Molecular characterization reveals genomic and transcriptomic subtypes of metastatic urothelial carcinoma

J. Alberto Nakauma-González[1,2,3,^], Maud Rijnders[3,^], Job van Riet[3], Michiel S. van der Heijden[4], Jens Voortman[5], Edwin Cuppen[6,7], Niven Mehra[8], Sandra van Wilpe[8], Sjoukje F. Oosting[9], L. Lucia Rijstenberg[10], Hans M. Westgeest[11], Ellen C. Zwarthoff[10], Ronald de Wit[3], Astrid A.M. van der Veldt[3], Harmen J. G. van de Werken[1,2,^,*], Martijn P. J. Lolkema[3,^,*], Joost L. Boormans[2,12,^,*]

1. Cancer Computational Biology Center, Erasmus MC Cancer Institute, Erasmus University Medical Center, Rotterdam 3015 GD, the Netherlands
2. Department of Urology, Erasmus MC Cancer Institute, Erasmus University Medical Center, Rotterdam 3015 GD, the Netherlands
3. Department of Medical Oncology, Erasmus MC Cancer Institute, Erasmus University Medical Center, Rotterdam 3015 GD, the Netherlands
4. Department of Medical Oncology, the Netherlands Cancer Institute, Amsterdam 1066 CX, the Netherlands
5. Department of Medical Oncology, Amsterdam UMC, Vrije Universiteit Amsterdam, Cancer Center Amsterdam, Amsterdam 1081 HV, the Netherlands
6. Center for Molecular Medicine and Oncode Institute, University Medical Center Utrecht, Utrecht 3584 CX, the Netherlands
7. Hartwig Medical Foundation, Amsterdam 1098 XH, the Netherlands
8. Department of Medical Oncology, Radboud University Medical Center, Nijmegen 6500 HB, the Netherlands
9. Department of Medical Oncology, University Medical Center Groningen, University of Groningen, Groningen 9713 GZ, the Netherlands
10. Department of Pathology, Erasmus MC University Medical Center Rotterdam, Rotterdam 3015 GD, the Netherlands
11. Department of Internal Medicine, Amphia hospital, Breda 4818 CK, the Netherlands
12. Lead contact

^ These authors contributed equally.

* Correspondence: h.vandewerken@erasmusmc.nl **(HJGvdW)**, m.lolkema@erasmusm.nl (**MPJL**), j.boormans@erasmusmc.nl **(JLB)**

Keywords

Molecular Profiling, Neoplasm Metastases, RNA-seq, Urologic Neoplasms, Whole Genome Sequencing.

1

Abstract

Metastatic urothelial carcinoma (mUC) is a lethal cancer for which few therapeutic options exist. To define the molecular landscape of mUC and to identify targets for therapy, we performed whole genome DNA- and RNA-sequencing on fresh-frozen metastatic tumor biopsies of 116 mUC patients. Driver genes resembled those reported for primary UC; yet, three putative driver genes unique to mUC were identified: *CNTNAP5*, *RARG* and *MGP*. Consensus clustering based on mutational signatures revealed two major genomic subtypes. The most prevalent subtype (67%) consisted almost exclusively of tumors with high APOBEC mutagenesis. Five RNA-based subtypes were identified, of which four resembled those reported for primary UC, and one had a non-specified phenotype. By integrating the genomic and transcriptomic data potential therapeutic options per subtype and individual patient are proposed. This study serves as a reference for subtype-oriented and patient-specific research on the etiology of mUC, and for novel drug development.

Introduction

Urothelial cancer (UC) is a molecularly and clinically heterogeneous disease. Non-muscle invasive bladder cancer (NMIBC) is characterized by excellent survival but high recurrence rates, whereas muscle-invasive bladder cancer (MIBC) has high metastatic potential and poor patient outcome despite aggressive local and systemic treatment (2). Comprehensive molecular profiling of UC has been restricted to NMIBC (3) and localized MIBC (4). At the genomic level, NMIBC is characterized by frequent *FGFR3* and *PIK3CA* mutations, whereas *TP53* mutations are uncommon (2). In MIBC, *TP53* is the most commonly mutated gene (5). *The Cancer Genome Atlas* (TCGA) initiative molecularly characterized 412 chemotherapy-naïve primary MIBC patients and found that a subgroup of patients had high Apolipoprotein B mRNA Editing Catalytic Polypeptide-like (APOBEC) signature mutagenesis and high mutational burden. The patients in this subgroup had an excellent 5-year overall survival rate of 75% (4). At the transcriptomic level, MIBC can be stratified into basal and luminal subtypes. A recent study proposed a consensus molecular classification of MIBC, consisting of six subtypes: basal/squamous, luminal non-specified, luminal papillary, luminal unstable, neuroendocrine-like (NE-like), and stroma-rich (6). These subtypes included distinct genomic alterations and clinical and pathological characteristics, which might guide therapeutic decision making.

A comprehensive multi-omics characterization of mUC has not yet been performed. A previous study reported the clonal evolution of mUC by whole-exome sequencing (WES) in a cohort of 32 chemotherapy-treated patients, and showed that APOBEC mutagenesis was clonally enriched in chemotherapy-treated mUC (7). Having more knowledge on the molecular characteristics of mUC is crucial for more robust and accurate patient stratification and for rational drug development paths that will eventually improve the outcome of this lethal cancer. In the present study, we conducted a comprehensive genomic and transcriptomic analysis of freshly obtained metastatic biopsies of 116 mUC patients, with the aim of identifying key molecular insights into tumorigenesis and defining molecular subtypes of mUC.

3

Results

Cohort description

The study population consisted of 210 patients with advanced or mUC who were prospectively enrolled in a multicenter clinical trial (NCT01855477; Data S1.1) and who were scheduled for 1st or 2nd line palliative systemic treatment. Whole genome sequencing (WGS) was successfully performed on DNA obtained from freshly obtained biopsies from metastatic sites in 116 mUC patients (124 samples), and matched RNAseq was available for 90 patients (97 samples, Figure 1A-B). Sequential biopsies of a metastatic lesion taken at the time of clinical or radiological disease progression from eight patients were additionally sequenced. Patient characteristics are provided in Table S1.1.
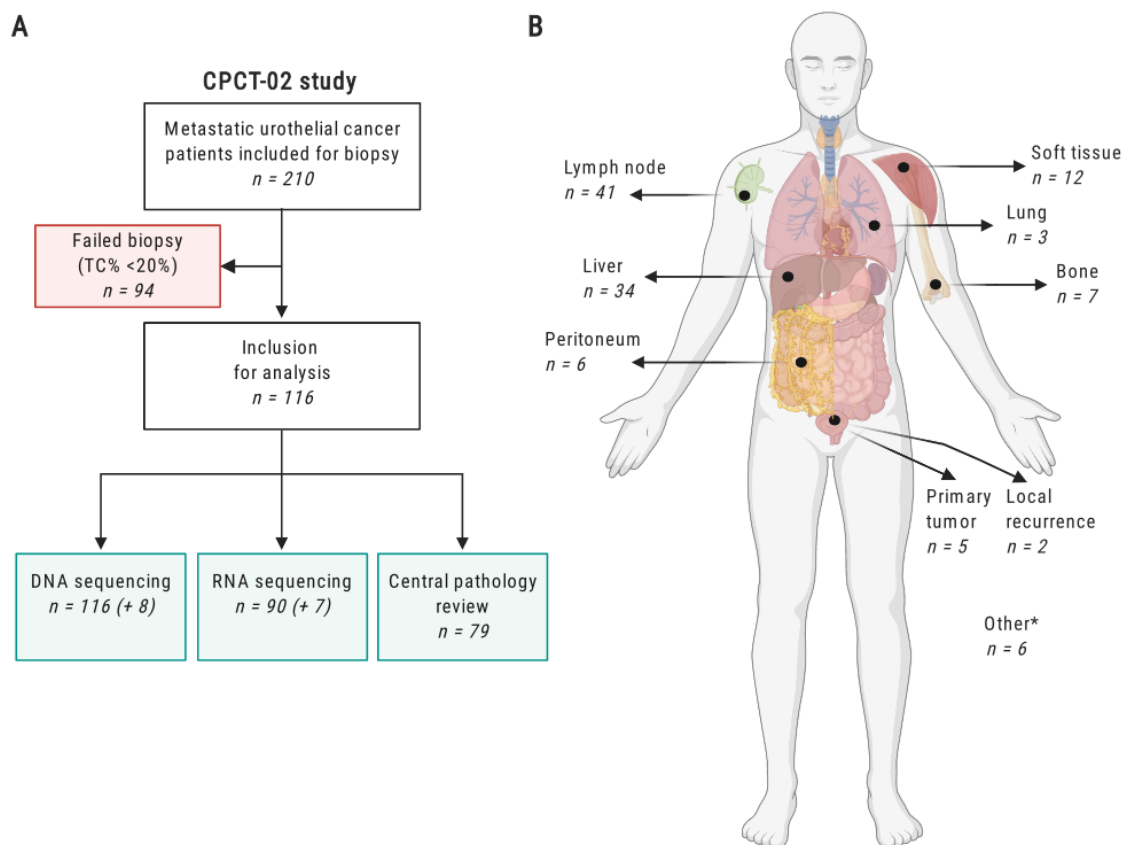
Figure 1 – Overview of the study design and biopsy sites of 116 patients with metastatic urothelial cancer

(A) Flowchart of patient inclusion. Patients with advanced or metastatic urothelial cancer who were
scheduled for systemic palliative treatment were selected from the prospective Center for

Personalized Cancer Treatment (CPCT-02) patient cohort (n = 210). Patients were excluded if the tumor cell percentage in the biopsy was <20%, resulting in WGS data and RNAseq for 116 and 90 patients, respectively. Tissue slides of 79 patients were available for central pathology review (primary tumor and/or metastatic biopsy). DNA +8 and RNA +7 indicate the numbers of patients from whom a second biopsy was obtained at disease progression.

(B) Overview of the number of biopsies per site analyzed by WGS. * Other biopsy sites included abdominal or pelvic masses (n = 3), adrenal gland (n = 1), and brain (n = 1), or unspecified biopsy site (n = 1).

Genomic landscape of mUC

Analysis of WGS (mean coverage 99 X) and matched blood samples (mean coverage 38 X) identified a median of 20,634 single nucleotide variants (SNVs), 1,018 small insertions/deletions (Indels) and 175 somatic multi-nucleotide variants (MNVs; Figure S1A). SNVs were more frequent in coding regions (7.63 SNVs per Megabase pair; SNVs/Mbp) than in the whole genome (7.22 SNVs/Mbp; Wilcoxon signed-rank test p = 0.0024; Figure S1B). However, Indels and MNVs were less frequent in coding regions (Wilcoxon signed-rank test p < 0.001 and p = 0.0072, respectively). Analysis of all SNVs revealed that 68% of all SNVs were clonal with a median of 74% per tumor (Table S1.5), and that 91% of the tumors were enriched for APOBEC associated mutations (73% high and 18% low enrichment of APOBEC mutagenesis; Figure 2A). The mean contribution of APOBEC COSMIC signatures (Signatures 2 and 13) in tumors with high APOBEC mutagenesis enrichment was 52% *versus* 15% in tumors with low APOBEC mutagenesis.

