



The Lund Molecular Taxonomy Applied to Non—Muscle-Invasive Urothelial Carcinoma



Nour-Al-Dain Marzouka,* Pontus Eriksson,* Carina Bernardo,* Carolyn D. Hurst,† Margaret A. Knowles,† Gottfrid Sjö Dahl,‡ Fredrik Liedberg,‡ and Mattias Höglund*

From the Division of Oncology,* Department of Clinical Sciences Lund, Lund University, Lund, Sweden; the Division of Molecular Medicine,† Leeds Institute of Medical Research at St James's, St James's University Hospital, Leeds, United Kingdom; and the Urology—Urothelial Cancer,‡ Department of Translational Medicine, Lund University, Skåne University Hospital, Malmö, Sweden

Accepted for publication
May 16, 2022.

Address correspondence to
Mattias Höglund, Ph.D.,
Department of Oncology, Clin-
ical Sciences, Lund University,
Medicon Village 404, Scheele-
vägen 2, 223 81 Lund, Sweden.
E-mail: mattias.hoglund@med.lu.se.

The precise classification of tumors into relevant molecular subtypes will facilitate both future research and optimal treatment. Here, the Lund Taxonomy system for molecular classification of urothelial carcinoma was applied to two large and independent cohorts of non—muscle-invasive tumors. Of 752 tumors classified, close to 100% were of the luminal subtypes, 95% urothelial-like (Uro; UroA, UroB, or UroC) and 5% genomically unstable. The obtained subtype structure organized the tumors into groups with specific and coherent gene mutation, genomic, and clinical profiles. The intrasubtype variability in the largest group of tumors, UroA, was caused by infiltration and proliferation, not considered as cancer cell type—defining properties. Within the UroA subtype, a *HOXB*/late cell-cycle gene expression polarity was found, strongly associated with *FGFR3*, *STAG2*, and *TP53* mutations, as well as with chromosome 9 losses. Kaplan-Meier analyses identified the genomically unstable subtype as a progression high-risk group, also valid in the subgroup of T1 tumors. Almost all progression events occurred within 12 months in this subtype. Also, a general progression gene signature was derived that identifies high- and low-risk tumors. All findings were demonstrated in two independent cohorts. The Lund Taxonomy system is applicable to both non—muscle- and muscle-invasive tumors and may be a useful biological framework for translational studies. (*J Mol Diagn* 2022, 24: 992–1008; <https://doi.org/10.1016/j.jmoldx.2022.05.006>)

Several studies have been directed toward the molecular subclassification of urothelial carcinomas.^{1–4} Most early studies focused on muscle-invasive and advanced tumors only. As several different classification systems were suggested, a collaboration was initiated with the aim to arrive at a common consensus classification system,⁵ published in 2020.⁶ Similar extensive data on gene expression, gene mutation, and genomic alterations for non—muscle-invasive bladder cancer (NMIBC) have however been limited. Thus, the publications of two large studies on NMIBC, the UROMOL⁷ and the Leeds⁸ cohorts, amounting up to 750 cases, many with associated gene mutations and genomic alterations, are important contributions to the field and provide an opportunity to deepen our molecular understanding of non—muscle-invasive urothelial tumors (stages Ta and T1). As a starting point, both studies applied genome-wide gene expression analyses, followed by various

clustering procedures, to identify four subsets of tumors, each as the basis for further analyses. Both groups also subdivide tumors according to fraction of genome altered (FGA) based on array comparative genomic hybridization or low-pass sequencing copy number data and arrive at three-⁷

Supported by the Swedish Cancer Society 2020/0709 and 2020/0710, the Swedish Research Council 2021-00859, the Bertha Kamprad Foundation, the Nilsson Cancer Foundation, the Hjälm Family Foundation for Medical Research U-2021-01, Lund Medical Faculty (ALF), the Cancer Research Fund at Malmö General Hospital, Skåne University Hospital Research Funds, Sten K. Johnson Foundation, the Royal Physiographic Society in Lund, Cancerfoundation Kronoberg 2019, Torsten Gester Research Foundation, Märta Winkler Foundation for Medical Research, Bergvall Foundation, and Skåne County Council's Research and Development Foundation.

N.M. and P.E. share first authorship.

Disclosures: None declared.

and four-tiered⁸ genomic groupings. The Lund Taxonomy (LundTax) was originally designed to classify both non-muscle- and muscle-invasive tumors.¹ Since its original publication, it has been extended to include five major subtypes, urothelial-like (Uro), genomically unstable (GU), basal/squamous-like (Ba/Sq), mesenchymal-like, and small cell/neuroendocrine-like. Uro is by far the largest group and demonstrates intrasubtype variability, motivating the further subdivision into UroA, UroB, and UroC.^{9–11} The three LundTax subtypes of a nonluminal type (Ba/Sq, mesenchymal-like, and small cell/neuroendocrine-like) make up around half of muscle-invasive cases but are rarely seen in non-muscle-invasive cases. The LundTax system differs from other classification systems as it aims to classify tumors based on the cancer cells only, achieved by a combination of gene expression clustering and immunohistochemistry.⁹ This approach revealed that some gene expression clusters contained samples of more than one cancer cell phenotype, a feature named convergence, and that two different clusters could be composed of identical cancer cell phenotypes, a feature named divergence.⁹ These processes were mainly driven by signals originating from noncancer cells or by strong proliferation signatures. To overcome these obstacles, the grouping of tumors obtained by gene expression clustering was adjusted with extensive immunohistochemical data⁹ to arrive at a subclassification of cancer cell phenotypes, independent of the presence of nontumor cells. An RNA expression-based classifier was then trained on the immunohistochemistry-adjusted subtypes.¹⁰ Consequently, the RNA classifier identifies subtypes defined by immunohistochemistry, naturally selecting differences in gene expression by the cancer cells and ignoring biopsy bulk features, such as immune and stromal cell infiltration.^{12,13} Herein, we apply the LundTax classifier to 535 cases from the UROMOL⁷ and 217 cases from the Leeds⁸ cohorts in parallel to test and cross-validate the potential value of the Lund Taxonomy to biologically stratify non-muscle-invasive urothelial carcinoma.

Materials and Methods

Data Sets and Classification

Transcriptomic and genomic data for samples of non-muscle-invasive bladder cancers were downloaded from the published UROMOL study⁷ as well as available clinical and follow-up information. The retrieved cohort contained uniformly remapped and requantified¹⁴ edgeR trimmed mean of M values (TMM) normalized RNA-sequencing (RNA-seq) expression data for 535 samples summarized at the gene level in log₂ counts per million format.⁷ A total of 438 samples were sequenced on the Illumina HiSeq 2000 platform using EpiCentre ScriptSeq (Illumina, San Diego, CA) library preparation,¹⁵ and an additional 97 tumors were sequenced on the Illumina NovaSeq6000 platform utilizing

KAPA RNA HyperPrep library preparation (Roche, Basel, Switzerland).⁷ For these cases, RNA-seq derived mutational load and gene mutations were available and used, as reported by Linskrog et al.⁷ For the genes *hTERT* ($n = 397$), *FGFR3* ($n = 421$), *RAS* ($n = 422$), and *PIK3CA* ($n = 429$), reported hot-spot DNA sequencing data were used. Mutational signatures were based on RNA-seq data and available for 437 samples. Copy number data were available for 303 samples.⁷ This cohort is referred to as the UROMOL cohort. Gene expression and copy number data for 217 non-muscle-invasive cancers were retrieved from Hurst et al.⁸ The microarray gene expression data were generated on the Affymetrix GeneChip Human Transcriptome Array 2.0 (Affymetrix, Santa Clara, CA) and processed through Affymetrix Expression Console Software, Affymetrix Power Tools, and the R2 Genomics Analysis and Visualization Platform and provided by the authors in normalized gene-level log₂ format.⁸ Clinical and follow-up data were available for 195 cases, and mutational data for the genes *FGFR3*, *PIK3CA*, *KDM6A*, *STAG2*, *RBI*, and *TP53* were obtained through DNA sequencing. This cohort is referred to as the Leeds cohort. Both cohorts were classified according to the LundTax system into UroA, UroB, UroC, GU, Ba/Sq, mesenchymal-like, or small cell/neuroendocrine-like using a previously established random forest-based single-sample predictor.^{12,13} For detailed information on cohort composition and criteria for clinical entities, see Supplemental Table S1 in Linskrog et al.⁷ and Supplemental Table S1 in Hurst et al.⁸ For heat map visualization, the log-transformed data were median centered.

Gene Signatures

Predefined signatures of early and late cell cycle genes, *FGFR3* co-expressed genes, ribosomal genes,¹ and urothelial differentiation genes⁹ were applied to the data. The genomic circuit score, established to distinguish Uro from GU cases, was calculated as $RB1 + FGFR3 + CCND1 - E2F3 - CDKN2A$ mRNA expression using log-transformed values.^{10,16} A HOX-bound gene signature¹⁷ and established cancer immune and stroma signatures, Immune141_UP and Stromal141_UP,¹⁸ were also applied. The late/early cell cycle ratio was calculated as follows: the cohort centered log-transformed median expression value of the late cell cycle signature genes minus the median expression value of the early cell cycle signature genes. To sort the samples according to median late cell cycle and anterior *HOXB* expression, the median of each signature was calculated for each sample using log-transformed data. Samples were then sorted on the basis of median *HOXB* expression when the *HOXB* median was larger than the late cell cycle median; otherwise, they were sorted by median late cell cycle expression.

