically, we found indications of a FOXM1/PLK1-axis separating the proliferative Genomically Unstable and SCC-like subtypes from the well differentiated Urobasal A subtype. The Genomically Unstable and SCC-like subtypes in turn are defined by pharmacologically actionable [6,10] transcription factor networks involving PPARG or EGFR/STAT3 activity respectively.
As many key determinants of UBC subtypes are de facto drug targets (Figure 1B), molecular subtypes will likely predict targeted treatment response. Subtype assessment should therefore be an integral part of future clinical and experimental trials.


**Figure legend**

Figure 1. (A) Schematic representation of subtype classification overlaps in Aine et al. For each classification system, the subtype distribution is proportional to classification results. Overlaps between classification systems have been matched to approximate classification results. Subtype colors were obtained from the respective original publications. (B) Subtype associated differentially expressed genes categorized by transcription factors (TFs), tumor markers, and actionable targets. UNC = University of North Carolina; MDA = M.D. Anderson Cancer Center; TCGA = The Cancer Genome Atlas; UroA = Urobasal A; Genom. Unst. = Genomically Unstable; Inf. Epi. = Infiltrated Epithelial; Inf. Mes. = Infiltrated Mesenchymal; UroB = Urobasal B; Sc/Ne = Small cell/Neuroendocrine; TFs = Transcription factors.

**Take home message**

Comparative analysis showed that bladder cancer classification systems identify overlapping subtypes but at different levels. Muscle-invasive bladder cancer shows remarkable heterogeneity and six subtypes were identified that differ in transcriptional networks, marker profiles, and expression of actionable targets.