Genes harboring more mutations in their coding sequence than expected by random chance were analyzed with dNdScv; the analysis revealed 18 significantly mutated genes (Table S1.7). Three of those genes – *CNTNAP5 (13%), RARG (6%)* and *MGP (4%)* – were unique to this mUC cohort (8). Regarding known driver genes, mUC lesions harbored more somatic mutations in *TP53* than numbers reported in TCGA (non-synonymous mutations and indels; 60% *vs* 49%, Fisher's exact test p = 0.021, Data S1.3).
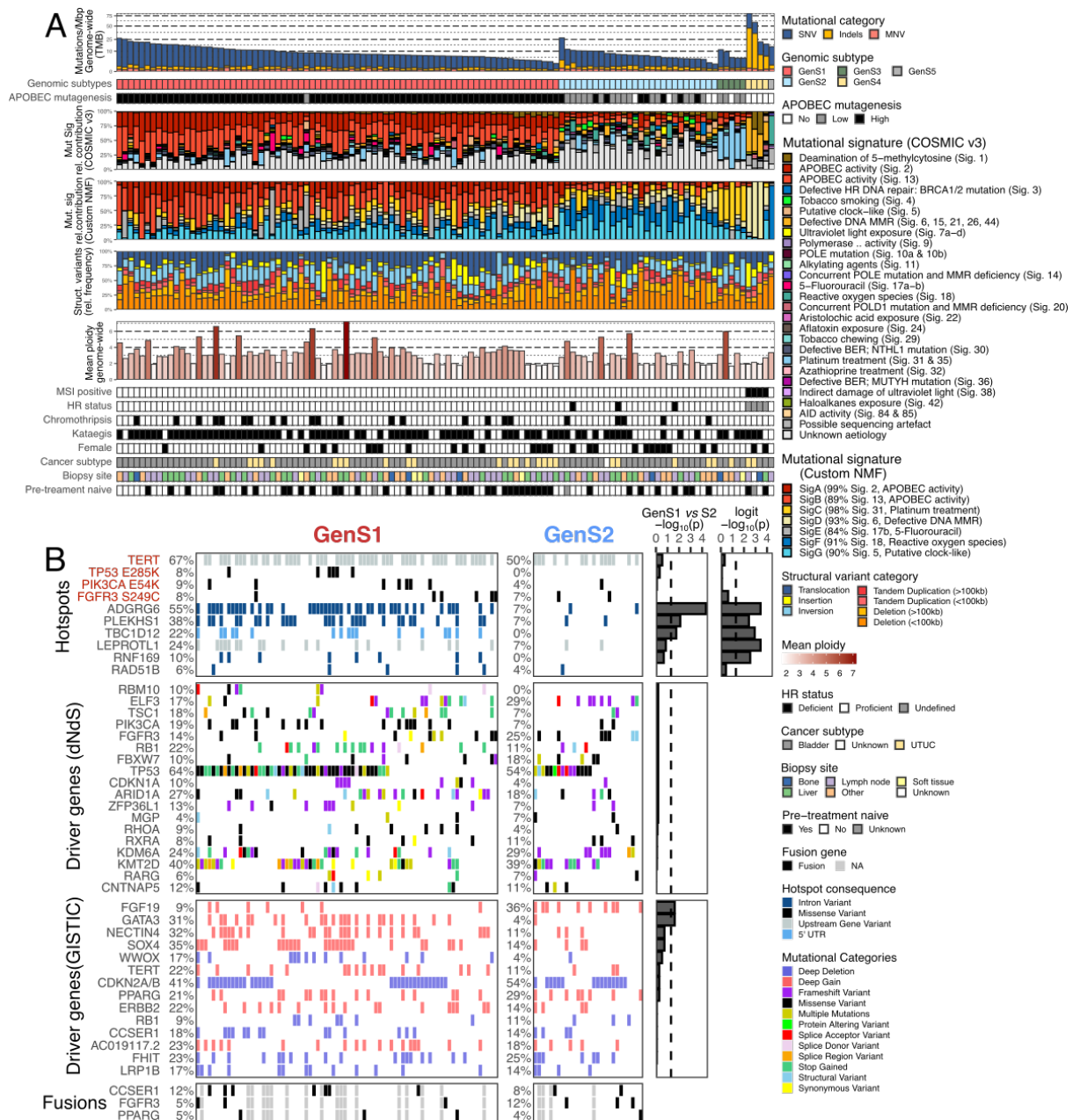
Figure 2 – Genomic landscape of metastatic urothelial carcinoma stratified by genomic subtypes

(A) Whole-genome sequencing data from biopsy samples of metastatic urothelial carcinoma were classified in genomic subtypes by hierarchical consensus clustering of the relative contribution of mutational signatures COSMIC v3 (9) grouped by etiology. The genomic features are displayed from top to bottom as follows: Genome-wide TMB; Genomic subtype (GenS1-5); APOBEC enrichment analysis showing tumors with no-, low- and high-APOBEC mutagenesis; mutational signatures grouped by etiology, except for APOBEC activity for which both signatures are shown separately; Relative contribution of seven *de novo* (custom) mutational signatures by deconvolution of SNVs in 96

tri-nucleotide context with NMF; Relative frequency of different types of structural variants; Mean ploidy; Tumors with MSI; HR deficiency status; Samples with at least one chromothripsis event; Samples with at least one kataegis event; Female patients; Primary origin of metastatic sample; Site of biopsy; Treatment-naïve patients.

(B) Overview of recurrent hotspot mutations, driver genes and gene fusions for the genomic subtypes GenS1 and GenS2. Name of genes affected by hotspot mutations in >5% of samples are displayed in red when the hotspot has a COSMIC id. Significantly mutated genes were estimated by dNdScv (10); all genes with q < 0.05 were considered driver genes. Recurrent focal copy number changes were estimated by GISTIC2 (11); genes in genomic regions with q < 0.05 were considered significant. Only affected genes present in >10% of the samples are shown. Gene fusions were detected from RNAseq data. Benjamini-Hochberg adjusted p-values of Fisher's exact test (for hotspot mutations and GISTIC2) and of logistic regression analysis corrected by mutational load (driver genes by dNdScv) are shown on the right to reflect the significance on the difference between GenS1 and GenS2. An additional logistic (logit) regression analysis was performed on hotspot mutations to show the linear relation with the number of APOBEC associated mutations. Bars beyond the dashed line ($-\log_{10}(0.05)$) are statistically significant. TMB = tumor mutational burden; Mbp = mutations per mega base pair; NMF = non-Negative Matrix Factorization; MSI = microsatellite instability; HR = Homologous Recombination; UTUC = upper tract urothelial carcinoma.

Structural variants (SVs) were common with a median of 259 (40,297 in total) per tumor. Deletion was the most frequent type of SV with a median of 92 per tumor (Figure S1D). Chromothripsis, a complex event that produces SVs in which chromosomes are shattered and rearranged, was detected in 20% of the tumors (Data S1.4). The genes most frequently affected by SVs were *CCSER1* (13%) and *AHR* (12%; Data S1.3).

Chromosomal arm and focal copy number alterations (CNA) were analyzed with GISTIC2. This revealed frequent deletion of chromosome 9 and amplification of chromosome 20 (Data S1.5A). In total, 49 genomic regions were significantly altered by focal CNAs which included several oncogenic genes (Table S1.8). The most frequently amplified genes were *SOX4* (28%), *GATA3* (22%), *PPARG* (22%), and ERBB2 (19%); the most frequently deleted genes were *CDKN2A/B* (43%), *FHIT* (24%), *CCSER1* (17%) and *LRP1B* (17%; Data S1.3).

7

Hotspot mutations in the *TERT* promoter were present in 62% of the tumors (Data S1.3; Table S1.11). Still, *TERT* expression did not differ between tumors with and without hotspot mutations (Figure S2B), in line with a previous report (12). However, differential gene expression analysis showed that tumors with hotspot mutations in the *TERT* promoter had downregulation of genes related to the muscle contraction pathway (Figure S2A, Table S1.12). Hotspot mutations in other driver genes concerned *FGFR3* S249C (8%), *PIK3CA* E54K (7%), and *TP53* E285K (5%). Furthermore, hotspot mutations were identified in non-coding regions of *ADGRG6 (40%)*, *PLEKHS1 (28%)*, *LEPROTL1* (18%), and *TBC1D12* (15%; Data S1.3) with no apparent association with gene expression and minimal transcriptomic effect (Figure S2A-B). The hotspot areas of *ADGRG6*, *PLEKHS1* and *TBC1D12* form hairpin loop structures in the DNA with specific tri-nucleotide sequences frequently mutated by APOBEC enzymes (Figure S2C). Unlike other known driver genes affected by hotspot mutations (*TERT*, *FGFR3*, *PIK3CA* and *TP53*), these genes were not significantly affected by other somatic mutations in the coding region or by CNAs, suggesting that hotspot mutations in *ADGRG6*, *PLEKHS1* and *TBC1D12* are likely passenger hotspots attributed to APOBEC activity as theoretically predicted (13).

Fusion gene analysis performed at the transcriptomic level (Table S1.14) detected 1394 fusion genes, of which 10% were also reported in the TCGA cohort (14). Seventy-six percent of all individual genes found involved in fusion events have previously been implicated in fusions (14). *FGFR3* gene fusions were present in seven out of 90 samples with only one *FGFR3-TACC3* fusion. Four *PPARG* fusions were detected, of which two *PPARG-TSEN2* fusions were confirmed at DNA level (Table S1.13). Other putative fusion events in cancer-related genes were found in *CCSER1* (n = 9), *ERBB4* (n = 5), *RB1* (n = 4), *MDM2* (n = 4), *TERT* (n = 3) and *STAG2* (n = 3).

A stratification based on the proposed etiology of SNV COSMIC signatures (Table S1.15) using unsupervised consensus clustering (1) revealed two major genomic subtypes (GenS; Figure 2; Data S1.6). Subtype GenS1 (67%) contained almost exclusively tumors with high APOBEC mutagenesis, which was reflected by a large contribution of APOBEC signatures 2 and 13. In addition, we performed deconvolution of SNV patterns by non-negative matrix factorization (NMF; Data S1.7), which confirmed APOBEC signatures as the main source of mutations in GenS1, with high contribution of *de novo* mutational signatures SigA (0.99 cosine similarity with APOBEC signature 2) and SigB (0.89 cosine similarity with APOBEC signature 13). Subtype GenS2 (24%) aggregated predominantly tumors with low APOBEC mutagenesis (16 out of 28), and was characterized by signatures of unknown etiology. *De novo* mutational signatures SigF (0.91 cosine similarity with SBS18 COSMIC

signature) and SigG (0.90 cosine similarity with SBS5 COSMIC signature) were dominant in GenS2. Analysis of the TCGA cohort (WES data) showed that GenS1 and GenS2 were also the two major genomic subtypes in localized UC (Figure S3A-B). The other three smaller subtypes (9% of the present cohort) were related to the platinum treatment signature (GenS3), which was not detected in the treatment naïve TCGA cohort, the defective DNA mismatch repair (MMR) signature and microsatellite instability (MSI, GenS4), and the reactive oxygen species signature (GenS5).

The origin of somatic driver mutations was independent of the genomic subtypes, although amplifications of *GATA3* and *FGF19* were enriched in GenS1 and GenS2, respectively (Figure 2B). Hotspot mutations occurred more frequently in GenS1. In particular, *ADGRG6*, *PLEKHS1* and *TBC1D12* were significantly more often mutated in GenS1. However, these hotspot mutations are potentially irrelevant byproducts caused by APOBEC mutagenesis as logistic regression analysis showed a correlation between APOBEC mutational load (C>T and C>G mutations in TCW context, W = A or T) and occurrence of these hotspot mutations (Figure 2B).

Other genomic differences between GenS1 and GenS2 (Data S1.2-3 and Data S1.8) included higher SNVs/Mbp in GenS1 and higher Indels/Mbp in GenS2, which pattern was also observed in the TCGA cohort (Figure S3C). All three tumors with homologous recombination (HR) deficiency identified were of subtype GenS2. Clinical characteristics such as sex, cancer subtype (bladder or upper tract UC), and pre-treatment status did not differ between GenS1 and GenS2. Thus, despite that two very different etiologies lead to UC development, the two mutagenic processes lead to similar profiles of somatically affected driver genes.

APOBEC mutagenesis is an active process that generates new mutations in mUC

APOBEC enzyme expression analysis revealed neither significant differences between GenS1 and GenS2 (Figure S4A), nor between tumors with and without APOBEC mutagenesis (Figure 3A). To investigate whether this outcome could be attributed to inactivation of APOBEC mutagenesis in mUC, we analyzed WGS data of eight tumors from patients who had undergone serial biopsies, and reconstructed their evolutionary paths (Figure 3C). The cancer cell fraction was lower in the branches of the evolutionary trees than in the trunk, suggesting that these mutations were novel and not widely spread in the cancer cell population (Figure 4D). The rate of novel APOBEC mutations (number of APOBEC mutations divided by the number of days elapsed between serial biopsies) was calculated using only mutations from branches corresponding to the second biopsy. Tumors with

high APOBEC mutagenesis accumulated more novel APOBEC mutations than other tumors (Figure 3F, Wilcoxon rank-sum test p = 0.036), matching with the presence of APOBEC signatures (Figure 3E).

Several studies have linked the expression of APOBEC3A/3B to APOBEC mutagenesis in UC (4, 15). When comparing the estimated relative activity of APOBEC3A and 3B in tumors with APOBEC mutagenesis, we observed differential expression of these enzymes (Figure 3G). Some tumors had high levels of *APOBEC3A* expression while the expression of *APOBEC3B* was low – or *vice versa*. This explains the lack of differential expression of APOBEC enzymes between tumors with different levels of APOBEC mutagenesis and suggests that APOBEC mutagenesis in some tumors may be triggered either by APOBEC3A or by APOBEC3B. Taking this into account, we combined the activity of both enzymes and calculated an APOBEC score (sum of *APOBEC3A* and *3B* expression). It appeared that that tumors with high APOBEC mutagenesis had a higher APOBEC score than other tumors (Wilcoxon rank-sum test p = 0.012; Figure 3B). APOBEC score was also high in GenS1 (Figure S4B). This confirmed the link between APOBEC (3A and 3B) enzyme expression and APOBEC mutations in mUC, as previously shown for MIBC by Robertson *et al.*

In tumors with high APOBEC mutagenesis, the mean ploidy and the number of genes affected by CNA were higher than in tumors without APOBEC mutagenesis (Wilcoxon rank-sum test p = 0.01 and p <0.001, respectively; Figure S4C-D). This phenomenon may indicate ongoing genomic instability in APOBEC-driven tumors.