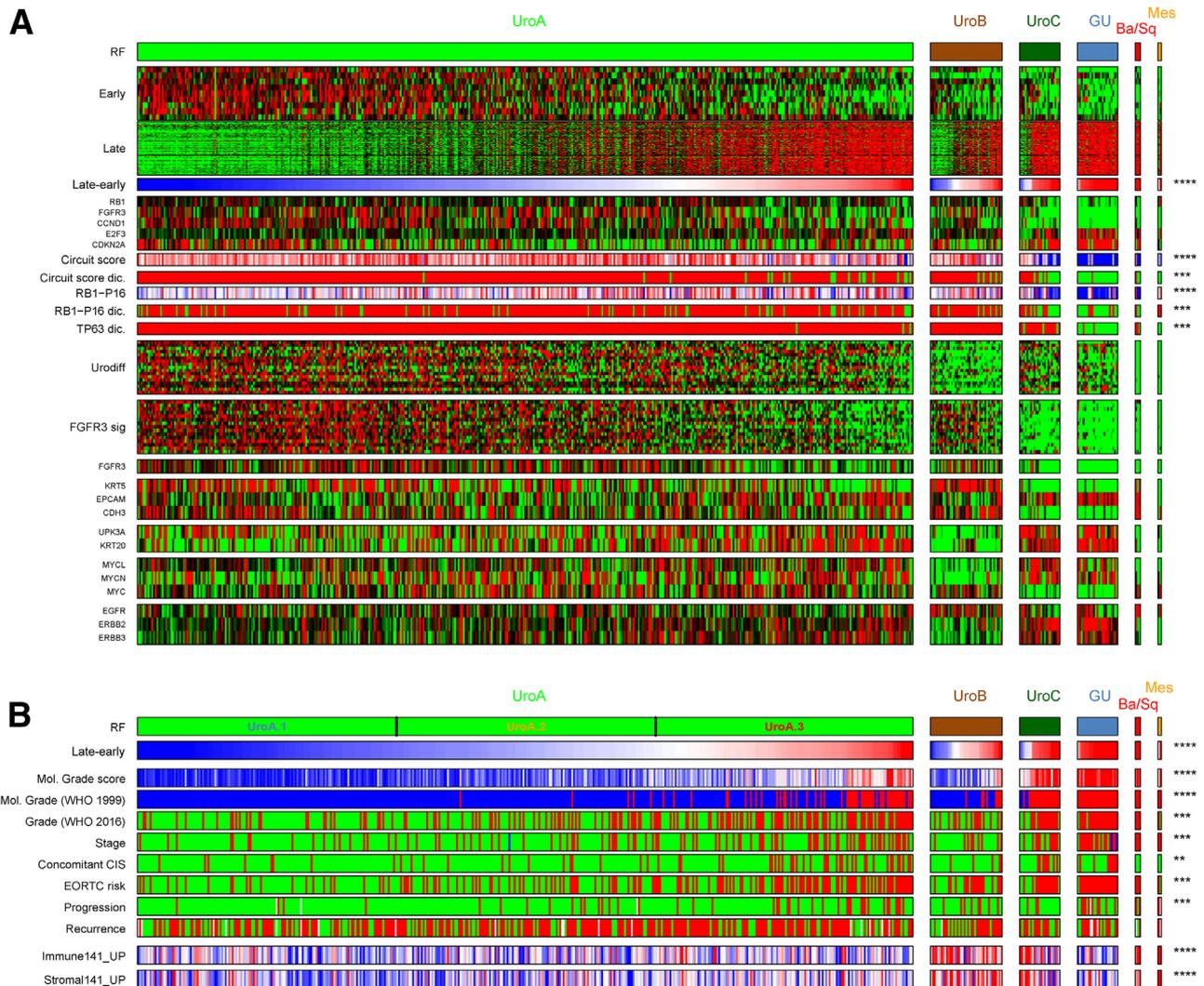


Figure 1 Molecular classification of the UROMOL cohort and clinical associations. **A:** Grouping of the cases according to the Lund Taxonomy classification and according to the ratio between the late and early cell cycle gene signature within each subtype. RF, molecular subtype according to the random forest algorithm. Early, the early cell cycle gene signature. Late, the late cell cycle gene signature. Late-early, the ratio between late/early gene signatures (blue, low ratio; red, high ratio). Circuit score genes: *RB1*, *FGFR3*, *CCND1*, *E2F3*, and *CDKN2A*. Circuit score: red, high score [urothelial-like (Uro)]; blue, low score [genomically unstable (GU)]. Circuit score dic., dichotomized circuit score. RB1-P16, ratio between *RB1/CDKN2A(p16)* expression [red, *RB1* high, *CDKN2A(p16)* low (Uro); blue, *RB1* low, *CDKN2A(p16)* high (GU)]. RB1-P16 dic., dichotomized *RB1/CDKN2A(p16)* expression ratio. TP63 dic, dichotomized TP63 expression (red, high; green, low). Urodif, urothelial differentiation signature. FGFR3 sig., FGFR3 associated gene signature. Individual genes as indicated (red, high relative expression; green, low relative expression). **B:** Clinical associations with the subtypes in the UROMOL cohort. RF, subtype according to the random forest algorithm with tentative division of UroA into three groups. Late-early, the ratio between late/early gene signatures (blue, low ratio; red, high ratio). Mol. Grade score, continuous score on the molecular-grade signature (blue, low score; red, high score); MolGrade [World Health Organization (WHO) 1999], dichotomized cohort/platform-adapted threshold for molecular-grade (blue, WHO 1999 grade 1 or 2; red, high WHO 1999 grade 3). Grade (WHO 2016): green, low grade; red, high grade. Stage, green, pathologic stage Ta; red pathologic stage T1. Concomitant carcinoma *in situ* (CIS): green, absent; red, present. European Organisation for Research and Treatment of Cancer (EORTC) risk: green, low EORTC risk score; red, high EORTC risk score. Progression: green, no progression; red, progression; white, no data. Recurrence: green, no recurrence; red, recurrence; white, no data. Immune141_UP, continuous score on the immune signature: blue, low score; red, high, score. Stromal141_UP, continuous score on the stromal signature: blue, low score; red, high score. $**P < 0.01$, $***P < 0.001$, and $****P < 0.0001$ (analysis of variance). Ba/Sq, basal/squamous-like; Mes, mesenchymal-like.

Molecular-Grade Signature

A rule-based single-sample molecular grade classifier was built based on the World Health Organization (WHO) 1999 grading system (G1, G2, and G3) and identifies G3 cases.¹⁹ For the training data, 314 Uro samples, the only subtype that includes G1, G2, as well as G3 cases, were retrieved from

two previous studies.^{1,9} Cases were divided as G3 versus G1 or G2. Genes were ranked within samples, and areas under the curve (AUCs) were calculated for each individual gene. Genes with an AUC >0.7 were selected and paired to all genes in the data sets separately to make rules. The selected genes were allowed to be part of rules more than one time. Scores were calculated for each possible rule,²⁰ and rules

with scores >0.45 in both data sets were selected. The final rule scores were calculated as the average of their score in the training data. The classifier calculates a weighted sample score based on the number of the TRUE rules and their scores in the training data set as follows: sample score = the sum of TRUE rule scores/sum of all scores in the classifier. The molecular grade for the sample is said to be G3 if the sample score is >0.5 , and G1/G2 otherwise. As thresholds for G3 versus G1/G2 are platform dependent, adjusted thresholds were established for each data set.

Establishing Thresholds

The R package *DistributionOptimization*²¹ was used to fit gaussian mixture models of data variables. Number of modes in *DistributionOptimization* function was set to two unless otherwise indicated. To dichotomize data, the R package *AdaptGauss*²² was used to determine the cutoff between the gaussian models.

Fraction of Genome Altered and Mutational Load

The reported fraction of genome altered (FGA) values for each case were used for the UROMOL cohort.⁷ The FGA values for cases in the Leeds cohort were calculated as the summed length of autosomes showing imbalances divided by the length of the autosomes. Mutational load for the UROMOL cases was obtained from Lindskrog et al.⁷

Consensus Clustering of UroA Samples

BioMart was used to obtain up-to-date gene symbols, biotype, and chromosome location.²³ Protein coding genes on chromosomes 1:22, XY, and MT, and with detected expression in $>40\%$ of samples, were kept and recentered. *ConsensusClusterPlus*²⁴ was applied to the top 5000 varying genes within the UroA subtype using $k = 2$ to 5. Analysis of variance (ANOVA) significant genes for the given solution were plotted in heat map format with gene clusters obtained by manual examination of hierarchical clustering cutoffs. Gene Ontology enrichment of gene clusters was examined using the PANTHER Gene Ontology enrichment tool.²⁵

Statistical Tests

All tests were performed in R version 4.1.1 software (<https://www.r-project.org>). Fisher exact tests were used to find the significance level in binary variables, and ANOVA was used for continuous variables. Progression-free survival in the UROMOL and Leeds cohorts was visualized by Kaplan-Meier curves with comparisons using the log-rank method. Survival data and outcome definitions were as defined in the original studies. Ratios on the progression signature were dichotomized into high- and low-risk categories based on the same gene signatures but with

ratio thresholds selected in each study separately. For clarity, survival curves were truncated at 5 years, after which only three progression events occurred.

Receiver Operating Characteristic Curve Analysis

The R package *caTools* and *ROCit* was used to identify genes associated with progression in normalized and centered data within the UROMOL cohort. Genes were marked as either up-regulated or down-regulated within the progression group. The AUC value of the mean expression was then examined for the top 100 up-regulated and down-regulated genes, as well as the ratio between the mean up and down expression. The AUC value was then examined in the independent Leeds cohort using the average expression value for the same identified genes and their ratio. To examine if the same information could be recapitulated from non-cohort-centered expression values (ie, independently for each sample), the average internal rank of the identified up-regulated and down-regulated genes for each sample was used to calculate the ratio (average rank of up genes divided by average rank of down genes). The ratio of average ranks for the same genes was then applied to the Leeds cohort.

Results

The rationale of this investigation is to apply LundTax for urothelial carcinomas to two large and independent cohorts of non-muscle-invasive tumors, the UROMOL and the Leeds cohorts. These cohorts are analyzed in parallel to be able to cross-validate our findings, and hence, the respective data sets will be treated separately. Furthermore, as the classification, and consequently the grouping of the cases, is based in gene expression data only, the independent mutation and genomic data are used to further validate the classification outcomes.

The UROMOL Cohort

Tumors in the UROMOL cohort ($n = 535$) were classified according to the LundTax system using a single-sample classifier resulting in 507 Uro cases, sub-stratified into 443 UroA, 41 UroB, and 23 UroC, and a group of 23 GU tumors. In addition, three samples were classified as Ba/Sq and two as mesenchymal-like. Cases were then organized within each subtype according to an expression ratio between an early and a late cell cycle gene signature (Figure 1A). Established gene signatures and individual genes, known to distinguish the Lund Taxonomic groups, were then used to verify the classification in a heat map (Figure 1A). The genomic circuit score clearly identified tumors classified as GU, both by a continuous and a dichotomized score (Supplemental Figure S1). The continuous and dichotomized RB1/CDKN2A expression ratio also indicated that GU samples were classified correctly (Figure 1A and Supplemental Figure S2). Similarly, the

expression of TP63, dichotomized into expressed and not expressed (Supplemental Figure S3), firmly established the nature of GU samples.¹¹ The urothelial differentiation signature was expressed by most of the UroA tumors and showed the lowest expression in UroB. The FGFR3 associated gene signature, as well as the *FGFR3* gene itself, was expressed in UroA and UroB but was almost absent in UroC and GU (Figure 1A). *KRT5* and *CDH3* (P-cadherin), both expressed by normal urothelial basal cells, showed mixed expression in UroA, high expression in UroB, but low expression in UroC and GU (Figure 1A). In accordance with previous results,^{9–11} an almost complete absence of *UPK3A* and *KRT20* expression in UroB tumors was found. UroB expressed the transcription factor *MYC*, but not *MYCL* nor *MYCN*, and UroC and GU the opposite, also in accordance with previous findings.¹⁰ Furthermore, *EGFR* expression was characteristic for UroB tumors and *ERBB2* expression for UroC and GU, validating previous subtype-specific findings.^{11,26} Thus, the classification of the UROMOL data into the Lund Taxonomy conforms well with previously described discriminative gene signatures.

To detect possible intrasubtype variability, the large UroA subtype was tentatively divided in three groups of equal sizes, UroA.1, UroA.2, and UroA.3, based on the proliferation ratio (Figure 1B). Molecular grade was associated with subtype and with the cell cycle axis within subtypes (Figure 1B); Supplemental Figure S4 shows information for establishing a cohort-adapted threshold. High molecular-grade tumors were predominantly seen in UroA.3, UroB, UroC, and GU. The grouping of the tumors was also aligned with the originally reported pathologic grades according to WHO 2016 (low grade and high grade). UroC and GU were of high grade using any grading system (70% and >90%, respectively). UroC and GU were also dominated by tumors of high European Organisation for Research and Treatment of Cancer (EORTC) risk scores (65% and 96%, respectively). More important, tumor progression predominantly occurred in GU (35%), followed by UroB (20%) patients. A similar association was also seen for pathologic stage, of which the majority (70%) of GU tumors were stage T1, as well as for concomitant carcinoma *in situ* (CIS), seen in 34% of both UroC and GU. The single case of pathologic CIS-only was classified as GU. Recurrence-free survival and recurrence rates were not associated with any subtype. Immune and stromal cell infiltration was predominantly seen in the UroB and UroC groups, whereas infiltration was randomly distributed within UroA. Hence, stratifying the UROMOL cohort according to the Lund Taxonomy identifies groups with distinct clinical and pathologic features.

Gene mutations in the UROMOL cohort were then analyzed. Rank and distribution plots of mutational load indicated a bimodal-like distribution that motivated dichotomizing of the data into high and low mutational load (Supplemental Figure S5). High mutational load cases were predominantly seen in UroC (22%) and GU cases (26%) but

also observed among UroA.3 (Figure 2). Key gene mutations, chromosomal aberrations, and mutational signatures were then analyzed. As expected, *hTERT* mutations were frequent (71% to 86%) and did not differ significantly among subtypes. *FGFR3* mutations were significantly higher in UroA (68%) and UroB (77%) and absent in GU tumors. *FGFR3* mutations were more frequent in UroA.1 and UroA.2 compared with UroA.3. *RAS* mutations were only detected in UroA and UroB cases, the same subtypes that showed substantial numbers of *FGFR3* mutations. *PIK3CA* mutations were seen in all subtypes. *KDM6A* mutations were particularly frequent in UroA (20%). *STAG2* mutations, associated with low-grade papillary tumors, were seen in UroA (11%), UroB (12%), and GU (4%), but not detected in UroC. However, *STAG1*, a paralog to *STAG2* with similar functions, displayed an almost inverted frequency distribution compared with *STAG2*. *PPARG* mutations were only seen in UroA and at low frequencies (1% to 6%). *RBI* mutations were almost absent in all subtypes, except for GU (17%). *TP53* mutations were predominantly seen in UroC (26%) and GU (35%) and differed significantly from the remaining subtypes. Mutation signature 5 and the APOBEC signature were almost equally frequent in UroA, UroB, and UroC, but differed significantly in the GU subtype, in which the APOBEC signature was seen in 89% and mutation signature 5 in only 11% of the cases. Taken together, the subtypes show distinct mutational profiles with respect to mutational load, specific gene mutations, and mutation signatures.

The analysis of genomic imbalances indicated that *TP53* losses (17p13) showed an increasing frequency with increasing cell cycle activity within UroA. *TP53* losses were particularly frequent in UroB and GU tumors (22% and 33%, respectively), but absent in UroC. UroC, on the other hand, showed a substantial fraction (40%) of cases with *MDM2* (12q15) gains/amplifications. *CDKN2A* (9p21) homozygous deletions were absent in UroC and GU but frequent in UroB (17%).^{16,27} As expected, *RBI* (13q14) losses were particularly frequent in GU cases (58%), because this subtype is characterized by frequent *RBI* mutations and lower expression levels. Gains/amplifications of *E2F3* (6p22) were particularly frequent in GU. Chromosome arm and 6p amplifications have previously been shown to be associated with UroC and GU cases but are rarely seen in UroA and UroB.¹⁰ Conversely, gains and amplifications of *CCND1* (11q13) were almost only detected in UroA and UroB and absent in UroC and GU. Hence, the LundTax subtypes in NMIBC show distinct profiles with respect to specific genomic aberrations.