10
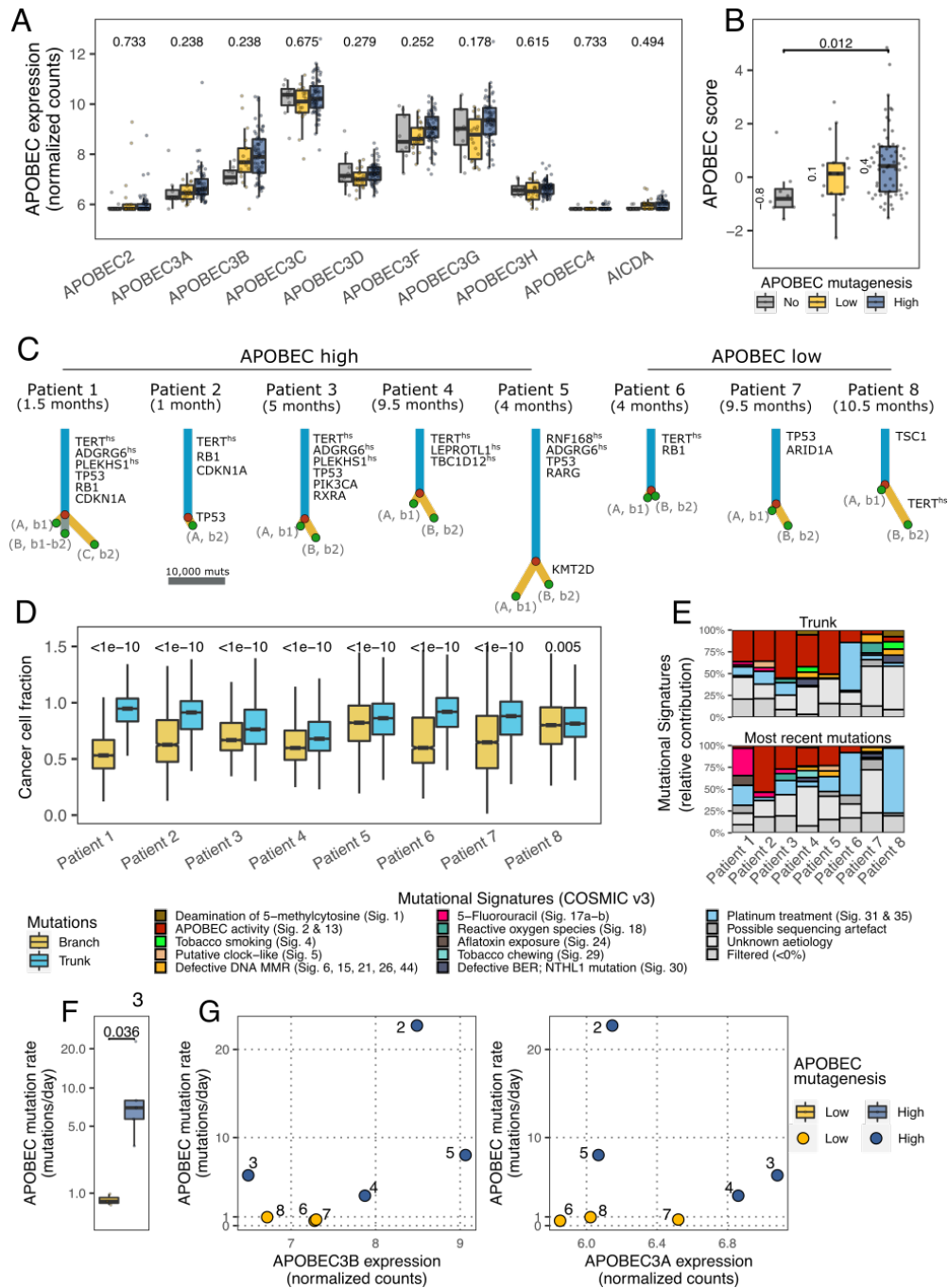
Figure 3 – APOBEC expression and cancer evolution of eight tumors from metastatic urothelial carcinoma patients with serial biopsies

(A) RNA-expression of different APOBEC enzymes compared between APOBEC and non-APOBEC mediated mutagenesis tumors. Kruskal-Wallis test p-values were Benjamini-Hochberg corrected.

(B) APOBEC score (sum of *APOBEC3A* and *APOBEC3B* expression) across groups of tumors with distinct level of enriched APOBEC mutagenesis. Wilcoxon rank-sum test p-value is shown for non-APOBEC and high APOBEC mediated mutagenesis tumors.

11

(C) Evolutionary trees from eight tumors (five with high APOBEC and three with low APOBEC mutagenesis) with two biopsies were reconstructed from single nucleotide variants. Mutations in driver genes (dNdScv) and hotspot sites (hs) are shown, and their locations are indicated (trunk or branch). Branches represent subclonal populations (A, B or C), indicating their presence in the first or second biopsy (b1 or b2). For patient 1, a subclonal population is present in both biopsies. The cancer cell fraction of each single nucleotide variant was calculated and clustered using DPClust for paired-biopsies. The evolutionary tree was reconstructed using the *sum rule* (17).

(D) Comparing the cancer cell fraction of somatic mutations from the trunk and branches. Wilcoxon rank-sum test was applied and p-values were Benjamini-Hochberg corrected.

(E) COSMIC v3 mutational signatures calculated from the trunk and from the branch exclusive to the second biopsy.

(F) The APOBEC mutation rate from novel (recent) mutations in the second biopsy was compared between low and high APOBEC mutagenesis tumors. Wilcoxon rank-sum test was applied.

(G) APOBEC mutation rate is displayed as a function of *APOBEC3A* and *APOBEC3B* expression. APOBEC expression was estimated as the mean expression of both biopsies per tumor. Numbers indicate patient number. RNAseq was not available for patient 1.

APOBEC associated mutations are randomly distributed across the genome in mUC

The substrate of APOBEC enzymes is single-stranded DNA (ssDNA; Roberts *et al.*, 2013), this has led to the hypothesis that APOBEC enzymes are mainly active during replication or in open chromatin and transcriptionally active genomic regions (18, 19). As our cohort contained primarily tumors with APOBEC mutagenesis, and WGS data of these tumors was available, we had the unique opportunity to explore the consequences of APOBEC enzymatic activity across the genome.

The total number of SNVs/Mbp varied across the genome, and non-APOBEC mutations followed the same pattern (Figure 4A). The frequency of non-APOBEC mutations decreased as the predicted DNA accessibility and overall gene expression level increased (Figure 4B). In contrast, the frequency of APOBEC mutations was constant across the genome, demonstrating that APOBEC mutagenesis was likely independent of genomic

12

regions (Figure 4A-B and Data S1.9A). The flat distribution of APOBEC mutations supported the hypothesis that these mutations had been generated during replication, when APOBEC enzymes have equal access to ssDNA across the genome (19).
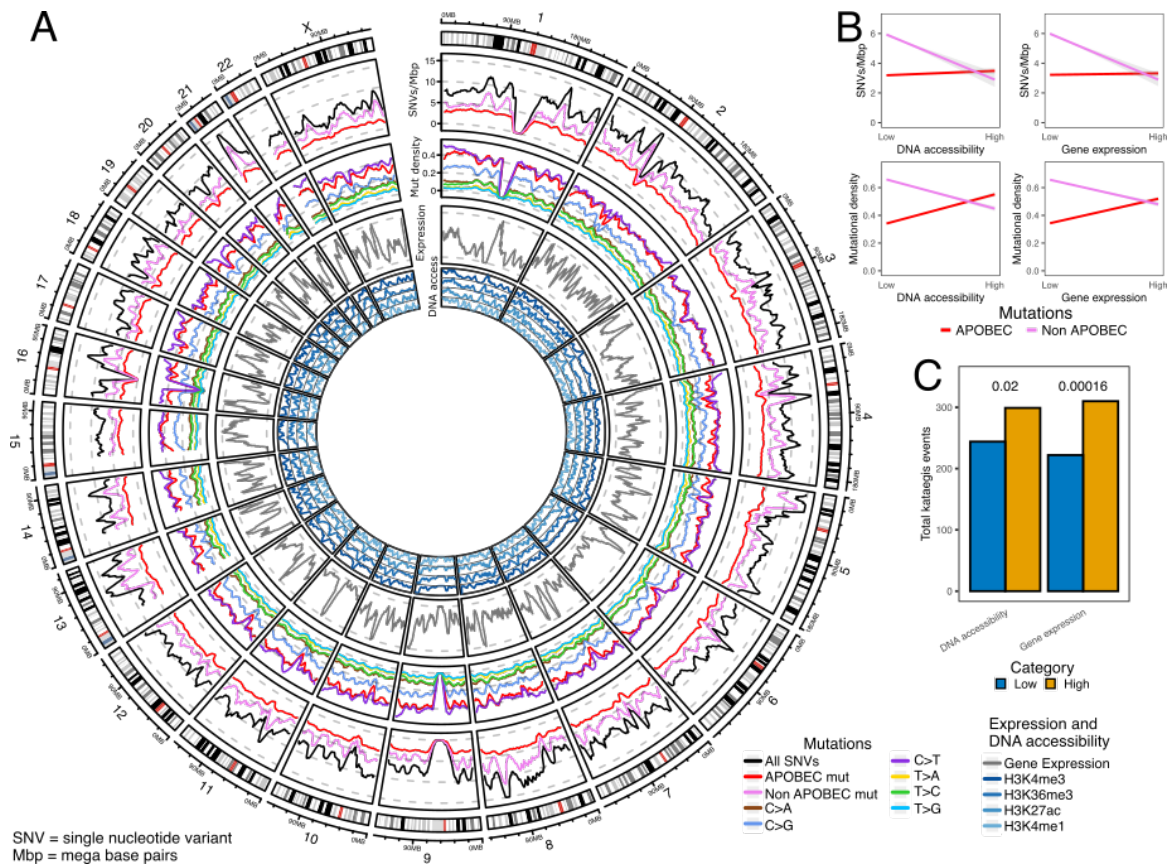
Figure 4 – Differences in the load of APOBEC associated mutations between high and low DNA accessibility regions in metastatic urothelial carcinoma genomes

(A) WGS data was analyzed to estimate the mean number of single nucleotide variants in windows of one mega base pairs across the entire genome. The Circos plot shows from outer to inner circles: the genomics ideogram from chromosome 1 to X where the centrosomes are indicated in red; Mutational load of APOBEC and non-APOBEC associated mutations; Relative contribution (mutational density) of the 6 single nucleotide variant types are displayed. The frequency of APOBEC associated mutations relative to the total is shown as well; Average RNA counts (expression) from 90 tumors with RNAseq

13

data; DNA accessibility estimation from different ChIPseq experiments in normal urothelial samples derived from the ENCODE (20). Peaks represent highly accessible DNA.

(B) Linear regression (with 95% confidence interval) of mutational load per mega base pairs (absolute) and mutational density (relative) for APOBEC and non-APOBEC associated mutations with DNA accessibility and expression data.

(C) Frequency of kataegis events (full cohort) in high and low DNA accessibility or in high and low gene expression regions. P-values of binomial test are shown for each comparison.

Localized hypermutation events (kataegis) were present in 70% of the samples (Figure 2A); they were more frequent when APOBEC mutagenesis was high (Data S1.9B). This higher frequency confirmed a link between kataegis and APOBEC activity (21), and we therefore expected to find kataegis events scattered across the genome. However, our data suggested that kataegis was more likely to happen in regions with high DNA accessibility and high transcriptional activity (Figure 4C). Thus, while general APOBEC mutagenesis seemed to occur primarily during replication, kataegis-like APOBEC events seemed to occur during both replication and at transcribed loci.

In summary, APOBEC mutagenesis was an ongoing process in mUC that equally affected the whole genome. In some tumors, it seemed to be triggered through activation of either APOBEC3A or APOBEC3B. Tumors with APOBEC mutagenesis were genomically less stable and displayed more kataegis events.

Transcriptomic subtypes of mUC

The consensus classifier of primary MIBC stratifies organ-confined UC of the bladder into one of six transcriptomic subtypes (6). Unlike primary bladder tumor samples, metastatic biopsies are derived from different organs with some degree of normal non-urothelial cell contamination. Using the consensus classifier would lead to misclassification of samples when no correction for organ-specific transcripts is applied. It would also limit the number of phenotypic subtypes that we can identify in mUC, which is crucial as the transcriptomic phenotypes of mUC are unknown. Therefore, it was mandatory to perform *de novo* subtyping for mUC samples.

14

High-quality RNAseq data was available for 90 (97 samples) out of 116 patients (Data S1.10). To reduce the bias introduced by biopsy location, we filtered the organ-specific transcripts prior to hierarchical consensus clustering (Figure S5). Five transcriptomic subtypes were identified (Figure 5). Several phenotypic markers were used to establish the phenotype of each subtype (Figure S6A).
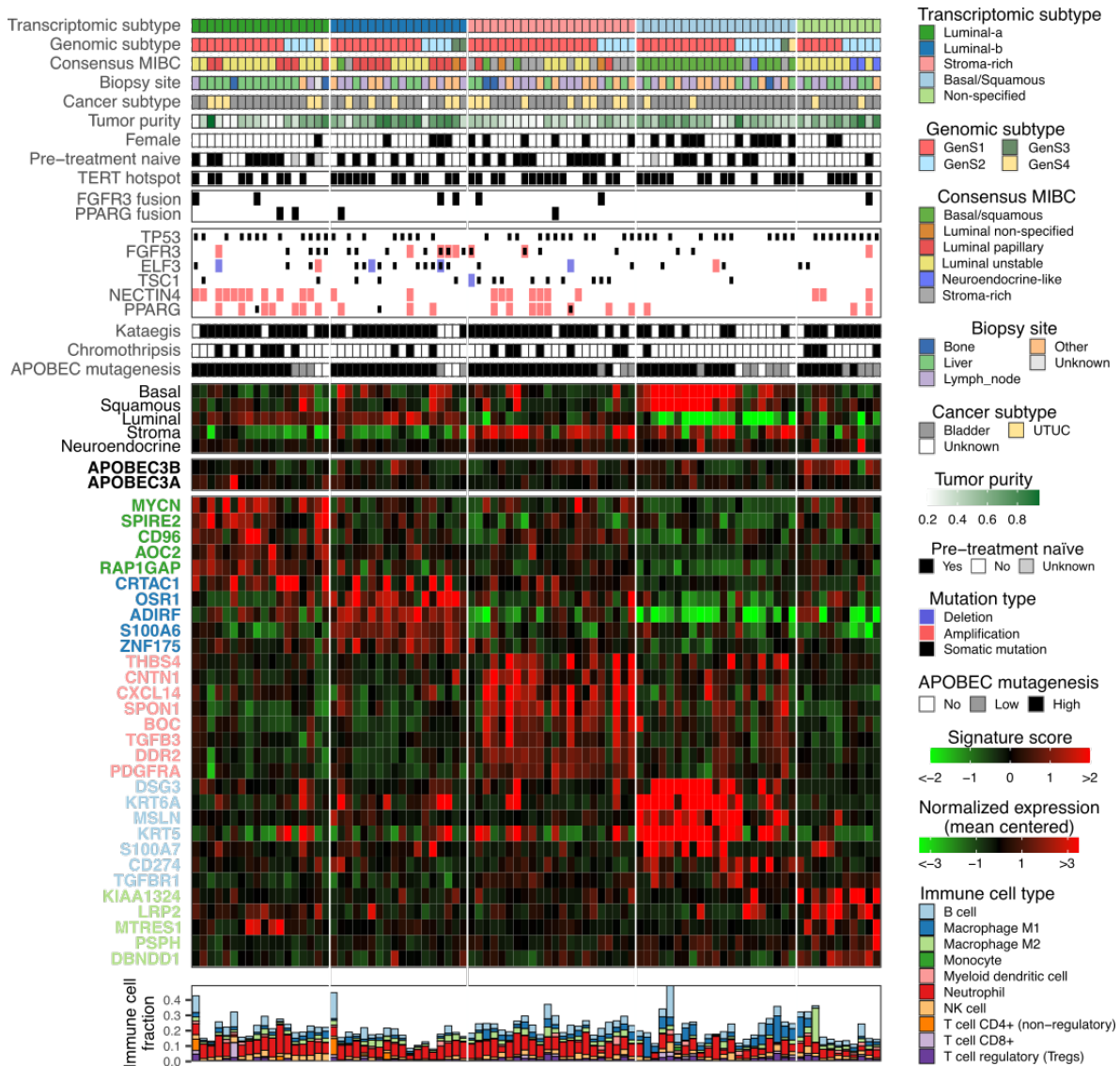
Figure 5 – **Genomic and transcriptomic characteristics of patients with metastatic urothelial carcinoma stratified by mRNA subtypes**

Transcriptomic profiles of 90 metastatic urothelial carcinoma patients were clustered using ConsensusClusterPlus (1). Five transcriptomic subtypes were identified: luminal-a, luminal-b, stroma, basal/squamous and non-specified phenotype. From top to bottom: Transcriptomic subtypes;

Genomic subtypes (GenS1-4); Transcriptional subtypes based on the consensus MIBC classifier (6); Site of sampled biopsy; Subtype of primary tumor (UTUC = upper tract urinary cancer); Estimated tumor cell percentage; Female patients; Pre-treatment naïve patients; Tumors with hotspot mutations in *TERT* promoter; Tumors with fusion genes detected with RNAseq data; Samples with alterations in selected genes; Tumors with one or more kataegis events; Tumors with one or more chromothripsis events; APOBEC mutagenesis enrichment; *APOBEC3B* and *APOBEC3A* expression; Signature score (mean expression of genes related to each phenotype) of basal, squamous, luminal, stroma and neuroendocrine markers; Top overexpressed genes in each mRNA subtype; Immune cell fractions estimated with immunedeconv (22), using the quanTIseq method (23).

The luminal subtypes, which are dominant in NMIBC (> 90% of the tumors, Lindskrog *et al.*, 2020), represented 51% of the tumors in the TCGA cohort *versus* 40% in the present cohort (p = 0.061, Figure S3D). Although the two subtypes detected in the present cohort were of clear luminal origin (Figure S6A), they did not reflect the luminal subtypes specified in the consensus MIBC clustering (Figure 5, top rows). The luminal-a subtype had high expression of *MYCN*, one of the MYC oncogene family members that regulates different species of RNA (24), and high expression of *CD96* (Figure 5). This subtype had low tumor purity, a high fraction of NK cells, a low clonal fraction (interpreted as high heterogeneity), and relatively high expression of *FGFR3*, *PPARG* and *NECTIN4* (Figure S6). *NECTIN4* was amplified in 61% of these tumors. The luminal-b subtype had high tumor purity, a low number of SVs, a low fraction of NK cells, a high expression of *FGFR3*, *PPARG,* and *S100A6* (Figure S6), and a higher proportion of *ELF3* (56%) and *FGFR3* (50%) DNA alterations compared to all other subtypes (Fisher's exact test p = 0.0023 and p = 0.0053, respectively).

In the stroma-rich subtype, genes known to be associated with stromal content and cancer-associated fibroblasts (*THBS4, CNTN1, CXCL14* and *BOC;* Mathew *et al.*, 2014; Kuroda *et al.*, 2019; Zhang *et al.*, 2019; Gu *et al.*, 2020) were differentially expressed (Figure 5). This subtype was highly concordant with the stratification of the consensus MIBC clustering (79% of tumors identified as stroma-rich by the consensus MIBC were in this group). However, the stroma-rich subtype was more prevalent in the present cohort than in the TCGA cohort (24% *vs* 9%, Fisher's exact test p < 0.001, Figure S3D). Tumors of the stroma-rich subtype showed high expression of *TGFB3*, a ligand of the TGF-β pathway (Figure 5), low tumor purity, a high signature score for

epithelial to mesenchymal transition (Figure S6), high expression of various collagens (Table S1.22), and a higher rate of *TSC1* DNA alterations (45% of the tumors, Figure 5) compared to the rest of the cohort (Fisher's exact test p < 0.001).

The basal/squamous subtype was also highly concordant to the similarly named MIBC consensus cluster (86% of tumors identified as basal/squamous by the consensus MIBC were in this group). Yet, the prevalence of this subtype was lower in the present cohort than in the TCGA cohort (23% *vs* 38%, Fisher's exact test p = 0.013; Figure S3D). This subtype was characterized by high expression of basal/squamous markers (*DSG3*, *KRT5, KRT6A* and *S100A7*) and enrichment in females (52%, Fisher's exact test p = 0.0043). *TGFBR1* (a receptor of the TGF-β pathway), *CD274* (the gene that encodes PD-L1) and *MSLN* (a tumor-associated antigen) were highly expressed in this subtype (Figure 5). *NECTIN4* amplifications were not found, and the *NECTIN4* expression level was low (Figure S6D). In line with a previous study (29), the expression of adipogenesis regulatory factor (*ADIRF*) was low (Figure 5). The immune cell compartment consisted of a large fraction of M1 macrophages and a low fraction of neutrophils (Figure S6B). This subtype was less affected by kataegis and chromothripsis events than were the other subtypes (Figure 5, Fisher's exact test p = 0.0006 and p = 0.019, respectively).

The non-specified subtype did not reflect any of the phenotypic markers associated to phenotypes of the consensus MIBC subtypes. This subtype had overexpression of *KIAA1324* (a diagnostic biomarker in different cancer subtypes; 30) and *LRP2* (Figure 5), and a high score of claudin-low markers, a low fraction of neutrophils, high numbers of Indels/Mbp, high numbers of SVs, and high levels of *APOBEC3B* expression (Figure S6). This subtype was enriched for patients with a pre-treatment history (Fisher's exact test p = 0.023, Figure 5).

In summary, transcriptomic profiling revealed that mUC can be stratified into five transcriptomic subtypes, of which the luminal-a and b, stroma-rich and basal/squamous subtypes are concordant to primary MIBC subtypes. The phenotype of the non-specified subtype appeared to be unique for mUC as it did not match any of the phenotypes of the consensus subtypes established for MIBC. A complete overview of driver genes, gene fusions and hotspot mutations per transcriptomic subtype is presented in Figure S7.

Altered canonical signaling pathways in different transcriptomic subtypes

Several canonical pathways involved in cell growth, proliferation and survival (33) were altered at the DNA level (Figure 6A). Of all subtypes, the luminal-a subtype showed most alterations in the Myc (72%) and TGF-β (72%) pathways (e.g. only 5% of basal/squamous tumors had TGF-β alterations). Perturbations in the TGF-β pathway were mainly driven by alterations in *TGFBR2*, *SMAD4*, *SMAD2* and *TGFBR1*. The most altered genes per pathway are displayed in Data S1.11.
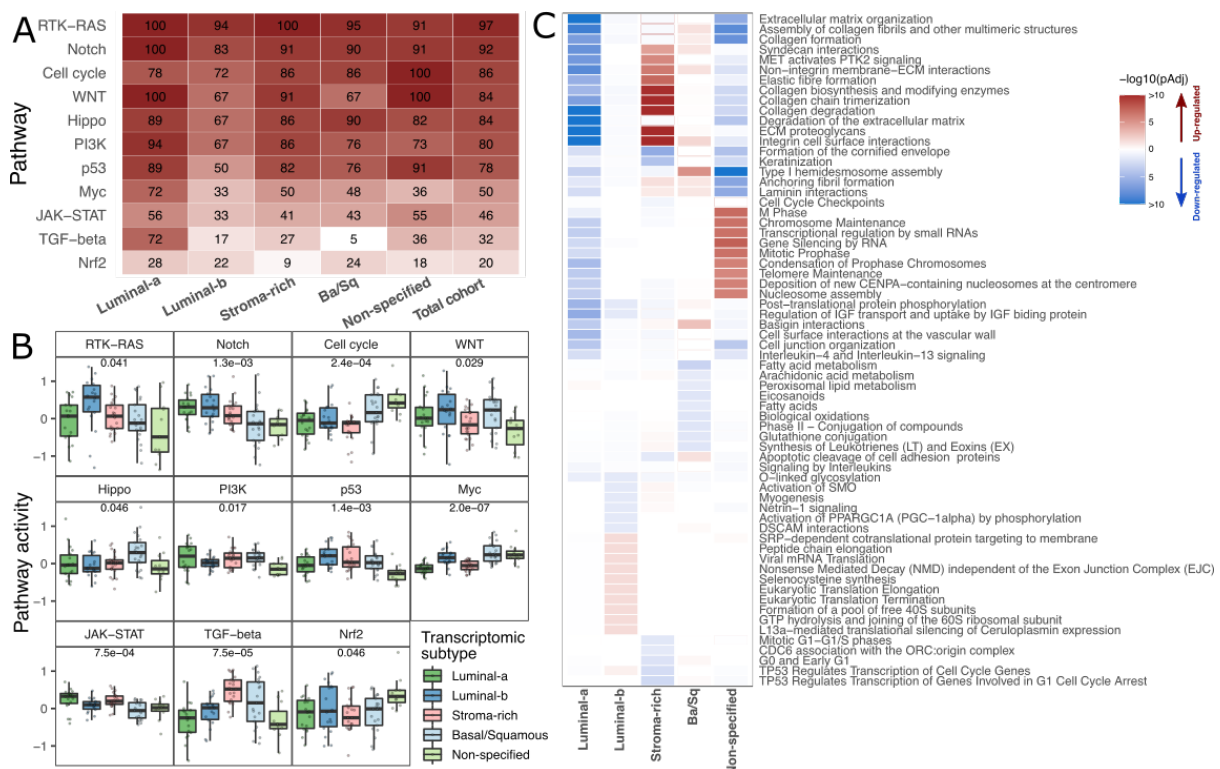
Figure 6 – Pathway alterations at genomic and transcriptomic level across mRNA-based subtypes of metastatic urothelial carcinoma

(A) The percentage of samples with DNA alterations in 11 canonical pathways is shown for each transcriptomic subtype (90 tumors in total) and for the entire cohort (n = 116). A patient was considered to have an altered pathway when at least one of the pathway-genes was altered either by non-synonymous mutations, structural variants or by deep copy number changes.

(B) Pathway activity was estimated as the mean expression of downstream genes targeted by each pathway. Only genes that were transcriptionally activated by these pathways were considered. Kruskal-Wallis test p-values were Benjamini-Hochberg corrected.

(C) Pathways up- (red) or down-regulated (blue) were estimated with reactomePA (34) from RNAseq data. Only the top ten up- and top ten down-regulated pathways per subtype are shown.

The luminal-b subtype was characterized by fewer alterations in Notch, Cell cycle, Hippo, PI3K, p53, Myc and JAK-STAT pathways. Alterations in the p53 pathway were common in the other subtypes, and most of them were the result of somatic mutations in *TP53* or amplification of *MDM2* (mutually exclusive p = 0.024). Amplification of *MDM2* has been previously reported in a pan-cancer study (33) as an alternative to *TP53* alterations to inactivate the p53 pathway through direct inhibition of p53 protein (35).