The global copy number data (Figure 3) revealed that the UroC and GU subtypes, as well as the UroA.3 subclass, showed large fractions of altered genome. Almost all UroC and GU tumors were triploid. In UroA, triploid cases were enriched in UroA.3 cases. UroA.1 was characterized by few chromosomal imbalances and, particularly, frequent absence of chromosome 9 deletions, otherwise frequent in the

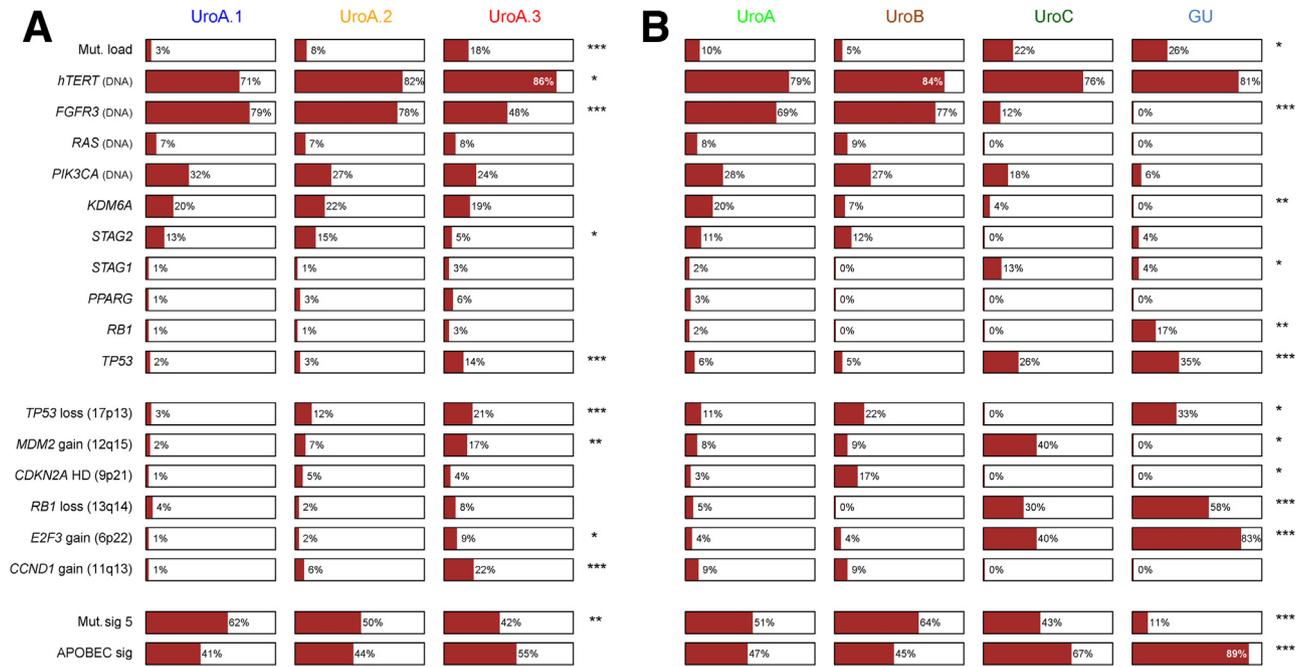


Figure 2 Distribution of mutational (Mut.) load, and selected mutations, genomic alterations, and mutation profiles in the UROMOL cohort. **A:** Distributions within the urothelial-like (Uro) A subtype tentatively grouped into UroA.1, UroA.2, and UroA.3. Brown, fraction of mutated, altered, cases. Mutational load dichotomized mutational load (brown, high mutational load; white, low mutational load). **B:** Distributions within each subtype. Brown, fraction of mutated cases. Mutational load, dichotomized mutational load (brown, high mutational load; white, low mutational load). Data: mutational load is based on RNA sequencing (RNA-seq). Mutations in *hTERT*, *FGFR3*, *RAS*, and *PIK3CA* are based on hot-spot DNA sequencing. Gene mutations for *KDM6A*, *STAG2*, *STAG1*, *PPARG*, *RB1*, and *TP53* are based on RNA-seq. Mutational signatures are based on RNA-seq. Fractions (percentages) were calculated on the basis of the total number tested for each variable. * $P < 0.05$, ** $P < 0.01$, and *** $P < 0.001$ (analysis of variance). GU, genomically unstable; HD, homozygous deletion; sig, signature.

cohort. The UROMOL genomic classes conformed well with the present order of cases. Thus, LundTax molecular subtypes also show a strong association with genomic complexity.

The Leeds Cohort

The 217 Leeds cases were classified according to the LundTax system, and 183 UroA, 6 UroB, 15 UroC, 12 GU, and 1 single case of the Ba/Sq subtype were identified (Figure 4A). As for the UROMOL data, tumors were organized according to the late/early cell cycle ratio within each subtype. The circuit score, and the circuit score genes, as well as the *RBI/CDKN2A* expression ratio, and *TP63* expression all clearly confirmed the GU subclass of tumors within the cohort. All the dichotomized thresholds were adapted to the cohort (Supplemental Figures S6–S8). The urothelial differentiation signature varied within UroA, according to the cell cycle ratio, and showed the lowest expression in the UroB subtype. The *FGFR3* gene signature, including *FGFR3* itself, showed decreased expression in UroC and GU. *KRT5* expression was variable in UroA, high in UroB, and almost absent in UroC and GU. The expression of the urothelial basal cell marker *CDH3* was almost absent in UroC and GU. Similar to the UROMOL data, UroB expressed *MYC*, whereas *MYCN* was highly

expressed in UroC and was higher than *MYC* in GU. Subtype-associated expression of *EGFR* was not apparent in the Leeds data.

As for the UROMOL cohort, UroA cases were tentatively grouped according to the late/early ratio in three equally sized groups, UroA.1, UroA.2, and UroA.3 (Figure 4B). The molecular-grade signature was used to identify high-grade tumors with a platform-adjusted threshold (Supplemental Figure S9). High-grade UroA tumors were limited to the UroA.3 category, whereas UroB, UroC, and GU all were of high grade. The molecular grading conformed well also with the original WHO 2016 grading, albeit less stringent than the WHO 1999 grading system. All UroB, UroC, and GU were of pathologic stage T1; and among the UroA, the majority of UroA.3 were of stage T1. Progression events were rare in UroA and common in UroC. Recurrences were evenly distributed among the UroA cases but were slightly more frequent within the UroC subtype compared with the other subtypes. As in the UROMOL cohort, infiltration by immune and stromal cells mainly occurred in UroB and UroC tumors. Hence, subgrouping the Leeds cohort according to the Lund Taxonomy identifies the same associations with clinical and pathologic features as for the UROMOL cohort.

FGFR3 mutations were particularly prevalent in UroA and UroB cases but almost absent in UroC and GU

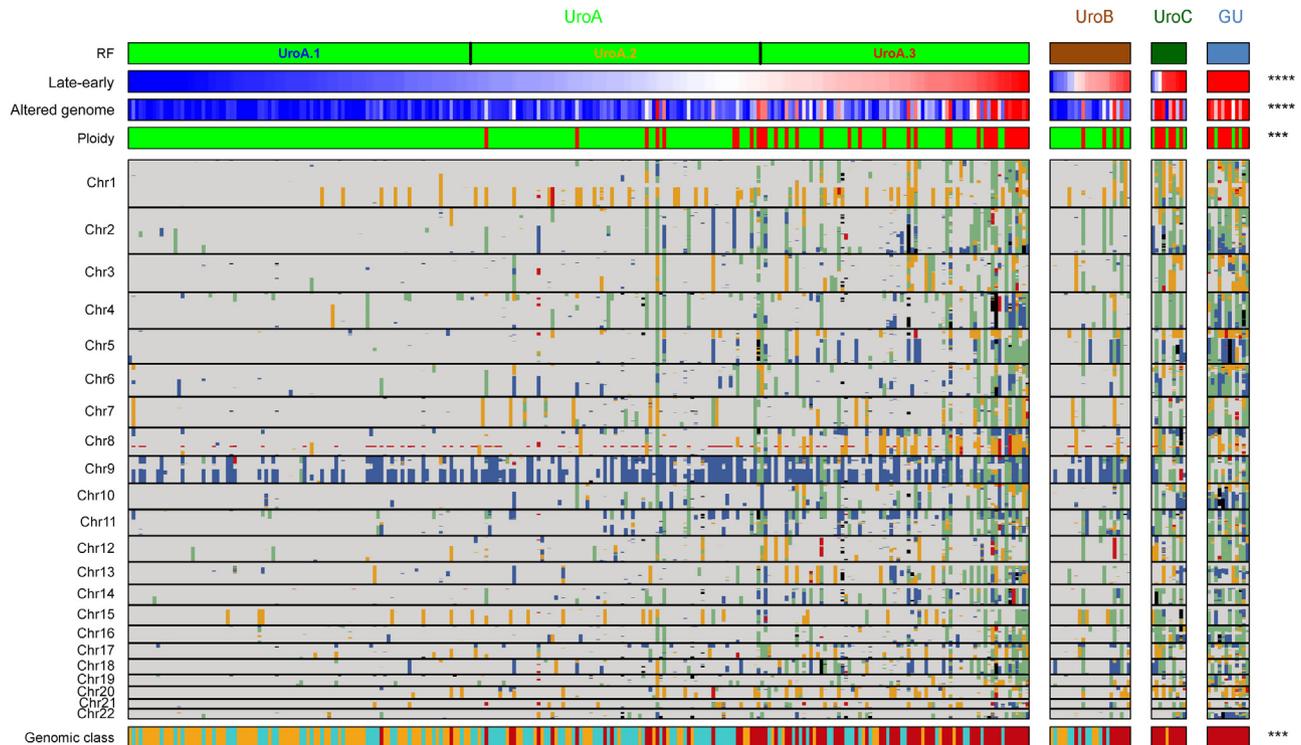


Figure 3 Genomic imbalance profiles in the UROMOL cohort, organized according to the Lund Taxonomy classifications system. RF, subtype according to the random forest algorithm with tentative division of urothelial-like (Uro) A into three groups as before. Late-early, the ratio between late/early gene signatures (blue, low ratio; red, high ratio). Altered genome, continuous variable of fraction of genome altered (blue, low fraction; red, high fraction). Ploidy: green, diploid; red, polyploid. Genomic imbalances: blue, loss; black, high balanced loss; orange, gain; red, high balanced gain; green, allelic imbalance without copy number alteration. Genomic class: genomic classes according to Lindskrog et al⁷ (2021): orange, class 1; blue, class 2; red class 3. **** $P < 0.001$, **** $P < 0.0001$ (analysis of variance). Chr, chromosome; GU, genomically unstable.

(Figure 5). Within UroA, UroA.1 (70%) and UroA.3 (37%) differed significantly in *FGFR3* mutation frequencies. *PIK3CA* mutations were seen in all subtypes, whereas *KDM6A* mutations were exclusively detected in UroA tumors with a strong enrichment in UroA.1 (46%). *STAG2* mutations were only observed in the Uro category of tumors. On the other hand, *RBI* and *TP53* mutations were significantly enriched in the UroC and GU tumors. Genomic loss of *TP53* (17p13) essentially followed the *TP53* mutation frequencies, being particularly frequent in UroA.3 and UroC tumors. *MDM2* (12q13) amplifications largely followed the pattern of *TP53* (17p13) losses but at lower frequencies. Both *RBI* (13q14) losses and *E2F3* (6p22) gains were strongly enriched in UroC and GU cases, whereas homozygous loss of *CDKN2A* (9p21) was almost absent in these subtypes but frequent in UroA.3 and UroB (16% and 17%, respectively). UroC differed radically from the remaining subtypes by showing a high frequency (40%) of *CCND1* (11q13) amplifications. Taken together, the classification of the Leeds cohort according to the LundTax system produces subtypes with distinct mutational and genomic profiles.

The copy number data were then plotted (Figure 6), and the FGA was calculated for each case. Data on ploidy were not available for this cohort. In general, FGA increased with increasing ratios on the late/early cell cycle axis. As for the

UROMOL data, UroC and GU showed large FGA, and cases not showing chromosome 9 losses were enriched in the UroA.1 group within the large UroA subtype. The original genomic classes defined by Hurst et al⁸ conformed well with the present organization of the data. Thus, a clear overall association between subtype, cell cycle activity, and fraction of genome altered is observed.