To assess the effect of genomic alterations on pathway activity, we calculated the mean expression of genes targeted by each pathway as *a proxy* of activity (Figure 6B). Myc and TGF-β pathway activities were low in the luminal-a subtype, corresponding with high frequencies of pathway alterations at the genomic level (Figure 6A). The luminal-b subtype showed the highest RTK-RAS and high WNT pathway activity. The stroma-rich subtype had high TGF-β pathway activity; the basal/squamous subtype had high activity of the Hippo, Myc and TGF-β pathways. The non-specified subtype had very low p53 pathway activity and very active cell cycle pathway signaling, two pathways that are usually co-altered (33).

In addition to the 11 oncogenic pathways described above, any pathway up- or down-regulated was analyzed by enriched pathway analysis with ReactomePA (Figure 6C). Up-regulation of pathways involved in collagen metabolism and extracellular matrix in the stroma-rich subtype corresponded with the stromal phenotype of this subtype (36). In the non-specified subtype, pathways related to cell cycle and chromosome integrity were up-regulated. Considering as well the high cell cycle activity (Figure 6B) and high frequency of mutations in the cell cycle pathway (Figure 6A), this up-regulation may suggest that the non-specified subtype is highly proliferative.

In summary, signaling pathway analysis showed the extent of heterogeneity between the transcriptomic subtypes, reflecting phenotypic characteristics of each group. The most striking difference was observed for

the TGF-β pathway, in which genomic alterations greatly affected the luminal-a subtype – with consequences

on pathway activity.

Discussion

We defined, for the first time, molecular subtypes of mUC based on whole genome and transcriptome characteristics through WGS and RNAseq analysis of metastatic biopsies of 116 mUC patients. We found three novel driver genes specific for mUC and, in line with findings for primary UC, we identified a central role of APOBEC mutagenesis. Furthermore, we concluded that mUC is a heterogeneous disease with various genomic and transcriptomic subtypes revealing the main mutational processes and phenotypes of this cancer.

The genomic landscape of mUC showed important similarities to that of primary UC. We validated our mUC findings in the TCGA cohort of primary UC, showing that aggregating mutational signatures by etiology is a robust approach to identify genomic subtypes. A recent study analyzed archived paraffin-embedded primary or metastatic tumor samples from MIBC patients who received palliative chemotherapy. By WES analysis, two major genomic subtypes were identified (37). The GenS2 subtype (enriched with SigG that correlates with COSMIC SBS5) in the present study largely overlapped with the SBS5 subtype reported by Taber *et al.*, 2020. Furthermore, they identified an APOBEC high signature that was similar to GenS1 in our study.

We identified *CNTNAP5*, *RARG* and *MGP* as exclusive driver genes of mUC. A pan-cancer study of the Hartwig Medical Foundation cohort previously identified driver mutations in *RARG* and *MGP* in mUC tumors (38). Only 3% of the samples in the TCGA localized UC cohort had *CNTNAP5* or *RARG* mutations, and *MGP* mutations were not detected, suggesting that these genes might be related to metastatic progression. *CNTNAP5* is a member of the contactin-associated protein family, which is involved in cell contacts and communication in the nervous system (39). Mutations in this gene have been associated to autism but not to cancer (40). Genomic alterations in the retinoic acid receptor gamma (*RARG*) gene have not been associated to cancer, except for fusion genes which are well-documented for leukemia (41). The matrix Gla protein (*MGP*) is misregulated in several cancers, and this misregulation is associated with tumorigenesis (42). The exact role of this gene in cancer is unknown, although several recent studies suggest that oncogenic signaling pathways are altered through *MGP* (43, 44). Other affected driver genes we found in the present study were similar to those found in primary UC. Frequent hotspot mutations in the non-coding region of *TERT*, *ADGRG6*, *PLEKHS1*, *LEPROTL1* and *TBC1D12* occurred similarly in NMIBC and MIBC (45, 46). Evidence suggests that clones with known driver genes emerge early during bladder cancer development and colonize distant areas of the bladder, which may explain the genomic similarity of mUC with primary UC (47).

WGS analysis revealed frequent SVs affecting *AHR* (aryl hydrocarbon receptor) and *CCSER1* (coiled-coil serine rich protein). SVs in *AHR* have not been described in cancer, but other molecular alterations in this gene have been associated with bladder cancer progression (48–50). *CCSER1* is located in a common fragile site region; thus, it is exposed to chromosomal rearrangements (51). Altered transcripts created through the deletion of specific exons in *CCSER1* have been associated with oncogenesis (51, 52); it is unclear, however, if SVs may have similar oncogenic effects in UC.

Previous studies that performed RNA-based subtyping showed that NMIBC is a homogeneous disease primarily of luminal origin (> 90%) and that MIBC is highly heterogeneous with multiple subtypes (3, 6). Here, we showed that mUC is a heterogeneous disease too, with similar subtypes as described for MIBC. The phenotypic similarity of MIBC and mUC suggests that despite ongoing mutagenesis, UC cell behavior does not change significantly during the metastatic process – all subtypes have metastatic potential. Furthermore, some patients with primary non-metastatic MIBC, as assessed by cross-sectional imaging, actually have systemic rather than localized disease. In a previous study, lymph node metastases were present in the cystectomy resection specimen of 25% of clinically node-negative patients. Due to the development of metastatic disease, patients with locally advanced bladder tumors (pT3 tumors) have a poor 5-year overall survival rate of only 35%, despite radical surgery (Stein *et al.*, 2001, Mari *et al.*, 2018).

In the present cohort, we did not identify a NE-like subtype at transcriptional level. The prevalence of this subtype in UC is low; in the TCGA cohort, only 2% of tumors were of the NE-like subtype. Central pathology revision of the metastatic biopsies by an expert genitourinary pathologist identified only three NE-like tumors in our cohort (Table S.3-4). The non-specified subtype we identified did not express any of the markers used to identify MIBC phenotypes (luminal, basal, squamous, stroma or NE), suggesting rewiring of its transcriptomic profile for adaptation. Studies in various cancers have shown that therapeutic pressure may trigger a phenotype-switching event (54), which could have happened in the non-specified phenotype as it was enriched for patients who had received systemic therapy prior to biopsy. Studies with larger numbers of paired biopsies (obtained before and after treatment) would be needed to explore this phenomenon in mUC.

APOBEC mutagenesis was widespread in mUC; the reconstruction of evolutionary paths from sequential biopsies of eight patients indicated that it was an ongoing process. This suggests that mUC is in continuous adaptation by generating novel mutations. A previous study indeed reported accumulating mutations in six

patients whose primary tumor and metachronous metastases were analyzed by WES (37). In our study, the accumulation of new mutations in the sequential biopsy specimen of one of eight patients led to the identification of new therapeutic targets (Data S1.12).

In a previous study (38), the genomic landscape of 85 (72 re-analyzed here) mUC patients was compared with that of other metastatic tumor types. This pan-cancer study concluded that mUC was characterized by high tumor mutational load (with no difference between mUC and primary UC), high CNAs, the highest number of driver genes among all cancer types analyzed, and actionable targets in 75% of the patients. In our study, we identified a potential targetable alteration in the genome of 98% of the patients (Figure 7A). In line with Priestley *et al.*, we found that 41% of patients could benefit from on-label therapies, and 63% from therapies approved by the US Food and Drug Administration for other tumor types. Additionally, we identified targets for therapies under investigation in clinical trials including basket trials in 109 of 116 patients. We identified four patients with MSI-high tumors that are potentially sensitive to immune checkpoint inhibitors (55). HR deficiency, observed in three patients, is a potential target for treatment with poly-ADP ribose polymerase inhibitors and/or double-stranded DNA break-inducing chemotherapy. At the RNA level, targetable *FGFR3* and *NTRK2* gene fusions were identified in eight patients.
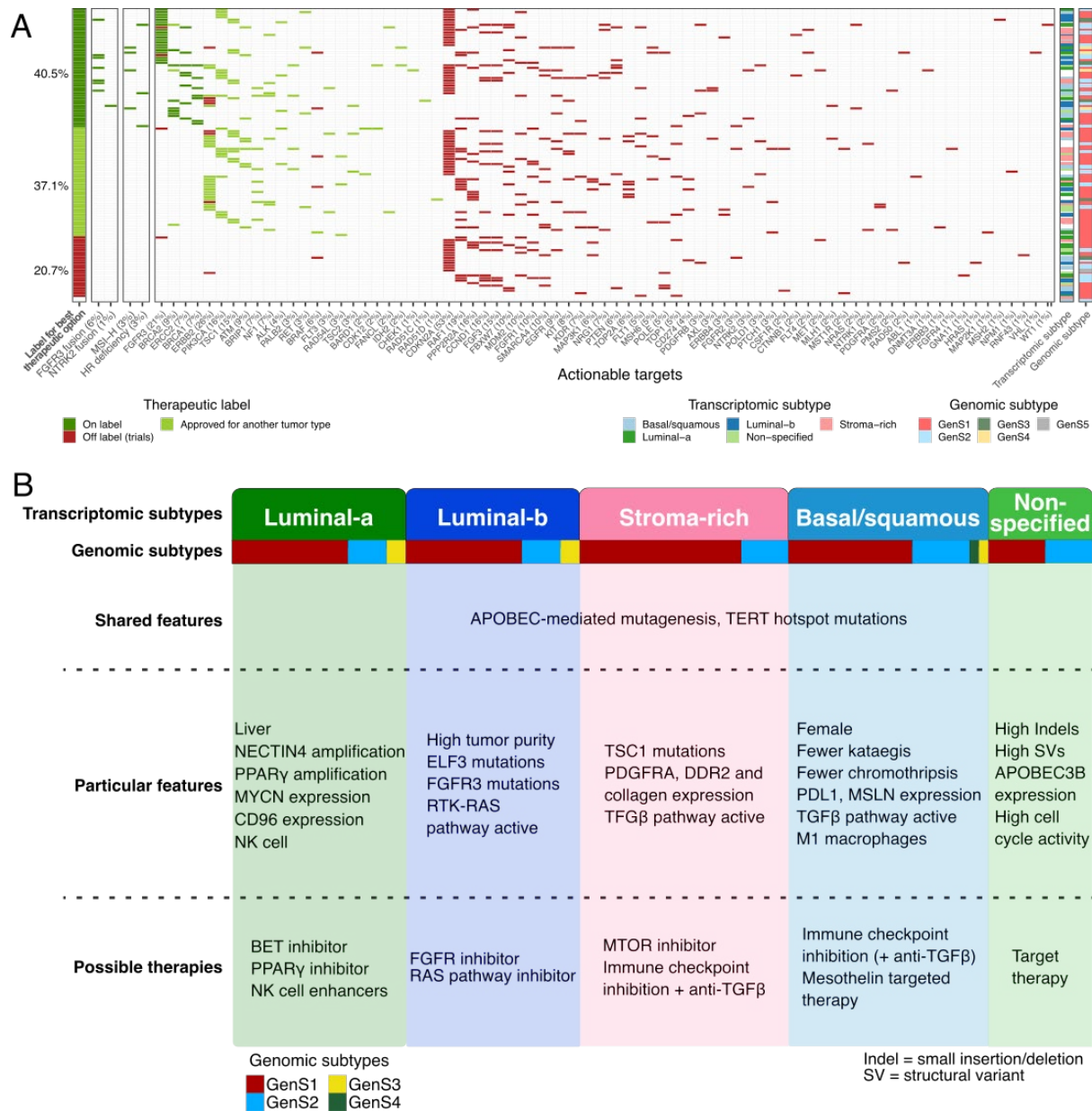
Figure 7 – Overview of actionable targets and possible treatments per transcriptomic subtype for metastatic urothelial carcinoma

(A) Per patient, overview of therapeutic targets based on gene fusions of RNA level, tumor with microsatellite instability or homologous recombination deficiency, and clinically-actionable genomic alterations for on- and off-label therapies of urothelial carcinoma. On the left, the most

effective treatment available per patient is shown. Bars on the right depict the genomic and the transcriptomic subtype per patient.

(B) Summary of molecular features found in the present study, and potential therapeutic implications for metastatic urothelial carcinoma per transcriptomic subtype.

In a previous study, the antibody-drug conjugate enfortumab vedotin targeting NECTIN4 induced objective clinical responses in 44% of mUC patients who experienced disease progression after chemotherapy and anti-PD1/L1 therapy (56). Currently preselection for this treatment is not required. However, we found significant variation in the expression of NECTIN4, suggesting that patients with the basal/squamous subtype may be less likely to experience clinical benefit, as they did not have *NECTIN4* amplifications, and the lowest *NECTIN4* expression level. Thus, subtype-specific treatment with enfortumab vedotin might result in better risk-benefit ratios. The 23 patients with *HER2* aberrations may be sensitive to HER2 targeting agents; especially some of the newer antibody-drug conjugates with DNA damaging payloads could represent an effective treatment (57, 58).

Based on the identified transcriptomic subtypes we suggested potential therapeutic targets per subtype (Figure 7B). The luminal-a subtype was characterized by *MYCN* and *PPARGC1B* overexpression. In pre-clinical studies, treatment with a BET- or PPARγ-inhibitor downregulated the expression levels of both genes, and had an antiproliferative effect on tumor cells (59, 60). The immune cell compartment of tumors of the luminal-a subtype was found rich in NK cells, which could be explained by the large fraction of liver biopsies, which are known to be enriched for NK cells (61), showing this subtype. Thus, other potential treatment strategies comprise of cytokine-mediated stimulation of NK cells and TLR agonists (62).

The luminal-b subtype was enriched with *FGFR3* mutations and had high expression of *FGFR3*, suggesting that this subtype may be susceptible to FGFR inhibitors. This subtype may also be sensitive to RAS pathway inhibitors as the RTK-RAS pathway activity was high (63).

The stroma-rich subtype was characterized by *TSC1* alterations that confer sensitivity to MTOR inhibitors, which have been approved for treatment of several tumor types (64, 65). Compared with the other subtypes, the stroma-rich subtype displayed the highest TGF-β pathway activity and overexpression of different

collagens. Previous studies have shown that TGF-β can stimulate cancer-associated fibroblasts to produce collagens (66, 67). Other studies found that TGF-β expression was associated with resistance to immune checkpoint inhibition in bladder cancer (68, 69). Results from pre-clinical studies suggest that addition of a TGF-β inhibitor may improve anti-PD1 efficacy (70).

The basal/squamous subtype has been found associated with high immune cell infiltration (significantly more M1 macrophages) and overexpression of PD-L1, which suggests that patients with tumors of this subtype are likely to benefit from immunotherapy (4). Since TGF-β pathway activity was also high in this subtype, combination therapy with a TGF-β inhibitor could be of added value. Furthermore, this subtype was characterized by overexpression of mesothelin, a known tumor antigen that is being investigated as a target for antibody-based, vaccine and CAR-T cell therapies in several tumor types (71).

As our studied cohort was heterogeneous regarding pre-treatment history and type of treatment initiated after biopsy collection, we were unable to reliably correlate the characteristics of the molecular subtypes to clinical endpoints such as overall survival. Additional studies in which biopsies are collected from uniformly treated mUC patients would be crucial to be able to properly correlate large scale genomic and transcriptomic data with clinical outcomes.

The findings of this study significantly add to our knowledge on the molecular biology of mUC; the identified potential therapeutic targets should be addressed in further research – with the ultimate aim to improve the management of mUC patients.

Methods

Patient cohort and study procedures

Between 07 June 2012 up to and including 28 February 2019, patients with advanced or mUC (n = 210) from 23 Dutch hospitals (Data S1.1, Figure 1A and Table S1.3) were included in the Netherlands nationwide study of the Center for Personalized Cancer Treatment (CPCT) consortium (CPCT-02 Biopsy Protocol, ClinicalTrial.gov no. NCT01855477) (1), which aimed to analyze the cancer genome and transcriptome of patients with advanced cancer. The study protocol was approved by the medical ethics review board of the University Medical Center Utrecht, the Netherlands. Patients eligible for inclusion were those aged ≥ 18 years old, with locally advanced or mUC, from whom a histological tumor biopsy could be safely obtained, and whom had an indication for initiation of a new line of systemic treatment with anti-cancer agents. Written informed consent was obtained from all participants prior to inclusion in the trial; the study complies with all relevant ethical regulations. Tumor biopsies and matched normal blood samples were collected following a standardized procedure described by the Hartwig Medical Foundation (https://www.hartwigmedicalfoundation.nl; 1). For the current analysis, biopsies were obtained from a safely accessible site, including lymph nodes, liver, bone and other organs (Figure 1). In five patients, a tumor biopsy was obtained from the primary bladder or upper urinary tract tumor as no safely accessible metastatic lesion was present. In two patients, a biopsy was obtained from a local recurrence after cystectomy and nephrectomy, respectively (Table S1.2). This study extends the pan-cancer analysis of Priestley *et al.*, in which WGS data of 72 mUC patients included in the current cohort were initially analyzed (Table S1.2).

Central pathology review

Tumor tissue slides for central pathological revision of the diagnosis of UC was available for 79/116 patients. Hematoxylin and eosin (H&E) stained slides from primary tumor tissues (cystectomy and transurethral resection specimens of the bladder, n = 23 patients), metastatic tumor biopsies (n = 15 patients), or both (n = 41 patients) were requested from the Nationwide Network and Registry of Histo- and Cytopathology in the Netherlands (PALGA) (2). Tissue slides and corresponding pathological reports were provided anonymously. All patient materials used for central pathology review were obtained within the CPCT-02 biopsy protocol or during routine patient care, and the use of these materials for research purposes was approved by the medical

ethics review board of the Erasmus University Medical Center, Rotterdam, the Netherlands (MEC-2019-0188). H&E slides were reviewed by an expert genitourinary pathologist (LLB), and used for re-evaluation of the diagnosis and description of aberrant histology (Table S1.3 and S1.4). Tumors were classified as pure UC (n = 66), or predominant UC with variant histology (n = 9 squamous, n = 3 neuro-endocrine, n = 1 micropapillary UC), and pure squamous cell bladder carcinomas (n = 3). In patients for whom both the primary and the metastatic tumor biopsy was available for review, the highest grade (WHO 1973 classification) was assigned, and presence of aberrant histology in one of the tissue samples was considered as positive.

Whole-genome sequencing and analysis

Whole-genome DNA sequencing, alignment and data processing

Sufficient amount of DNA was extracted (50-200 ng) from fresh-frozen tumor and blood tissue following standard protocols from Qiagen. DNA was fragmented by sonication for NGS Truseq library preparation and sequenced paired-end reads of 2x 150 bases with the Illumina HiSeqX platform. Alignment, somatic alterations, ploidy, sample purity and copy numbers estimations were performed as previously described (1). WGS was aligned to the reference human genome GRCH37 with BWA-mem v.0.7.5a (3), and duplicate reads were marked for filtering. Indels were realigned using GATK IndelRealigner (v3.4.46) (4). Recalibration of base qualities for SNVs and Indels was performed with GATK BQSR (5), and SNV and Indel variants were evaluated with Strelka v.1.0.14 (6) using matched blood WGS as normal reference (Table S1.5).In-house pipeline (7) was used to further annotate somatic mutations with Ensembl Variant Effect Predictor (VEP, v99, cache 99_GRCh37) (8) using GENCODE v33 in combinations with the dbNSFP plugin (v3.5, hg19) (9) for gnomAD (10) population frequencies. SNVs, Indels and MNVs variants were removed if the following filters were not passed: default Strelka filters (PASS-only), gnomAD exome (ALL) allele frequency < 0.001, gnomAD genome (ALL) < 0.005 and number of reads < 3. In addition, structural variants and copy number changes were estimated using GRIDDS, PURPLE and LINX suit v2.25 (11). Structural variants that passed the default QC filters (PASS-only) and Tumor Allele Frequency TAF ≥ 0.1 were annotated as "somatic structural variants" if there was overlap with coding region. Mean read coverages of tumor and reference samples were estimated using Picard Tools (v1.141; CollectWgsMetrics) based on GRCh37 (https://broadinstitute.github.io/picard/). Genomic and coding

tumor mutational burden (TMB; mutations per megabase pair (Mbp)) were calculated considering SNVs, Indels and MNVs. The total number of somatic mutations in coding region was divided by 28.71 Mbp (protein-coding region size) and in the whole genome by 2,858.67 Mbp (genomic alignment size).

Detection and characterization of recurrent copy number alterations

Ploidy and CNAs were estimated as described by Priestley *et al.*; and following the custom pipeline described by van Dessel *et al.*, 2019, recurrent focal and broad CNAs were estimated with GISTIC2.0 v2.0.23 (12). CNAs were classified as shallow or deep according to the threshold in GISTIC2 calls. Significant recurrent focal CNAs were identified when q ≤ 0.05 and annotated with genes overlapping these regions, which were considered drivers (Table S1.8 and S1.10). The estimated copy number per gene was also estimated (Table S1.9).

Detection of driver genes using dN/dS ratios

We used the dNdScv R package (v0.0.0.9) to detect cancer driver genes under strong positive selection (13). This method used 192 mutation rates representing all combinations in trinucleotide context. Mutation rates of each gene were corrected by the global mutation rate. The ratio of non-synonymous over synonymous mutations was calculated with maximum-likelihood methods, and statistical significance was estimated. Genes with either qglobal_cv ≤ 0.05 or qallsubs_cv ≤ 0.05 were considered drivers of mUC (Table S1.7 and S1.10).

Other known driver genes that were frequently mutated were considered, and are shown in Data S1.3.

APOBEC enrichment and mutagenesis

For each sample, the total number of C>T or C>G (G>A or G>C) mutations was calculated ($C^{mut\,(C>T,C>G)}$). From these mutations, the total number of APOBEC mutations was estimated by counting all mutations in TCW (WGA) context (TCW^{Cmut}). The total number of TCW (WGA) motifs and total C (G) nucleotides in the reference genome were estimated ($TCW^{context}$ and $C^{context}$, respectively). Using this information and following Roberts *et al.*, 2013 , a contingency table was constructed; one-sided Fisher's exact test was applied to calculate the overrepresentation of APOBEC mutations. P-values were Benjamini-Hochberg corrected. Tumors with adjusted p-values lower than 0.01 were considered APOBEC enriched.

The magnitude of APOBEC enrichment *E* was estimated as (14)

$$E = \frac{\text{TCW}^{Cmut} \cdot \text{C}^{context}}{\text{TCW}^{context} \cdot \text{C}^{mut(C>T,C>G)}}. \tag{1}$$

APOBEC enriched tumors (always $E > 1$) were classified as high APOBEC mutagenesis when $E \geq 2$, and as low APOBEC mutagenesis when $E < 2$. Tumors without APOBEC enrichment were considered tumors with no APOBEC mutagenesis (Table S1.6).

Clonality fraction estimation

Mutations start as a single copy in the DNA, and multiple copies of the mutated nucleotide may appear if affected by CNAs events. Correcting for tumor purity and CNA, the number of copies n_{SNV} of each SNV was calculated as follows (15)

$$n_{SNV} = \frac{f_m}{p}[pC_t + (1-p)C_h], \tag{2}$$

where f_m is the relative frequency of the mutant variant reads, p is the tumor purity, C_t is the copy number affecting the region where a particular SNV was located, and C_h is the healthy copy number (2 for autosomes and 1 for allosomes).

Equation (2) is equivalent to the cancer cell fraction (CCF) with $n_{SNV} \approx 1$ in haploid and heterozygous-diploid regions; i.e., the fraction of tumor cells carrying a particular mutation. For regions with CNAs, $n_{SNV} > 1$, we must estimate the fraction of cancer cells carrying a particular SNV. As described previously (16, 17), we assume that all SNVs are present in the major copy number C_M; hence $n_{SNV} \leq C_M$ covering mutations that were acquired after copy number change events or present only in the minor copy number. Given the number of reference and mutant reads, and assuming binomial distribution, we estimated the expected number of allelic copies (n_{chr}) carrying the observed SNV resulting from f_m values when the mutation is present in 1, 2, 3,…, N_{chr} allelic copies. In some cases (sequencing noise) $n_{SNV} > C_M$, which was corrected with $N_{chr} = max(C_M, n_{SNV})$. We also corrected each f_m value with normal cell contamination – multiplying it by p. The resulting estimated n_{chr} with the maximum likelihood serves to calculate the CCF as n_{SNV}/n_{chr}.

Dirichlet process from the DPClust R package (https://github.com/Wedge-lab/dpclust) with 250 iterations and 125 burn in iterations was applied to the CCF distribution to estimate the fraction of clonal and subclonal SNVs per tumor (18). Multiple distributions (clusters) were obtained, representing different cancer cell populations.

The mean of the distributions was used to classify clusters of SNVs as clonal or subclonal. Clusters of SNVs with mean distribution > 0.8 were considered clonal (Table S1.5).

Mutational signatures and genomic subtypes

The mutational pattern of each sample was established by categorizing SNVs according to their 96-trinucleotide context. The contribution of each of the 67 mutational signatures from COSMIC v3 (as deposited on May 2019) (19) was subsequently estimated with MutationalPatterns v1.4.2 (Table S1.16) (20). To reduce the noise attributed to mutational signatures with very low contribution, mutational signatures were grouped into 26 proposed etiology categories (Table S1.15) derived from Alexandrov *et al.* (2020), Petljak *et al.* (2019), Angus *et al.* (2019) and Christensen *et al.* (2019). All 26 proposed etiology contributions were used, and hierarchical clustering was applied on 1-Pearson's correlation coefficient, 80% resampling and 1,000 iterations from the R package ConsensusClusterPlus v1.48.0 (25). Considering average stability of each cluster and the size of the clusters (favoring large clusters) after each partition, samples were grouped into five distinct clusters (Data S1.6).

Independently, mutational patterns were deconvoluted to estimate *de novo* mutational signatures. Non-negative Matrix Factorization from the NMF R package (v0.21.0) was used with 1000 iterations (26). Evaluating different metrics provided by the NMF R package (high cophenetic correlation coefficient, high dispersion coefficient, high silhouette consensus, high sparseness basis and low sparseness coefficients), seven *de novo* signatures were recovered from the mutational patterns (Data S1.7). Cosine similarity was applied to compare the *de novo* signatures with mutational signatures from COSMIC v3.