To summarize, the classification of the UROMOL and the Leeds cohorts shows highly similar results with respect to subtype classification, expression of class-defining signatures, key gene signatures, as well as gene mutations and genomic data. All the LundTax class-defining features previously described fit with the present classifications. More important, although classification was based on gene expression only, it still organized gene mutations and specific genomic alterations in biologically coherent groups. Hence, the LundTax subtypes in NMIBC show distinct genomic profiles that mirror differences already observed in muscle-invasive bladder cancer (MIBC).^{9,10}

UroA Tumors

As the Lund classification algorithm produces a dominant group of UroA tumors, the extent to which a further subdivision of this subtype is motivated was investigated. To answer this in an unsupervised manner, consensus clustering

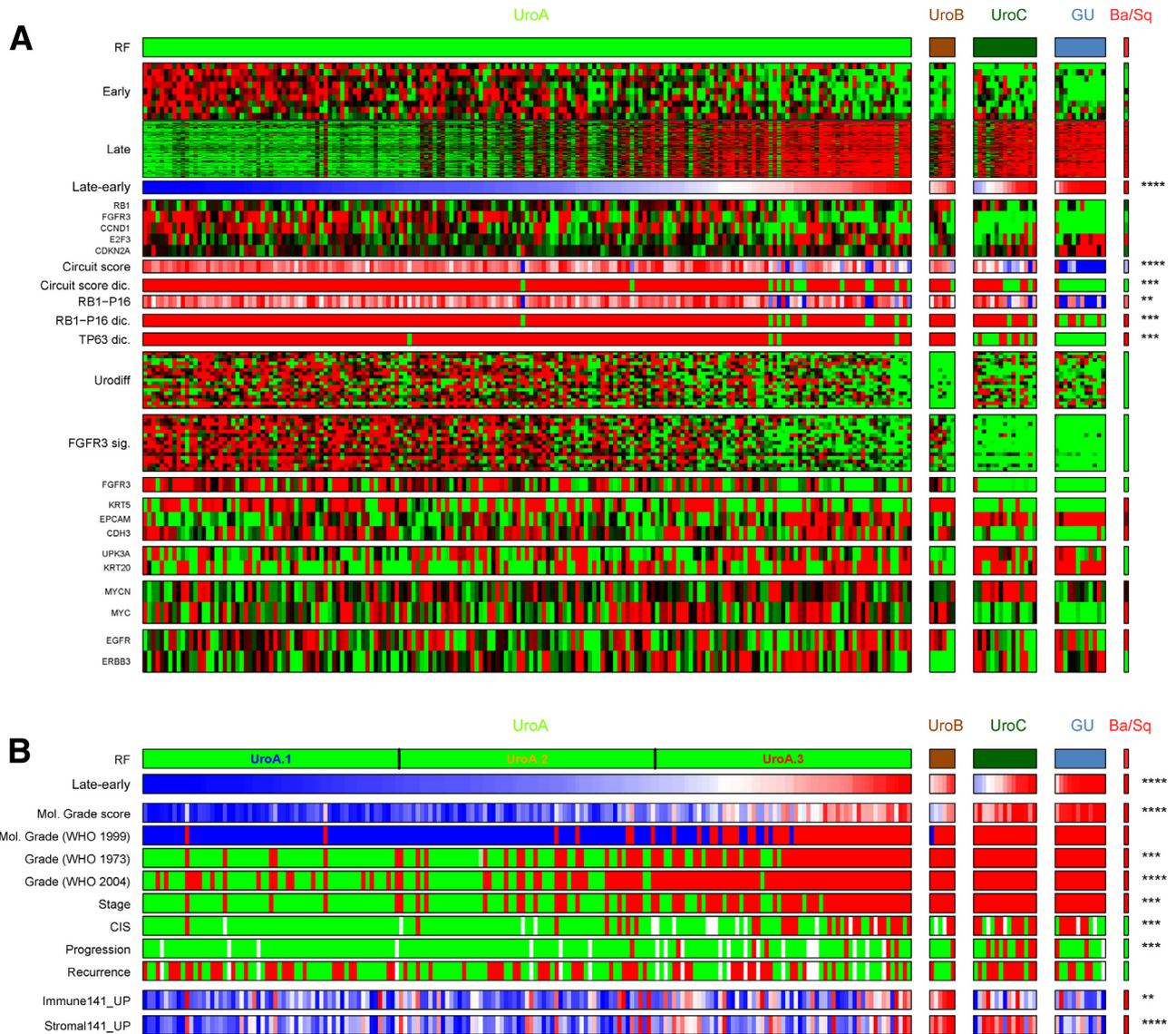


Figure 4 Molecular classification of the Leeds cohort and clinical associations. **A:** Grouping of the cases according to the Lund Taxonomy classification and according to the ratio between the late/early cell cycle gene signature within each subtype. RF, subtype according to the random forest algorithm. Early, the early cell cycle gene signature. Late, the late cell cycle gene signature. Late-early, the ratio between late/early gene signatures (blue, low ratio; red, high ratio). Circuit score genes: *RB1*, *FGFR3*, *CCND1*, *E2F3*, and *CDKN2A*. Circuit score: score of circuit genes: red, high score [urothelial-like (Uro)]; blue, low score [genomically unstable (GU)]. Circuit score dic., dichotomized circuit score [red, high (Uro); green, low (GU)]. RB1-P16, ratio between RB1/CDKN2A(p16) expression [red, RB1 high, CDKN2A(p16) low; blue, RB1 low, CDKN2A(p16) high]. RB1-P16 dic., dichotomized RB1/CDKN2A(p16) expression score [red, high (Uro), green, low (GU)]. TP63 dic., dichotomized TP63 expression [red, high (Uro); green, low (GU)]. Urodiff, urothelial differentiation signature. FGFR3 sig., the FGFR3 associated gene signature. Individual genes as indicated (red, high relative expression; green, low relative expression). **B:** Clinical associations with the subtypes in the Leeds cohort. RF, subtype according to the random forest algorithm with tentative division of UroA into three groups. Late-early, the ratio between late/early gene signatures (blue, low ratio; red, high ratio). Mol. Grade score, continuous score on the molecular-grade signature (blue, low score; red, high score). Mol. Grade [World Health Organization (WHO) 1999], dichotomized cohort/platform-adapted threshold (blue, WHO 1999 grade 1 or 2; red, WHO 1999 grade 3). Grade (WHO 1973): green, grade 1 or 2; red, grade 3. Grade (WHO 2016): green, low grade; red, high grade. Stage: pathologic stage (green, T_a; red, T₁). Concomitant carcinoma *in situ* (CIS): green, absent; red, present; white, no data. Progression: green, no progression; red, progression; white, no data. Recurrence: green, no recurrence; red, recurrence; white, no data. Immune141_UP, continuous score on the immune signature (blue, low score; red, high score). Stromal141_UP, continuous score on the stromal signature (blue, low score; red, high score). ***P* < 0.01, ****P* < 0.001, and *****P* < 0.0001 (analysis of variance). Ba/Sq, basal/squamous-like.

of all UroA cases in the respective cohorts was applied. We used $k = 2$ to 5 solutions, but none of the coclustering matrices, silhouette scores, or the cumulative distribution function scores indicated an optimal K-solution (Supplemental Figures S10 and S11). Furthermore, the

clustering was sensitive to variance filter levels and number of included genes. Taken together, this indicates that any remaining transcriptional variation within UroA is not distinct enough to motivate further subdivisions. However, to investigate the cause for the observed intrasubtype

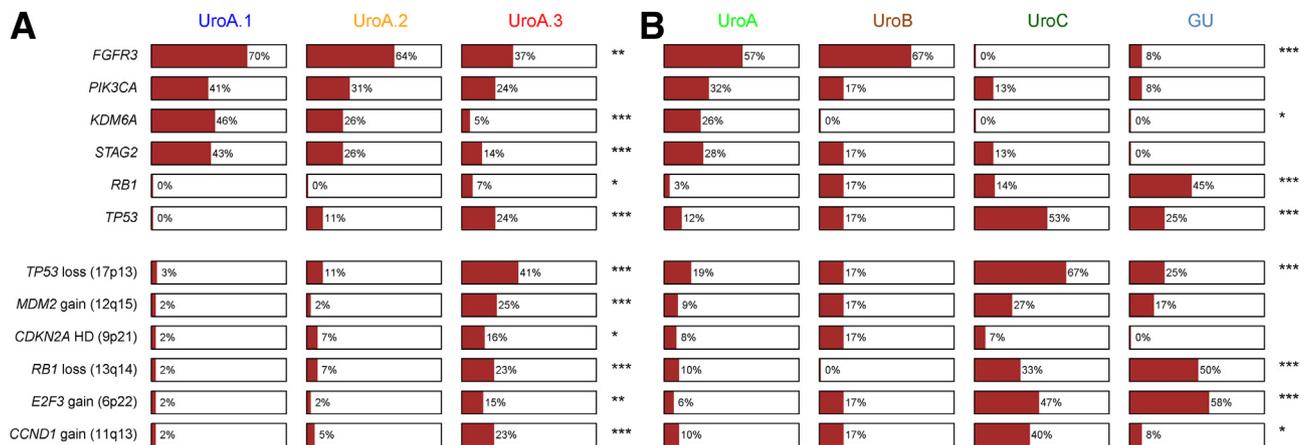


Figure 5 Distribution of selected mutations and genomic alterations in the Leeds cohort. **A:** Distributions within the urothelial-like (Uro) A subtype tentatively grouped into UroA.1, UroA.2, and UroA.3. **B:** Distributions within in each subtype. Brown, fraction of mutated cases. * $P < 0.05$, ** $P < 0.01$, and *** $P < 0.001$ (analysis of variance). GU, genomically unstable; HD, homozygous deletion.

transcriptional variation, a high k -value, $k = 5$, for the UROMOL data and a lower value, $k = 3$, for the Leeds data was tentatively selected in accordance with differences in cohort sizes (UROMOL, $N = 443$; and Leeds, $N = 183$), and then applied ANOVA to identify genes causing the variation. In the UROMOL data, a total of seven ANOVA signatures, S1 through S7, were identified (Figure 7A). A

systematic functional analysis using Gene Ontology term enrichment for biological processes revealed that S1 was enriched for cell cycle ($P < 3 \times 10^{-44}$) and related processes. Signature S2 did not show any strong associations with Gene Ontology terms. Signature S3 showed moderate associations with hormone metabolic process ($P < 3 \times 10^{-7}$) and blood vessel development

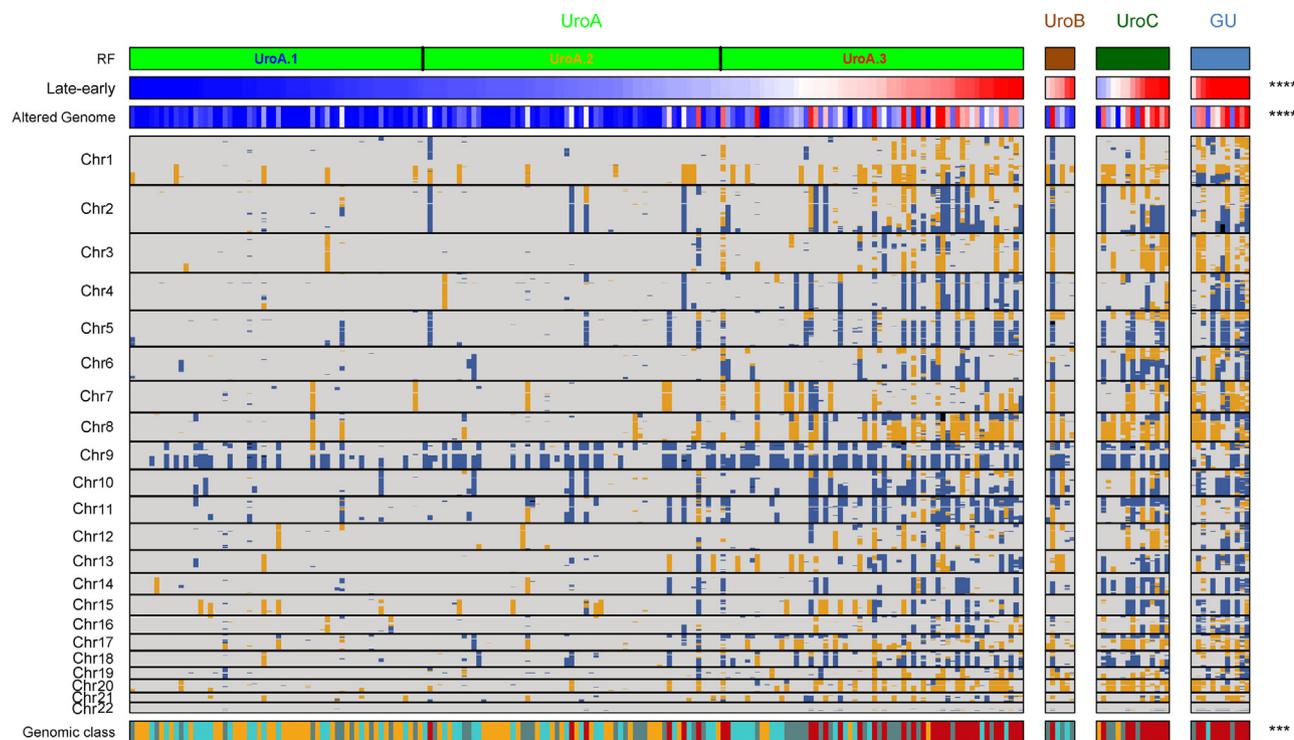


Figure 6 Genomic imbalance profiles in the Leeds cohort, organized according to the Lund Taxonomy classifications system. RF, subtype according to the random forest algorithm with tentative division of urothelial-like (Uro) A into three groups as before. Late-early, the ratio between late/early gene signatures (blue, low ratio; red, high ratio). Altered genome, continuous variable of fraction of genome altered (blue, low fraction; red, high fraction). Genomic imbalances: blue, loss; black, high balanced loss; orange, gain; red, high balanced gain; green, allelic imbalance without copy number alteration. Genomic class: genomic classes according to Hurst et al⁸ (2021): orange, class 1; blue, class 2; green, class 3; and red, class 4. *** $P < 0.001$, **** $P < 0.0001$ (analysis of variance). Chr, chromosome; GU, genomically unstable.

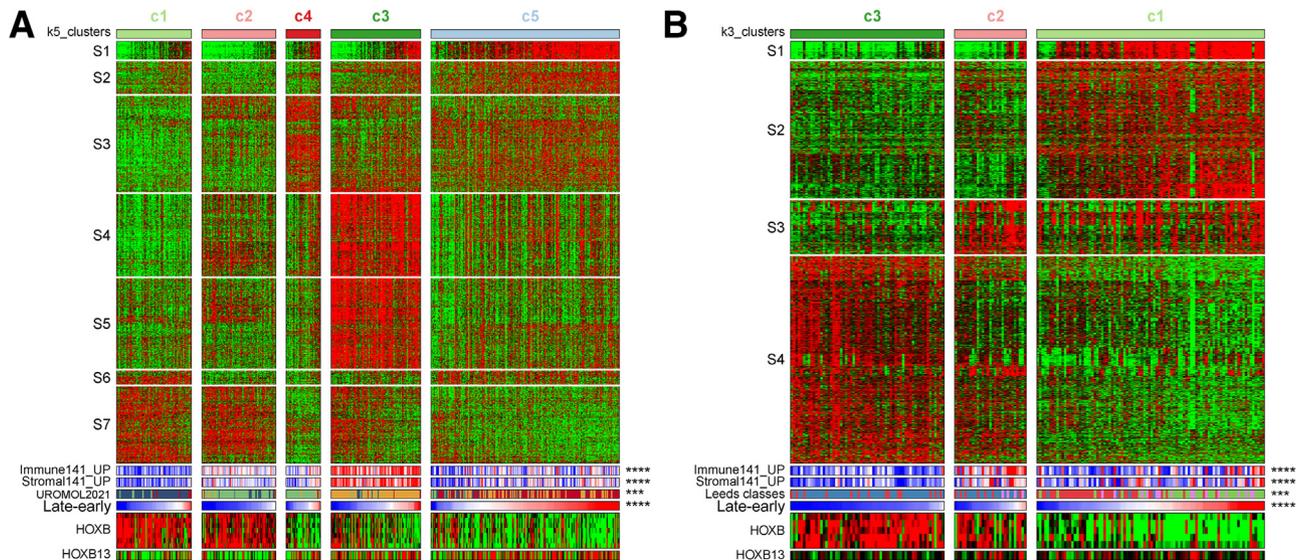


Figure 7 Consensus clustering of urothelial-like (Uro) A cases. **A:** Consensus clustering of the UROMOL UroA cases. Clusters, indicated as c1 through c5. Signatures S1 through S7 obtained by selection of analysis of variance significant genes and reordering by hierarchical clustering. Immune141_UP, continuous score on the immune signature (blue, low score; red, high score). Stromal141_UP, continuous score on the stromal signature (blue, low score; red, high score). UROMOL class, molecular class according to Lindeskrog et al⁷ (2021): green, class 1; red, class 2a; orange, class 2b; blue, class 3. Late-early, the ratio between late/early gene signatures (blue, low ratio; red, high ratio). HOXB, anterior HOXB genes and posterior HOXB13 (red, high relative expression; green, low relative expression). **B:** Consensus clustering of the Leeds UroA cases. Clusters, indicated as c1 through c3. Signatures S1 through S4 obtained by selection of analysis of variance significant genes and reordering by hierarchical clustering. Immune141_UP, continuous score on the immune signature (blue, low score; red, high score). Stromal141_UP, continuous score on the stromal signature (blue, low score; red, high score). Leeds classes, molecular class according to Hurst et al⁸ (2021): blue, E1; red, E2, green, E3; purple, class 4. Late-early, the ratio between late/early gene signatures (blue, low ratio; red, high ratio). HOXB, anterior HOXB genes and posterior HOXB13 (red, high relative expression; green, low relative expression). **** $P < 0.001$, **** $P < 0.0001$ (analysis of variance).

($P < 4 \times 10^{-7}$). Signature S4 was strongly associated with immune system process ($P < 5 \times 10^{-107}$), and signature S5 was associated with extracellular matrix organization ($P < 9 \times 10^{-17}$) and regulation of cell motility ($P < 3 \times 10^{-9}$). The distribution of the immune and stroma signatures (immune141_UP and stromal141_UP) was investigated for any natural thresholds by which not infiltrated and infiltrated cases could be separated. Both distributions were close to normal and hence noninformative ($H \approx 0$) (Supplemental Figure S12). Signature S6 was caused by coordinated expression of genes of the PCDHGA (protocadherin γ subfamily A) gene family. Signature S7 did not exhibit a clear-cut enrichment theme, but included the anterior HOXB mRNAs (HOXB2, HOXB3, HOXB5, HOXB6, HOXB8, and HOXB9), as well as KRT5. The HOXB gene cluster was almost exclusively expressed in tumor clusters c1 and c2 (Figure 7A).

In the Leeds data, four major ANOVA gene signatures, S1 through S4, explained the $k = 3$ clustering of the cases (Figure 7B). Signature S1 was enriched for cell cycle genes (cell cycle; $P < 4 \times 10^{-74}$). Signature S2 was as well enriched for cell cycle (mitotic cell cycle; $P < 1 \times 10^{-13}$) but also enriched for metabolic processes. Signatures S1 and S2 more or less defined tumor cluster c3. Signature S3 was highly enriched for immune response genes (immune response; $P < 4 \times 10^{-47}$) and defined tumor cluster c2 but was also seen in cluster c3. Signature S4 did not exhibit a

clear-cut enrichment profile but included RXRA, RARB, and GATA3, as well as HOXB2, HOXB5, HOXB6, HOXB8, and HOXB9 involved in normal differentiation, KRT5 and CD44 expressed by normal basal cells, and ITGA2, ITGA3, ITGB4, and ITGA6 involved in epithelial cell interaction networks. The S4 signature thus suggests highly structured/differentiated tumor cells. As for the UROMOL data, the immune and stroma signature scores were close to normal distributions, making a distinction between noninfiltrated and infiltrated categories arbitrary (Supplemental Figure S13). Taken together, the observed UroA intrasubtype variation in both the UROMOL and the Leeds cohorts is driven by a varying intensity of immune cell-related gene signatures, indicating the presence of infiltrating nontumor cells in the biopsy, and a variation in proliferation, none of which is considered cancer cell class-defining properties.

A HOXB/Cell Cycle Polarity in Non-Muscle-Invasive Urothelial Carcinomas

The above analyses suggest a polarity within UroA between high HOXB/low cell cycle gene expression and the opposite pattern. To investigate this further, cases were re-organized within each subtype from a high to a low HOXB/cell cycle ratio; Figures 8 and 9 show the UROMOL and Leeds cohorts, respectively. Within UroA, UroA.hox expresses the

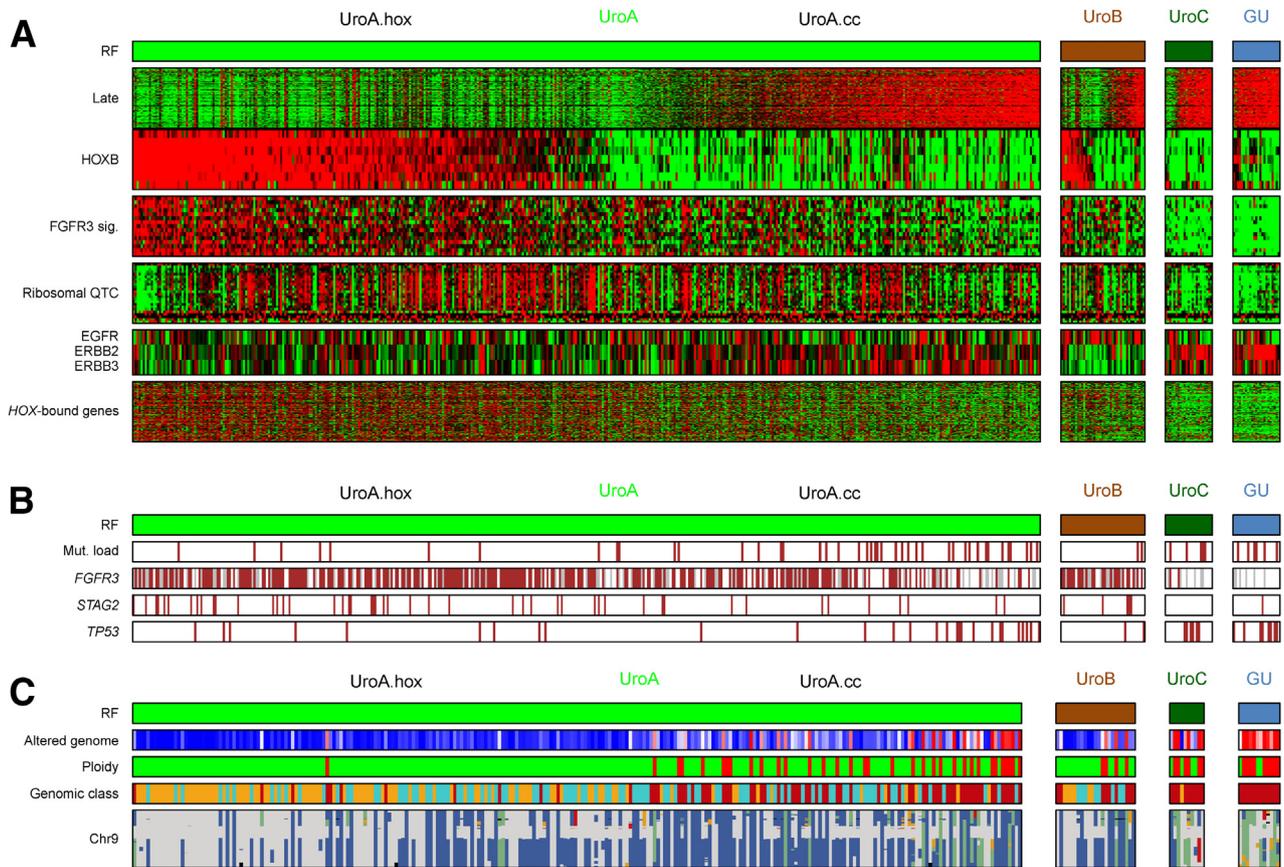


Figure 8 A *HOXB*/cell cycle gene expression polarity within non-muscle-invasive urothelial carcinoma. The UROMOL cohort. **A:** Gene expression profiles. RF, subtype according to the random forest algorithm. Cases within each subtype order from high *HOXB*, low cell cycle expression, to low *HOXB*, high cell cycle expression. Late, late cell cycle signature. *HOXB*, expression of *HOXB2* through *HOXB6*, *HOXB8*, and *HOXB9*. *FGFR3* sig., the *FGFR3* associated gene signature. Ribosomal QTC, the ribosomal gene expression signature. Individual genes (green, relative low expression; red, relative high expression). *HOX*-bound genes, the *HOX*-bound gene signature. **B:** Mutational (Mut.) load and gene mutations. RF, subtype according to the random forest algorithm. Cases within each subtype order from high *HOXB*, low cell cycle expression, to low *HOXB*, high cell cycle expression. Mut. Load: brown, high mutational load; white, low mutational load. Gene mutations: brown, mutation; white, no mutation; gray, no data. **C:** Genomic alterations. RF, subtype according to the random forest algorithm. Cases within each subtype order from high *HOXB*, low cell cycle expression, to low *HOXB*, high cell cycle expression. Altered genome, continuous variable of fraction of genome altered (blue, low fraction; red, high fraction). Genomic class: genomic classes according to Linskrog et al⁷ (2021): orange, class 1; blue, class 2; red, class 3. Ploidy: green, diploid; red, polyploid. Chr9, chromosome 9 imbalances (blue, loss; black, high balanced loss; orange, gain; red, high balanced gain; green, allelic imbalance without copy number alteration). GU, genomically unstable; Uro, urothelial-like.