Detection of chromothripsis

Catastrophic-like events, such as chromothripsis, were detected with Shatterseek v0.4 (27) using default parameters. Absolute copy numbers (as derived by PURPLE) were rounded to the nearest integer; only structural variants with TAF ≥ 0.1 at either end of the breakpoint were considered, and chrY was excluded. The following filters were applied and chromothripsis was considered when: a) the number of intra-chromosomal structural variants ≥ 25; b) the maximum number of oscillating CN segments with two states ≥ 7 or with three states ≥ 14; c) the size of the chromothripsis event ≥ 20 Mbp; d) random distribution of breakpoints $p \leq 0.05$; and e) *chromosomal breakpoint enrichment p ≤ 0.05* (Table S1.19).

MicroSatellite Instabilty (MSI) status

As previously described (1), MSI status was determined by estimating the MSI score as the number of indels (length < 50 bp) per Mbp occurring in homopolymers of five or more bases, dinucleotide, trinucleotide and tetranucleotide sequences of repeat count above five. Tumors with MSI score > 4 were considered MSI positive (Table S1.21).

Detection of homologous recombination (HR) deficiency

The Classifier for Homologues Recombination Deficiency (CHORD; v2.0) with default parameters was used to identify tumors with HR proficiency and deficiency (28). Four samples had very high number of indels corresponding with MSI samples (Figure 2A) and were discarded for the HR deficiency analysis (Table S1.20).

Detection of kataegis

Following the method described by van Dessel *et al.* (2019), kataegis events were estimated using SNVs. Each chromosome was divided into segments (maximum 5000 segments) of five or more consecutive SNVs. Segments were considered a kataegis event when the mean intermutational distance was ≤ 2000 bp (Table S1.18).

Mutational load across genomic regions

The genome was divided in regions (bins) with one Mbp size. The number of SNVs was counted in each bin, and the mean number of SNVs was estimated from the entire cohort. These values represented the average SNVs/Mbp reflecting the mutational load in each genomic region. APOBEC mutations were identified as C>T and C>G mutations in TCW trinucleotide context (14), where W is A or T. The relative density of APOBEC mutations was calculated as the proportion of APOBEC mutations over the total SNVs/Mbp at each genomic bin position. The average SNVs/Mbp and relative APOBEC density were smoothed by applying a moving average with k = 9 bins. This approach was used per sample and for mean values from the entire cohort.

Genomic alteration of oncogenic pathways

Eleven oncogenic pathways were analyzed for somatic alterations. The list of genes was modified from Sanchez-Vega *et al.* (2018) and Leonard (2001) (Table S1.23). Altered pathways were defined when at least one of the pathway-genes was affected by any somatic mutation (SNV, Indel, MNV, SV or deep CNA; excluding synonymous mutations).

Inventory of clinically-actionable somatic alterations and putative therapeutic targets

Current clinical relevance of somatic alterations in relation to putative treatment options or resistance mechanisms and trial eligibility was determined based upon the following databases: CiViC (31) (Nov. 2018), OncoKB (32) (Nov. 2018), CGI (33) (Nov. 2018) and the iClusion (Dutch) clinical trial database (Sept. 2019, Rotterdam, the Netherlands). The databases were aggregated and harmonized using the HMF knowledgebase-importer (v1.7; https://github.com/hartwigmedical/hmftools/tree/master/knowledgebase-importer). Subsequently, we curated the linked putative treatments and selected treatments for which level A (biomarker for approved therapy or in guidelines) or level B (biomarker on strong biological evidence or used in clinical trials) evidence was available (1). Genomic alterations that confer resistance to certain therapies were excluded for the analysis. Treatment strategies including anti-hormonal therapy (as used for breast and prostate cancer), surgical resection, or radioiodine uptake therapy were excluded. Furthermore, closed trials (according to www.clinicaltrials.gov), and trials with only pediatric patients or patients with hematological malignancies were excluded (Table S1.24). The data base was complemented with *FGFR3* and *NTRK2* gene fusions (at RNA level) and patients with MSI high and HR deficient tumors. On-label treatments included

chemotherapy (cisplatin, gemcitabine, doxorubicin, mitomycin, and valrubicin) and the FGFR3 inhibitor erdafitinib. Off-label treatments included treatments that are on-label for other tumor types (FDA approved drugs according to the US national cancer institute; https://www.cancer.gov/about-cancer/treatment/drugs/cancer-type), and treatments available in clinical trials or basket trials. When patients had more than one possible treatment, on-label treatment was the preferred treatment, followed by on-label treatments for other tumor types.

DNA accessibility estimation (ChIPseq)

All available ChIPseq data for healthy urinary bladder (H3K4me1, H3K4me3, H3K36me3 and H3K27ac) were downloaded from the ENCODE portal (https://www.encodeproject.org) to our local server. The *bed.gz* files were imported with *narrowPeak* format for analysis. The signal of each experiment was divided in regions of one Mbp, and a moving average with k = 9 bins was applied. The scale of the signal was normalized; hence the sum of all regions at each chromosome is one. This step was taken to compensate for the bias observed in peak intensity signals across different chromosomes, possible due to technical issues in the ChIPseq technology, e.g. hyper-ChIPable regions or mappability (34).

High DNA accessible regions (open chromatin) were determined as such if the ChIPseq signal value of the region was above the median. Otherwise, the region was considered as low DNA accessible (condensed chromatin). This procedure was applied on each chromosome. Figure 5-B-C and Data S1.9A were generated using results from H3K4me1 ChIPseq. Using other ChIPseq experiments showed the same result.

Whole-transcriptome sequencing and analysis

RNA-sequencing, alignment and data pre-processing

Total RNA was extracted using the QIAGEN QIAsymphony kit (Qiagen, FRITSCH GmbH, Idar-Oberstein, Germany). Samples with a minimum of 100 ng total RNA were sequenced according to the manufacturer's protocols. Paired-end sequencing of RNA was performed on the Illumina NextSeq 550 platform (2x75bp) and Illumina NovaSeq 6000 platform (2x150bp).

Prior to alignment, samples were visually inspected with FastQC (v0.11.5). Sequence adapters (Illumina TruSeq) were trimmed using Trimmomatic v0.39 (35) at the following settings:

ILLUMINACLIP:adapters.fa:2:30:10:2:keepBothReads MINLEN:36. The trimmed paired-end reads were aligned to the human reference (GRCh37) using STAR v2.7.1a (36) with genomic annotations from GENCODE hg19 release 30 (37). Multiple lanes and runs per sample were aligned simultaneously and given respective read-group identifiers for use in downstream analysis to produce two BAM files per sample, consisting of genome- and transcriptome-aligned reads respectively.

STAR was performed using the following command:

STAR --genomeDir <genome> --readFilesIn <R1> <R2> --readFilesCommand zcat --outFileNamePrefix <outPrefix> --outSAMtype BAM SortedByCoordinate --outSAMunmapped Within --chimSegmentMin 12 --chimJunctionOverhangMin 12 --chimOutType WithinBAM --twopassMode Basic --twopass1readsN -1 --runThreadN 10 --limitBAMsortRAM 10000000000 --quantMode TranscriptomeSAM --outSAMattrRGline <RG>

After alignment, duplicate reads were marked and alignment quality metrics (flagstat) were generated using Sambamba v0.7.1 (38). For each genome-aligned sample, the uniformity of read distributions across transcript lengths was assessed using tin.py v 2.6.6 (39) from the RSeQC library v3.0.0 (40).

FeatureCounts v1.6.3 (41) was applied to count the number of overlapping reads per gene using genomic annotations from GENCODE (hg19) release 30 (37); only primary (uniquely mapped) reads were counted per exon and summarized per gene in a strand-specific manner:

featureCounts -T 50 -t exon -g gene_id --primary -p -s 1 -a <gencode> -o <output> <genomic BAMs>

RSEM v1.3.1 (42) was applied to quantify RNA expression into transcripts per million (TPM) values using transcript annotations from GENCODE (hg19) release 30 (37):

rsem-calculate-expression --bam --paired-end --strand-specific --alignments -p 8 <transcriptome BAM> <RSEM Index> <output>

Transcriptome expression data mapped to genomic regions

MultiBamSummary from deepTools (43) (v1.30.0) was used to read BAM files and estimate number of reads in genomic regions with a size of one Mbp. The average raw read count per Mbp was calculated, and a moving average with k = 9 bins was applied. The scale of the read counts was normalized following the method for

DNA accessibility regions, and high transcriptional regions were defined as such when the expression value of one region was above the median. This procedure was applied on each chromosome.

Transcriptomic subtypes: clustering samples by RNAseq data

Transcripts were normalized using DESeq2 v1.24.0 (44) with variance stabilizing transformation. Only transcripts with base mean above 100 were kept. The top 50% most variably expressed genes (7,030 transcripts) were used for cluster analysis.

Samples were grouped according to their biopsy site: liver (n = 31), lymph node (n = 30) and bone (n = 5), other (n = 23) and unknown (n=1). Differential analysis was performed to compare tumors from a specific biopsy site (liver, lymph node and bone) against all other tumors using DESeq2 v1.24.0 (44) with Wald test p-values estimation. Tissue-specific transcripts with $\log_2$ Fold Change ($\log_2$FC) > 1 and Benjamini-Hochberg corrected p-value < 0.05 were considered differentially expressed and tagged as tissue-specific. A total of 754 transcripts were tissue-specific, and were removed from the data set (Figure S5).

The remaining 6,276 normalized transcripts were grouped using hierarchical clustering with 1-Pearson's correlation coefficient, 80% resampling and 1,000 iterations from the R package ConsensusClusterPlus v1.48.0 (25). The mean cluster consensus value was obtained as a measure of cluster stability. Increasing the number of clusters will increase the stability by creating smaller clusters. Taking this into account, the criteria for selecting five clusters was based on cluster stability and the size of the clusters by not allowing clusters with <5 samples (Table S1.17).

All normalized transcripts (excluding biopsy specific transcripts) were used as input to classify each tumor into one of the six molecular classes identified in primary UC using the ConsensusMIBC classifier v1.1.0 (45).

To identify transcripts that contribute most to each cluster, we followed the same strategy used to identify tissue-specific transcripts. The top five transcripts with the highest $\log_2$FC and with Benjamini-Hochberg adjusted p-values lower than 1×10^{-5} are shown in Figure 5. Other differentially expressed genes are included in Figure 5 for their clinical relevance (*TGFB3*, *DDR2*, *PDGFRA*, *CD274* and *TGFBR1*). All differentially expressed genes per cluster with adjusted $p < 1 \times 10^{-5}$ and $\log_2$FC > 1 are listed in Table S1.22.

Phenotypic markers and signature score

Marker genes for basal (*CD44, CDH3, KRT1, KRT14, KRT16, KRT5, KRT6A, KRT6B, KRT6C*), squamous (*DSC1, DSC2, DSC3, DSG1, DSG2, DSG3, S100A7, S100A8*), luminal (*CYP2J2, ERBB2, ERBB3, FGFR3, FOXA1, GATA3, GPX2, KRT18, KRT19, KRT20, KRT7, KRT8, PPARG, XBP1, UPK1A, UPK2*), neuroendocrine (*CHGA, CHGB, SCG2, ENO2, SYP, NCAM1*), cancer-stem cell (*CD44, KRT5, RPSA, ALDH1A1*), EMT epithelial-mesenchymal (*ZEB1, ZEB2, VIM, SNAI1, TWIST1, FOXC2, CDH2*) and claudin-low (*CLDN3, CLDN7, CLDN4, CDH1, SNAI2, VIM, TWIST1, ZEB1, ZEB2*) were used for signature score (46). Stroma (*FAP*) and Interferon and CD8+ effector T cells (*IFNG, CXCL9, CD8A, GZMA, GZMB, CXCL10, PRF1, TBX21*) markers were also included (47). All normalized expression values were median centered, and the mean expression of each group of genes was defined as signature score (Figure 5 and S6A).