FGFR3, ribosomal, and *HOX* binding gene signatures at higher levels than UroA.cc, in both cohorts. A tendency for coordinated increased expression of *ERBB2* and *ERBB3* in UroA.cc was observed, particularly in the UROMOL data (Figures 8A and 9A). UroB shows a similar *hox/cc* polarity as UroA, particularly evident in the UROMOL data. Both UroC and GU were dominated by cases with low expression of *HOXB* genes (Figures 8A and 9A). High mutational load was enriched in the UroA.cc cases, as were *TP53* mutations. Conversely, *FGFR3* and *STAG2* mutations were enriched in the tentative UroA.hox cluster (Figures 8B and 9B). The *HOXB*/cell cycle re-organization was also strongly associated with the fraction of genome altered and resulted in a good alignment with the original UROMOL and Leeds genomic classes (Figure 8C and 9C). Virtually all polyploid UROMOL UroA cases were on the UroA.cc side, a partition also seen within UroB. In addition, both UROMOL and Leeds UroA.hox groups were enriched for cases not showing

loss of chromosome 9. Consequently, there may be a biologically relevant polarity within the UroA and UroB groups of tumors related to *HOXB* expression and cell cycle activity.

Progression Events

Kaplan-Meier curves were produced for progression events in the LundTax classified UROMOL cohort (Figure 10A). The GU subtype showed the worst outcome, with most of the progression events occurring within 1 year. In the Leeds data, UroA showed the best prognosis (Figure 10B). In the merged UROMOL and Leeds cohort ($n = 720$) GU showed the worst outcome (Figure 10C). Among the UROMOL T1 tumors, GU and UroB showed the highest risk for progression, of which almost all GU progressions occurred within 1 year (Figure 10D). The data for the Leeds cohort were not as clear (Figure 10E). In the merged T1 cohort ($n = 216$), about 30% of the T1 GU cases progressed within 1

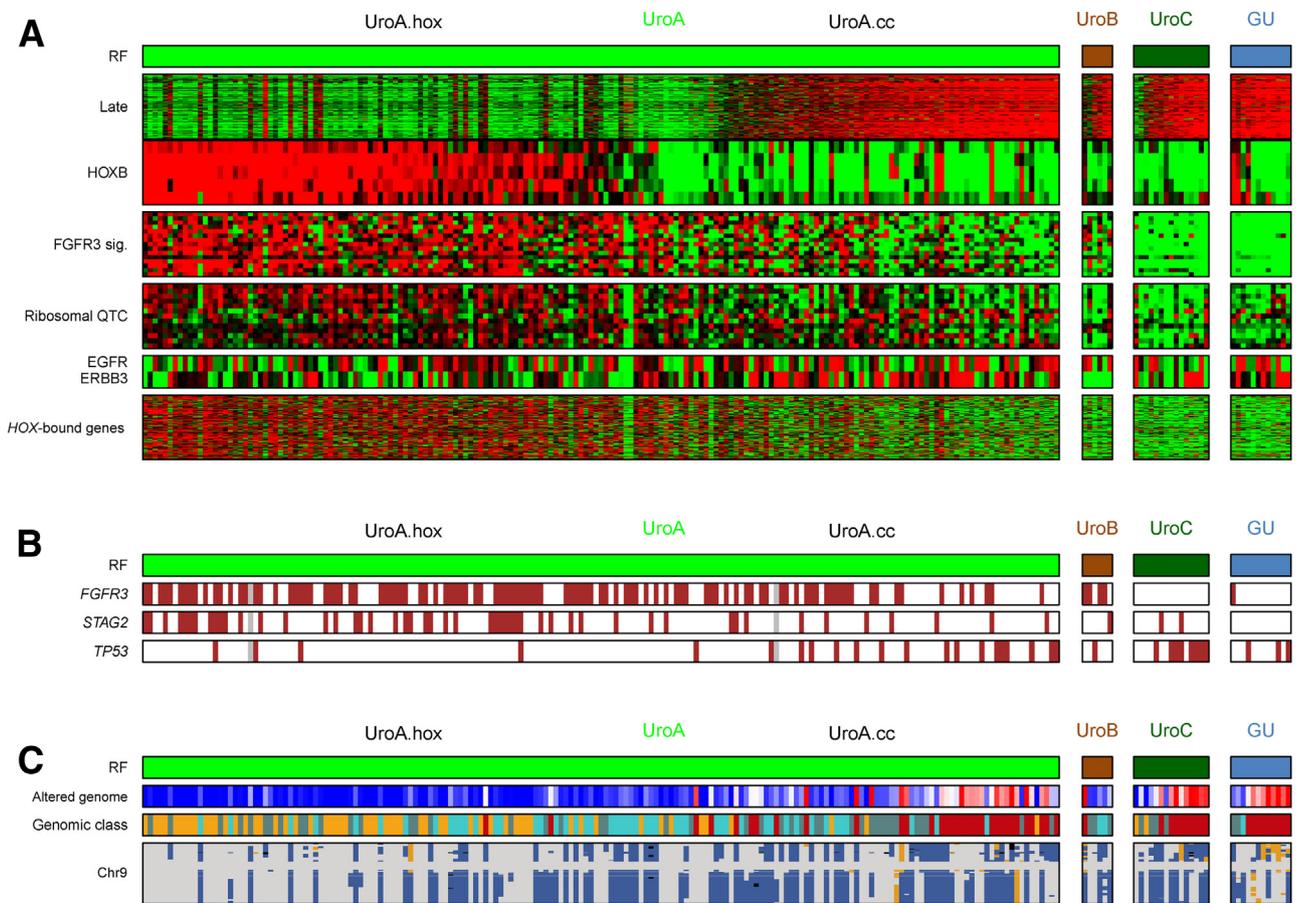


Figure 9 A *HOXB*/cell cycle gene expression polarity within non-muscle-invasive urothelial carcinoma. The Leeds cohort. **A:** Gene expression profiles. RF, subtype according to the random forest algorithm. Cases within each subtype order from high *HOXB*, low cell cycle expression, to low *HOXB*, high cell cycle expression. Late, late cell cycle signature. *HOXB*, expression of *HOXB2*, *HOXB5*, *HOXB6*, *HOXB8*, and *HOXB9*. *FGFR3* sig., the *FGFR3* associated gene signature. Ribosomal QTC, the ribosomal gene expression signature. Individual genes (green, relative low expression; red, relative high expression). *HOX*-bound genes, the *HOX*-bound gene signature. **B:** Mutational load and gene mutations. RF, subtype according to the random forest algorithm. Cases within each subtype order from high *HOXB*, low cell cycle expression, to low *HOXB*, high cell cycle expression. Gene mutations: brown, mutation; white, no mutation; gray, no data. **C:** Genomic alterations. RF, subtype according to the random forest algorithm. Cases within each subtype order from high *HOXB*, low cell cycle expression, to low *HOXB*, high cell cycle expression. Altered genome: continuous variable of fraction of genome altered (blue, low fraction; red, high fraction). Genomic class: genomic classes according to Hurst et al⁸ (2021): orange, class 1; blue, class 2; green, class 3; red, class 4. Chr9, chromosome 9 imbalances (blue, loss; black, high balanced loss; orange, gain; red, high balanced gain; green, allelic imbalance without copy number alteration). GU, genomically unstable; Uro, urothelial-like.

year (Figure 10F), in line with previously reported data.²⁸ As the Lund Taxonomy subtypes are based on biological properties only, clinical data are not included in their definitions; an alternative approach was applied to identify high-risk tumors using the complete UROMOL cohort as discovery data and the Leeds cohort for validation. Genes either positively or negatively associated with progression by receiver operating characteristic curve analyses were identified in the UROMOL data. The mean expression values of the 100 top-ranking up-regulated and the 100 top-ranking down-regulated genes were then used to separate progressors from non-progressors in the independent Leeds data (Supplemental Table S1). AUC values of 0.80 and 0.87 were obtained for the up and down signatures, respectively. As these signatures were derived from cohort-centered data, and not applicable to the single sample situation, a ratio between the up and down signatures within each sample

was produced; this ratio resulted in an AUC = 0.86 in the independent Leeds data (Figure 10G). When applied to uncentered Leeds data, using the mean rank-ordered values of the signatures, equivalent to single-sample analyses, an AUC of 0.83 was obtained. The optimal thresholds in the UROMOL and Leeds cohorts were then determined by receiver operating characteristic curve analysis in the respective cohorts, cases were categorized and merged, and Kaplan-Meier curves were produced. This identified a group with low progression risk, both in the complete (Figure 10H) and in the T1 restricted cohorts (Figure 10I).

Discussion

In this study, the LundTax classification system for urothelial carcinoma was applied on two large cohorts of

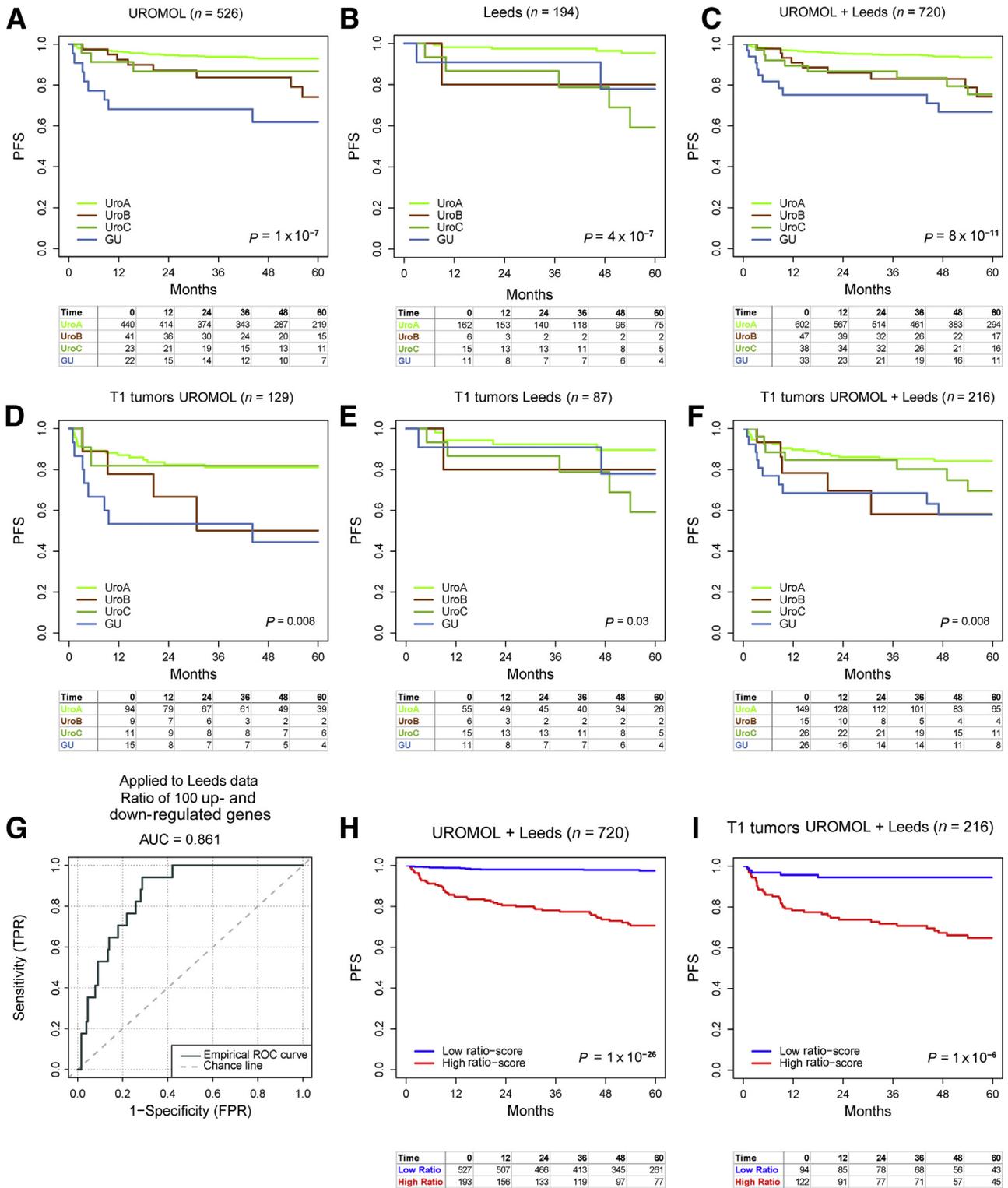


Figure 10 Progression events. **A:** Kaplan-Meier visualization of time to progression within the UROMOL cohort. **B:** Kaplan-Meier visualization of time to progression within the Leeds cohort. **C:** Kaplan-Meier visualization of time to progression in the merged UROMOL and Leeds cohorts. **D:** Kaplan-Meier visualization of time to progression in the pathologic stage T1 UROMOL cohort. **E:** Kaplan-Meier visualization of time to progression in the pathologic stage T1 Leeds cohort. **F:** Kaplan-Meier visualization of time to progression in the merged pathologic stage T1 UROMOL and Leeds cohorts. **G:** Receiver operating characteristic (ROC) curve analysis of the progression signature derived from the UROMOL data when applied to the independent Leeds cohort. **H:** Kaplan-Meier visualization of time to progression in the merged UROMOL and Leeds cohorts, classified as low risk (blue) and high risk (red), according to the progression signature. **I:** Kaplan-Meier visualization of time to progression in the merged UROMOL and Leeds cohorts of T1 tumors only classified as low risk (blue) and high risk (red), according to the progression signature. AUC, area under the curve; FPR, false positive rate; GU, genomically unstable; PFS, progression-free survival; TPR, true positive rate; Uro, urothelial-like.

non—muscle-invasive tumors, the UROMOL⁷ and the Leeds⁸ cohorts. Both cohorts were analyzed in parallel using data on genome-wide gene expression, gene mutation and copy number, as well as clinical follow-up data. However, the studies differ with respect to the means by which these data were obtained. The gene expression data in the UROMOL are based on RNA-seq, whereas the Leeds data are based on microarray hybridization. The variance across the Leeds could thus produce less contrasts. Gene mutation data were in most cases obtained by RNA-seq analyses in the UROMOL data set, but by DNA-sequencing analyses in the Leeds data. Although this may introduce biases and produce different detection levels and sensitivities to noise, almost all of the findings were validated in both data sets. The only major discrepancies seen were in the UroC subtype, where the Leeds UroC showed higher frequencies of *TP53* and *CDKN2A* deletions and *CCND1* amplifications than the UROMOL UroC. However, the genomic data for UROMOL UroC were based on only 10 cases (Figure 3) and may thus be uncertain.

Both Lindskrog et al⁷ and Hurst et al⁸ apply clustering based on gene expression data and hierarchical clustering and nonnegative matrix factorization, respectively, to define subtypes/classes. Lindskrog et al⁷ arrive at a four-tiered system, with classes 1, 2a, 2b, and 3. The UROMOL classes appear analogous to the tumor clusters I, IIa, IIb, and III described in an early microarray study on NMIBC.²⁹ Hurst et al⁸ also arrive at a four-tiered system with the subtypes negative matrix factorization 1 through negative matrix factorization 4. As both approaches use genome-wide gene expression as their primary source for grouping, the obtained groups/classes of tumors will be heavily influenced by infiltration of nontumor cells as well as by tumor cell proliferation. The Lund Taxonomy approach differs in this regard as infiltration and proliferation are not considered class-defining properties; signals from infiltrating cells are not tumor cell intrinsic, and proliferation behaves as a normally distributed variable, where only arbitrary thresholds may be established (ie, it is a quantitative and not a qualitative trait) (Supplemental Figures S14 and S15). Instead, the LundTax system is focused on defining crucial features of the cancer cells proper.⁹ To achieve this, the gene expression (mRNA) determined classes were adjusted by immunohistochemistry and tumor clusters identified as infiltrated were deconstructed into groups of different cancer cell phenotypes.^{9,10} By training our mRNA-based classification algorithms on such immunohistochemistry-adjusted data, the algorithms naturally ignore infiltration, as well as proliferation, as class-defining features.^{9,10,12,13} In short, mRNA expression is used to identify immunohistochemistry-adjusted classes of tumors with distinct cancer cell phenotypes, classes that show good correspondence with the molecular pathology of the cancer cells²⁶ as well as with key gene mutations and genomic alterations.¹⁰

The classification of the two cohorts showed that, in contrast to MIBC, practically all NMIBC tumors belong to

the luminal class of tumors: the urothelial-like divided into the three variants (UroA, UroB, and UroC) and the genomically unstable (GU). The proportions of these subtypes were similar in the two independent cohorts. Several previously defined class-defining signatures confirmed the classification of UroA, UroB, UroC, and GU. The independent gene mutation and genomic data were then used to validate the grouping. *FGFR3* mutations were, as previously reported, frequent in UroA and UroB but almost absent in UroC and GU. *RBI* and *TP53* mutations were, on the other hand, significantly enriched in UroC and GU cases. Specific genomic alterations followed the gene mutation pattern as losses of *TP53* (17p13) and *RBI* (13q14) and gains of *E2F3* (6p22) were almost exclusively seen in UroC and GU tumors. Furthermore, the organization of tumors into LundTax subtypes, and according to cell cycle activity within each subtype, more or less recapitulated the genomic copy number classes originally described for the UROMOL and Leeds cohorts, respectively. Hence, the distribution of gene mutations and genomic alterations emphasizes the identity of the LundTax subtypes determined by gene expression only.

The fact that UroA constituted >80% of the cases motivated a thorough analysis of this group. An unsupervised approach was used, where consensus clustering with predefined k values of 2 to 5 was utilized. None of the k values produced convincing coclustering matrices. The best interpretation of this is that the data cannot be further divided into well-defined and qualitatively distinct groups (ie, the within UroA variation is too small). However, to investigate which factors contributed to the remaining transcriptional variation, tentative k-solutions for each cohort were selected and then ANOVA was used to identify genes responsible for grouping of the data. Three major themes were identified: cell cycle signatures, infiltration signatures, and one less distinct signature, including genes associated with low-grade and well-differentiated tumors. The Lund Taxonomy does not consider infiltration and proliferation to be cancer cell class—defining features and thus further divisions are not motivated. However, a common theme in the last signature was the coordinated expression of the anterior *HOXB* genes. This gene set was expressed in specific tumor clusters in both cohorts after consensus clustering. Organization of the tumors along a UroA.hox/UroA.cc axis gave some insight into both the biology and the genomics of the UroA subtype. Although the two cohorts are of excellent quality, they are composed of non—muscle-invasive tumors only, and hence the gene expression variance across the data, compared with MIBC, is small. Normalization of such data makes it possible to detect subtle changes not detected when using a cohort with full urothelial carcinoma spectrum. Consequently, the variation in *HOXB* expression, and the identification of a *HOXB*/cell cycle polarity, should not be interpreted as a class-defining property. The detected down-regulation of *HOXB* genes could simply be caused by an increasing proportion of cells involved in cell

proliferation, a dynamic transition rather than a shift in subtype, as increased proliferation reduces the expression of differentiation genes. Consequently, the *hox/cc* polarity is an example of within-subtype variation. Exactly these two clusters were observed as clusters MS1a and MS1b when the Lund Taxonomy was first defined but merged to make up the UroA subtype.¹

Progression is the most clinically relevant outcome measure in non-muscle-invasive bladder cancer. There is an association between progression and the GU subtype.²⁸ This finding is now validated in the UROMOL and Leeds cohorts, with overall progression rates of 36% and 18%, respectively, and 53% and 18%, respectively, for T1 tumors. As an alternative to associate progression with specific subtypes, systematic receiver operating characteristic curve analyses were performed to identify high-risk tumors irrespective of molecular subtype. However, only 25% to 35% of the high-risk tumors eventually progress to muscle invasive growth, a relatively low positive predictive value but similar to previously published progression signatures.^{30,31} An explanation for the uncertain predictions could be that the risk for progression is caused by more than one factor (eg, pathologic grade, proliferation level, tumor genomics, or immunologic status of the patient) and that single high informative factors/variables would be hard to identify.⁷ However, the question may not be how to predict progression events but rather if it is possible using one single index tumor. Recurring tumors differ from preceding tumors with respect to genomic alterations and gene mutations.³² In particular, most recurring tumors do not originate from previous overt tumors, but rather represent clonally related *de novo* tumors, originating from a shared field of genomically destabilized urothelium.^{32,33} It could thus be more important to characterize the nature of this field (eg, the degree of field heterogeneity)^{34–36} to estimate risk for progression. In this scenario, an overt index tumor is one single tumor occurrence from a possibly large and evolving field of destabilized urothelium.³³ Irrespectively, the provided receiver operating characteristic curve-derived risk ratio identifies one group of tumors that almost never produce progressions, not even among T1 tumors.

As for the classification of urothelial carcinoma, the LundTax system is valid for both non-muscle-invasive and muscle-invasive tumors.^{1,9,10} Furthermore, the rule-based single-sample classification algorithm used in the present investigation is applicable to both non-muscle-invasive and muscle-invasive cases.^{12,13} It is important to emphasize that non-muscle-invasive and muscle-invasive variants of the LundTax subtypes do not differ in their class-defining properties.^{11,37} Consequently, it would not make sense to have separate molecular subtypes for different tumor stages. It is more appropriate to have the same subtypes allowed to occur in different frequencies, depending on pathologic stage. For direct comparisons between the UROMOL and Leeds classification systems with the LundTax classification, see [Supplemental Figures](#)

[S16](#) and [S17](#). The focus on cancer cell phenotypes for classification does not exclude the clinical importance of factors like proliferation and infiltration. On the contrary, they are important for the patient outcomes. However, if these features are mixed and used for classification purposes by, for example, using clustering algorithms on genome-wide gene expression data as only source, subtypes would sometimes be defined by cancer cells only and sometimes by whole biopsy properties.⁹ As an alternative, we promote a classification system based on cancer cell phenotypes with good alignment with cancer cell molecular pathology.^{9–13} Although there is a complete lack of published phase 3 clinical trials in which molecular classification is used to stratify treatment in NMIBC, new treatments (eg, checkpoint inhibitors in the BCG-unresponsive setting)³⁸ will likely have different outcomes depending on molecular subtype, as seen in the metastatic setting in the IMvigor 210 trial.³⁹ Furthermore, molecular subtyping is of clinical value to predict chemotherapy response.^{40–42} This approach is likely to be productive also to the NMIBC setting, where adjuvant intravesical chemotherapy frequently is used. Hence, biologically relevant molecular classification will be important for future translational studies of NMIBC tumors. To describe a given tumor, we suggest a report based on one single RNA-seq analysis from which data on molecular subtype are derived as well as indices for proliferation, immune and stromal infiltration, molecular grade, and progression risk, combined with pathologic grade and stage, as well as with clinical data. By keeping these variables separated, it will be possible to identify which characteristic has independent value in each specific clinical setting. Such a classification system will give the most coherent and systematic description of the individual tumors.

Acknowledgments

We thank Sia Viborg Lindskrog and Lars Dyrskjøt for critically reviewing the manuscript.

Author Contributions

N.M., P.E., and G.S. performed bioinformatic analyses; C.D.H. and M.A.K. provided access to unpublished data; M.H., N.M., P.E., G.S., F.L., C.B., and M.H. drafted the manuscript; N.M., P.E., C.B., C.D.H., M.A.K., G.S., F.L., and M.H. provided intellectual input and critically read the manuscript; F.L. and M.H. provided funding. The study was conceptualized and designed by M.H.

Supplemental Data

Supplemental material for this article can be found at <http://doi.org/10.1016/j.jmoldx.2022.05.006>.

References

- Sjödahl G, Lauss M, Lövgren K, Chebil G, Gudjonsson S, Veerla S, Patschan O, Aine M, Fernö M, Ringnér M, Månsson W, Liedberg F, Lindgren D, Höglund M: A molecular taxonomy for urothelial carcinoma. *Clin Cancer Res* 2012, 18:3377–3386
- Damrauer JS, Hoadley KA, Chism DD, Fan C, Tiganelli CJ, Wobker SE, Yeh JJ, Milowsky MI, Iyer G, Parker JS, Kim WY: Intrinsic subtypes of high-grade bladder cancer reflect the hallmarks of breast cancer biology. *Proc Natl Acad Sci U S A* 2014, 111:3110–3115
- Choi W, Porten S, Kim S, Willis D, Plimack ER, Hoffman-Censits J, Roth B, Cheng T, Tran M, Lee IL, Melquist J, Bondaruk J, Majewski T, Zhang S, Pretzsch S, Baggerly K, Siefker-Radtke A, Czerniak B, Dinney CP, McConkey DJ: Identification of distinct basal and luminal subtypes of muscle-invasive bladder cancer with different sensitivities to frontline chemotherapy. *Cancer Cell* 2014, 25:152–165
- Robertson AG, Kim J, Al-Ahmadie H, Bellmunt J, Guo G, Cherniack AD, et al: Comprehensive molecular characterization of muscle-invasive bladder cancer. *Cell* 2017, 171:540–556.e25
- Lerner SP, McConkey DJ, Hoadley KA, Chan KS, Kim WY, Radvanyi F, Höglund M, Real FX: Bladder cancer molecular taxonomy: summary from a consensus meeting. *Bladder Cancer* 2016, 2:37–47
- Kamoun A, de Reyniès A, Allory Y, Sjödahl G, Robertson AG, Seiler R, Hoadley KA, Groeneveld CS, Al-Ahmadie H, Choi W, Castro MAA, Fontugne J, Eriksson P, Mo Q, Kardos J, Zlotta A, Hartmann A, Dinney CP, Bellmunt J, Powles T, Malats N, Chan KS, Kim WY, McConkey DJ, Black PC, Dyrskjöt L, Höglund M, Lerner SP, Real FX, Radvanyi F; Bladder Cancer Molecular Taxonomy Group: A consensus molecular classification of muscle-invasive bladder cancer. *Eur Urol* 2020, 77:420–433
- Lindskog SV, Prip F, Lamy P, Taber A, Groeneveld CS, Birkenkamp-Demtröder K, et al: An integrated multi-omics analysis identifies prognostic molecular subtypes of non-muscle-invasive bladder cancer. *Nat Commun* 2021, 12:2301–2318
- Hurst CD, Cheng G, Platt FM, Castro MAA, Marzouka NS, Eriksson P, Black EVI, Alder O, Lawson ARJ, Lindskog SV, Burns JE, Jain S, Roulson JA, Brown JC, Koster J, Robertson AG, Martincorena I, Dyrskjöt L, Höglund M, Knowles MA: Stage-stratified molecular profiling of non-muscle-invasive bladder cancer enhances biological, clinical, and therapeutic insight. *Cell Rep Med* 2021, 2:100472
- Sjödahl G, Eriksson P, Liedberg F, Höglund M: Molecular classification of urothelial carcinoma: global mRNA classification versus tumour-cell phenotype classification. *J Pathol* 2017, 242:113–125
- Marzouka NA, Eriksson P, Rovira C, Liedberg F, Sjödahl G, Höglund M: A validation and extended description of the Lund taxonomy for urothelial carcinoma using the TCGA cohort. *Sci Rep* 2018, 8:3737
- Bernardo C, Eriksson P, Marzouka NA, Liedberg F, Sjödahl G, Höglund M: Molecular pathology of the luminal class of urothelial tumors. *J Pathol* 2019, 249:308–318
- Marzouka NA, Eriksson P: multiclassPairs: an R package to train multiclass pair-based classifier. *Bioinformatics* 2021, 37:3043–3044
- Eriksson P, Marzouka NA, Sjödahl G, Bernardo C, Liedberg F, Höglund M: A comparison of rule-based and centroid single-sample multiclass predictors for transcriptomic classification. *Bioinformatics* 2021, 38:1022–1029
- Patro R, Duggal G, Love MI, Irizarry RA, Kingsford C: Salmon provides fast and bias-aware quantification of transcript expression. *Nat Methods* 2017, 14:417–419
- Hedegaard J, Lamy P, Nordentoft I, Algaba F, Høyer S, Ulhøi BP, et al: Comprehensive transcriptional analysis of early-stage urothelial carcinoma. *Cancer Cell* 2016, 30:27–42
- Lindgren D, Sjödahl G, Lauss M, Staaf J, Chebil G, Lövgren K, Gudjonsson S, Liedberg F, Patschan O, Månsson W, Fernö M, Höglund M: Integrated genomic and gene expression profiling identifies two major genomic circuits in urothelial carcinoma. *PLoS One* 2012, 7:e38863
- Eriksson P, Aine M, Veerla S, Liedberg F, Sjödahl G, Höglund M: Molecular subtypes of urothelial carcinoma are defined by specific gene regulatory systems. *BMC Med Genomics* 2015, 8:25
- Yoshihara K, Shahmoradgolji M, Martínez E, Vegesna R, Kim H, Torres-Garcia W, Treviño V, Shen H, Laird PW, Levine DA, Carter SL, Getz G, Stenke-Hale K, Mills GB, Verhaak RGW: Inferring tumour purity and stromal and immune cell admixture from expression data. *Nat Commun* 2013, 4:2612
- Liedberg F, Lauss M, Patschan O, Aine M, Chebil G, Cwikiel M, Engilbertsson H, Gudjonsson S, Kollberg P, Sjödahl G, Månsson W, Lindgren D, Höglund M; Lund Bladder Cancer Group: The importance of being grade 3: WHO 1999 versus WHO 2004 pathologic grading. *Eur Urol* 2012, 62:620–623
- Geman D, d'Avignon C, Naiman DQ, Winslow RL: Classifying gene expression profiles from pairwise mRNA comparisons. *Stat Appl Genet Mol Biol* 2004, 3:9
- Lerch F, Ultsch A, Lötsch J: Distribution Optimization: an evolutionary algorithm to separate Gaussian mixtures. *Sci Rep* 2020, 10:648
- Ultsch A, Thrun MC, Hansen-Goos O, Lötsch J: Identification of molecular fingerprints in human heat pain thresholds by use of an interactive mixture model R toolbox (AdaptGauss). *Int J Mol Sci* 2015, 16:25897–25911
- Durinck S, Spellman P, Birney E, Huber W: Mapping identifiers for the integration of genomic datasets with the R/Bioconductor package biomaRt. *Nat Protoc* 2009, 4:1184–1191
- Wilkinson DM, Hayes DN: ConsensusClusterPlus: a class discovery tool with confidence assessments and item tracking. *Bioinformatics* 2010, 26:1572–1573
- Mi H, Huang X, Muruganujan A, Tang H, Mills C, Kang D, Thomas PD: PANTHER version 14: more genomes, a new PANTHER GO-slim and improvements in enrichment analysis tools. *Nucleic Acids Res* 2019, 47:D419–D426
- Eriksson P, Sjödahl G, Chebil G, Liedberg F, Höglund M: HER2 and EGFR amplification and expression in urothelial carcinoma occurs in distinct biological and molecular contexts. *Oncotarget* 2017, 8:48905–48914
- Rebouissou S, Héroult A, Letouzé E, Neuzillet Y, Laplanche A, Ofualuka K, Maillé P, Leroy K, Riou A, Lepage ML, Vordos D, de la Taille A, Denoux Y, Sibony M, Guyon F, Lebre T, Benhamou S, Allory Y, Radvanyi F: CDKN2A homozygous deletion is associated with muscle invasion in FGFR3-mutated urothelial bladder carcinoma. *J Pathol* 2012, 227:315–324
- Patschan O, Sjödahl G, Chebil G, Lövgren K, Lauss M, Gudjonsson S, Kollberg P, Eriksson P, Aine M, Månsson W, Fernö M, Liedberg F, Höglund M: A molecular pathologic framework for risk stratification of stage T1 urothelial carcinoma. *Eur Urol* 2015, 68:824–832
- Lindgren D, Liedberg F, Andersson A, Chebil G, Gudjonsson S, Borg A, Månsson W, Fioretos T, Höglund M: Molecular characterization of early-stage bladder carcinomas by expression profiles, FGFR3 mutation status, and loss of 9q. *Oncogene* 2006, 25:2685–2696
- Dyrskjöt L, Zieger K, Real FX, Malats N, Carrato A, Hurst C, Kotwal S, Knowles M, Malmström PU, de la Torre M, Wester K, Allory Y, Vordos D, Caillaud A, Radvanyi F, Hein AM, Jensen JL, Jensen KM, Marcussen N, Orntoft TF: Gene expression signatures predict outcome in non-muscle-invasive bladder carcinoma: a multicenter validation study. *Clin Cancer Res* 2007, 13:3545–3551
- Dyrskjöt L, Reinert T, Algaba F, Christensen E, Nieboer D, Hermann GG, Mogensen K, Beukers W, Marquez M, Segersten U,

- Høyer S, Uhløi BP, Hartmann A, Stöhr R, Wach S, Nawroth R, Schwamborn K, Tulic C, Simic T, Junker K, Harving N, Petersen AC, Jensen JB, Keck B, Grimm MO, Horstmann M, Maurer T, Steyerberg EW, Zwarthoff EC, Real FX, Malats N, Malmström PU, Ørntoft TF: Prognostic impact of a 12-gene progression score in non-muscle-invasive bladder cancer: a prospective multicentre validation study. *Eur Urol* 2017, 72:461–469
32. Marzouka NA, Lindgren D, Eriksson P, Sjö Dahl G, Bernardo C, Liedberg F, Axelson H, Höglund M: Recurring urothelial carcinomas show genomic rearrangements incompatible with a direct relationship. *Sci Rep* 2020, 10:19539
33. Höglund M: Bladder cancer, a two phased disease? *Semin Cancer Biol* 2007, 17:225–232
34. Strandgaard T, Nordentoft I, Lamy P, Christensen E, Thomsen MBH, Jensen JB, Dyrskjöt L: Mutational analysis of field cancerization in bladder cancer. *Bladder Cancer* 2020, 6:253–264
35. Thomsen MBH, Nordentoft I, Lamy P, Vang S, Reinert L, Mapendano CK, Høyer S, Ørntoft TF, Jensen JB, Dyrskjöt L: Comprehensive multiregional analysis of molecular heterogeneity in bladder cancer. *Sci Rep* 2017, 7:11702
36. Lawson ARJ, Abascal F, Coorens THH, Hooks Y, O'Neill L, Latimer C, Raine K, Sanders MA, Warren AY, Mahbubani KTA, Bareham B, Butler TM, Harvey LMR, Cagan A, Menzies A, Moore L, Colquhoun AJ, Turner W, Thomas B, Gnanapragasam V, Williams N, Rassl DM, Vöhringer H, Zumalave S, Nangalia J, Tubfo JMC, Gerstung M, Saeb-Parsy K, Stratton MR, Campbell PJ, Mitchell TJ, Martincorena I: Extensive heterogeneity in somatic mutation and selection in the human bladder. *Science* 2020, 370:75–82
37. Sjö Dahl G, Eriksson P, Patschan O, Marzouka NA, Jakobsson L, Bernardo C, Lövgren K, Chebil G, Zwarthoff E, Liedberg F, Höglund M: Molecular changes during progression from non-muscle-invasive to advanced urothelial carcinoma. *Int J Cancer* 2020, 146:2636–2647
38. Balar AV, Kamat AM, Kulkarni GS, Uchio EM, Boormans JL, Roumiguié M, Krieger LEM, Singer EA, Bajorin DF, Grivas P, Seo HK, Nishiyama H, Konety BR, Li H, Nam K, Kapadia E, Frenkl T, de Wit R: Pembrolizumab monotherapy for the treatment of high-risk non-muscle-invasive bladder cancer unresponsive to BCG (KEYNOTE-057): an open-label, single-arm, multicentre, phase 2 study. *Lancet Oncol* 2021, 22:919–930
39. Mariathasan S, Turley SJ, Nickles D, Castiglioni A, Yuen K, Wang Y, et al: TGFβ attenuates tumour response to PD-L1 blockade by contributing to exclusion of T cells. *Nature* 2018, 554:544–548
40. Sjö Dahl G, Abrahamsson J, Holmsten K, Bernardo C, Chebil G, Eriksson P, Johansson I, Kollberg P, Lindh C, Lövgren K, Marzouka NA, Olsson H, Höglund M, Ullén A, Liedberg F: Different responses to neoadjuvant chemotherapy in urothelial carcinoma molecular subtypes. *Eur Urol* 2022, 81:523–532
41. Taber A, Christensen E, Lamy P, Nordentoft I, Prip F, Lindskrog SV, Birkenkamp-Demtröder K, Okholm TLH, Knudsen M, Pedersen JS, Steiniche T, Agerbæk M, Jensen JB, Dyrskjöt L: Molecular correlates of cisplatin-based chemotherapy response in muscle invasive bladder cancer by integrated multi-omics analysis. *Nat Commun* 2020, 11: 4858
42. Seiler R, Ashab HAD, Erho N, van Rhijn BWG, Winters B, Douglas J, Van Kessel KE, Franssen van de Putte EE, Sommerlad M, Wang NQ, Choerung V, Gibb EA, Palmer-Aronsten B, Lam LL, Buerki C, Davicioni E, Sjö Dahl G, Kardos J, Hoadley KA, Lerner SP, McConkey DJ, Choi W, Kim WY, Kiss B, Thalmann GN, Todenhöfer T, Crabb SJ, North S, Zwarthoff EC, Boormans JL, Wright J, Dall'Era M, van der Heijden MS, Black PC: Impact of molecular subtypes in muscle-invasive bladder cancer on predicting response and survival after neoadjuvant chemotherapy. *Eur Urol* 2017, 72:544–554