Pathway activity score

Genes transcriptionally activated by the eleven canonical pathways analyzed in this study were used to estimate pathway activity score (Figure 6B). All normalized expression values were median centered, and the mean expression of each group of genes was defined as activity score. Activity score was estimated for TGFβ pathway (*ACTA2, ACTG2, ADAM12, ADAM19, CNN1, COL4A1, CCN2, CTPS1, RFLNB, FSTL3, HSPB1, IGFBP3, PXDC1, SEMA7A, SH3PXD2A, TAGLN, TGFBI, TNS1, TPM1*) (48), cell cycle pathway (*MKI67, CCNE1, BUB1, BUB1B, CCNB2, CDC25C, CDK2, MCM4, MCM6, MCM2*) (47), WNT pathway (*EFNB3, MYC, TCF12, VEGFA*) (48), Notch pathway (*HES1, HES5, HEY1*) (49), PI3K pathway (*AGRP, BCL2L11, BCL6, BNIP3, BTG1, CAT, CAV1, CCND1, CCND2, CCNG2, CDKN1A, CDKN1B, ESR1, FASLG, FBXO32, GADD45A, INSR, MXI1, NOS3, PCK1, POMC, PPARGC1A, PRDX3, RBL2, SOD2, TNFSF10*) (50), hippo pathway (*TAZ, YAP1*) (51), p53 pathway (*CDKN1A, RRM2B, GDF15, SUSD6, BTG2, DDB2, GADD45A, PLK3, TIGAR, RPS27L, TNFRSF10B, TRIAP1, ZMAT3, BAX, BLOC1S2, PGF, POLH, PPM1D, PSTPIP2, SULF2, XPC*) (52), Nrf2 pathway (*GCLM, NQO1, PHGDH, PSAT1, SHMT2*) (53), MYC pathway (*TFAP4, BMP7, CCNB1, CCND2, CCNE1, CDC25A, CDK4, CDT1, E2F1, GATA4, HMGA1, HSP90AA1, JAG2, CDCA7, LDHA, MCL1, NDUFAF2, MTA1, MYCT1, NPM1, ODC1, SPP1, PIN1, PTMA, PRDX3, PRMT5, DNPH1, TFRC, EMP1, PMEL, C1QBP*) (54), RTK-RAS pathway (*SPRY2, SPRY4, ETV4, ETV5, DUSP4, DUSP6, CCND1, EPHA2, EPHA4*) (55) and JAK-STAT pathway (*IRGM, ISG15, GATA3, FCER2, THY1, NFIL3, ARG1, RETNLB, CLEC7A, CHIA, OSM, BCL2L1, CISH, PIM1, SOCS2, GRB10*) (56).

Pathway enrichment analysis

All differentially expressed genes with adjusted $p < 0.05$ and absolute $log_2FC > 1$ in each of the transcriptomic subtypes were used for pathway enrichment analysis. Using reactomePA (57), the top ten (sorted by adjusted p-value) up- and down-regulated pathways were selected, and these are shown in Figure 6C.

Immune cell infiltration

To quantify immune cell fractions in each sample (Figure 5 and S6B), we analyzed RSEM read counts of all transcripts with the R package immunedeconv (58), using the quanTIseq method (59).

Detection of fusion genes

In-frame gene fusions were detected at DNA level by the GRIDDS, PURPLE, LINX suite v2.25 (11), and reported relevant if they appear in the ChimerDB 4.0 (Table S1.13) (60). At RNA level, Arriba (https://github.com/suhrig/arriba/) was used to infer gene fusion events with the option to discard known false positives from a list provided by Arriba. High confidence fusions were retained, and only events where at least one transcript is protein coding were kept (Table S1.14). Fusion genes previously identified by other studies, mostly from the TCGA data, were evaluated with ChimerDB 4.0 (60). All "deletion/read-through" events were discarded as possible false positives unless they were supported by the ChimerDB 4.0 data base. Medium confidence fusions were included in the final list if one of the fused genes appears in a high confidence event.

APOBEC mutation rate and APOBEC expression in tumors with multiple sequential biopsies

A second metastatic tumor biopsy was taken in eight patients from the same (n = 5) or a different (n = 3) metastatic lesion, and analyzed by WGS (n = 8) and RNAseq (n = 7). Each patient's first and second biopsies shared a high proportion of mutations (SNVs, Indels and MNVs), confirming the clonal relation of the sampled sites (Figure S4A). Dirichlet process from the DPClust R package (https://github.com/Wedge-lab/dpclust) with 250 iterations and 125 burn in iterations was applied to the CCF distribution of paired-biopsies to estimate the subclonal (clusters) composition of each tumor. All unique mutations in each biopsy were considered a subclone; only subclones with >5% of SNVs were considered relevant. Small populations of subclones (<5% of

SNVs) were merged to the nearest subclone. The evolutionary tree (Figure S4A) was reconstructed using the *sum rule* (61).

The CCF of somatic mutations in the branches was lower than that in the trunk (Figure 4B), suggesting that these mutations are recently acquired mutations. To estimate the rate of novel APOBEC associated mutations, only unique SNVs from the second biopsy were kept, as these somatic alterations may probably correspond to new mutations acquired during the time frame between the biopsies. As the time elapsed between the first and the second biopsy varied between tumors, we normalized the number of recent APOBEC associated mutations by dividing the total over the number of days elapsed between the biopsies. The value estimated is proportional to the mutation rate of APOBEC associated mutations (mutations per day).

In seven tumors, RNAseq data was available. Expression of *APOBEC3A* and *APOBEC3B* in Figure 3D is the mean normalized expression of the two biopsies.

Analysis of the *The Cancer Genome Atlas* primary bladder cancer cohort

Public open data of the TCGA bladder cancer cohort, including somatic mutations detected by Mutect (SNVs and Indels) of 412 tumors, GISTIC copy number changes at gene level of 410 tumors and RNAseq (HTSeq counts; Affymetrix SNP6 arrays) data available for 410 tumors were analyzed (Figure S3). Some samples had very few mutations, and only tumors with total SNVs > 50 were considered in this analysis (367/412). The same method applied on our mUC cohort was applied on the TCGA data to deconvolute mutational signatures and to identify genomic subtypes. Twelve genomic subtypes were identified, but several of them formed small groups with very specific mutational signature patterns, including one sample with very high POLE signature. All genomic subtypes with < 1% of the total cohort were grouped together in GenS0. We compared the genomic subtypes between mUC and the TCGA cohort using cosine similarity.

Transcript counts were normalized with DESeq2 (44) following the same procedure as used for the mUC cohort. All tumors were from primary UC, and organ-specific transcript were not discarded. The consensus MIBC classifier (45) was applied to infer the transcriptomic subtype of each tumor.

Code availability

All custom code and scripts are available at https://bitbucket.org/ccbc/dr31_hmf_muc/ and

https://github.com/hartwigmedical/.

Statistical analysis

Analysis was performed using the statistical analysis platform R v3.6.1 (65).

Data availability

WGS data, RNA-seq data and corresponding clinical data have been requested from Hartwig Medical

Foundation and were provided under data request number DR-031. All data is freely available for academic

use from the Hartwig Medical Foundation through standardized procedures. Request forms can be found at

https://www.hartwigmedicalfoundation.nl (1).

ChIPseq data experiments are freely available through The ENCODE Project Consortium (62) and the Roadmap

Epigenomics Consortium (63) on the ENCODE portal (https://www.encodeproject.org) (64). We downloaded

files with the following identifiers: ENCSR065IQH, ENCSR054BKO, ENCSR632OWD and ENCSR449TNC.

TCGA data for the bladder cancer cohort was downloaded through the portal: https://www.cancer.gov/tcga.

References to main text

1. MD Wilkerson, DN Hayes, ConsensusClusterPlus: A class discovery tool with confidence assessments and item tracking. *Bioinformatics* **2010**;26:1572–1573.

2. MA Knowles, CD Hurst, Molecular biology of bladder cancer: New insights into pathogenesis and clinical diversity. *Nat Rev Cancer* **2015**;15:25–41.

3. SV Lindskrog, FF Prip, P Lamy, A Taber, S Clarice, I Nordentoft, *et al.*, An integrated multi-omics analysis identifies clinically relevant molecular subtypes of non-muscle-invasive bladder cancer. *medRxiv* **2020**;2020.06.19.20054809.

4. AG Robertson, J Kim, H Al-Ahmadie, J Bellmunt, G Guo, AD Cherniack, *et al.*, Comprehensive Molecular Characterization of Muscle-Invasive Bladder Cancer. *Cell* **2017**;171:540-556.e25.

5. A Giannopoulou, A Velentzas, E Konstantakou, M Avgeris, S Katarachia, N Papandreou, *et al.*, Revisiting Histone Deacetylases in Human Tumorigenesis: The Paradigm of Urothelial Bladder Cancer. *Int J Mol Sci* **2019**;20:1291.

6. A Kamoun, A de Reyniès, Y Allory, G Sjödahl, A Gordon Robertson, R Seiler, *et al.*, A Consensus Molecular Classification of Muscle-invasive Bladder Cancer. *Eur Urol* **2019**;77:420–433.

7. BM Faltas, D Prandi, ST Tagawa, AM Molina, DM Nanus, C Sternberg, *et al.*, Clonal evolution of chemotherapy-resistant urothelial carcinoma. *Nat Genet* **2016**;48:1490–1499.

8. F Martínez-Jiménez, F Muiños, I Sentís, J Deu-Pons, I Reyes-Salazar, C Arnedo-Pac, *et al.*, A compendium of mutational cancer driver genes. *Nat Rev Cancer* **2020**;20:555–572.

9. JG Tate, S Bamford, HC Jubb, Z Sondka, DM Beare, N Bindal, *et al.*, COSMIC: The Catalogue Of Somatic Mutations In Cancer. *Nucleic Acids Res* **2019**;47:D941–D947.

10. I Martincorena, KM Raine, M Gerstung, KJ Dawson, K Haase, P Van Loo, *et al.*, Universal Patterns of Selection in Cancer and Somatic Tissues. *Cell* **2017**;171:1029-1041.e21.

11. CH Mermel, SE Schumacher, B Hill, ML Meyerson, R Beroukhim, G Getz, GISTIC2.0 facilitates sensitive and confident localization of the targets of focal somatic copy-number alteration in human cancers.

Genome Biol **2011**;12:R41.

12.   Y Allory, W Beukers, A Sagrera, M Flández, M Marqués, M Márquez, *et al.*, Telomerase reverse transcriptase promoter mutations in bladder cancer: High frequency across stages, detection in urine, and lack of association with outcome. *Eur Urol* **2014**;65:360–366.

13.   R Buisson, A Langenbucher, D Bowen, EE Kwan, CH Benes, L Zou, *et al.*, Passenger hotspot mutations in cancer driven by APOBEC3A and mesoscale genomic features. *Science* **2019**;364:eaaw2872.

14.   YE Jang, I Jang, S Kim, S Cho, D Kim, K Kim, *et al.*, ChimerDB 4.0: an updated and expanded database of fusion genes. *Nucleic Acids Res* **2019**;48:D817–D824.

15.   AP Glaser, D Fantini, Y Wang, Y Yu, KJ Rimar, JR Podojil, *et al.*, APOBEC-mediated mutagenesis in urothelial carcinoma is associated with improved survival, mutations in DNA damage response genes, and immune response. *Oncotarget* **2018**;9:4537–4548.

16.   SA Roberts, MS Lawrence, LJ Klimczak, SA Grimm, D Fargo, P Stojanov, *et al.*, An APOBEC cytidine deaminase mutagenesis pattern is widespread in human cancers. *Nat Genet* **2013**;45:970–976.

17.   HX Dang, BS White, SM Foltz, CA Miller, J Luo, RC Fields, *et al.*, ClonEvol: clonal ordering and visualization in cancer sequencing. *Ann Oncol* **2017**;28:3076–3082.

18.   MD Kazanov, SA Roberts, P Polak, J Stamatoyannopoulos, LJ Klimczak, DA Gordenin, *et al.*, APOBEC-Induced Cancer Mutations Are Uniquely Enriched in Early-Replicating, Gene-Dense, and Active Chromatin Regions. *Cell Rep* **2015**;13:1103–1109.

19.   JII Hoopes, LMM Cortez, TMM Mertz, EPP Malc, PAA Mieczkowski, SAA Roberts, APOBEC3A and APOBEC3B Preferentially Deaminate the Lagging Strand Template during DNA Replication. *Cell Rep* **2016**;14:1273–1282.

20.   CA Davis, BC Hitz, CA Sloan, ET Chan, JM Davidson, I Gabdank, *et al.*, The Encyclopedia of DNA elements (ENCODE): data portal update. *Nucleic Acids Res* **2018**;46:D794–D801.

21.   S Nik-Zainal, LB Alexandrov, DC Wedge, P Van Loo, CD Greenman, K Raine, *et al.*, Mutational processes molding the genomes of 21 breast cancers. *Cell* **2012**;149:979–993.

22. G Sturm, F Finotello, F Petitprez, JD Zhang, J Baumbach, WH Fridman, *et al.*, Comprehensive evaluation of transcriptome-based cell-type quantification methods for immuno-oncology in *Bioinformatics*, (Oxford University Press), pp. i436–i445.

23. F Finotello, C Mayer, C Plattner, G Laschober, D Rieder, H Hackl, *et al.*, Molecular and pharmacological modulators of the tumor immune contexture revealed by deconvolution of RNA-seq data. *Genome Med* **2019**;11:34.

24. A Baluapuri, E Wolf, M Eilers, Target gene-independent functions of MYC oncoproteins. *Nat Rev Mol Cell Biol* **2020**;21:255–267.

25. K Kuroda, M Yashiro, T Sera, Y Yamamoto, Y Kushitani, A Sugimoto, *et al.*, The clinicopathological significance of Thrombospondin-4 expression in the tumor microenvironment of gastric cancer. *PLoS One* **2019**;14:e0224727.

26. Y Gu, T Li, A Kapoor, P Major, D Tang, Contactin 1: An important and emerging oncogenic protein promoting cancer progression and metastasis. *Genes (Basel)* **2020**;11:1–22.

27. Q Zhang, N Zhou, W Wang, S Zhou, A novel autocrine CXCL14/ACKR2 Axis: The achilles' heel of cancer metastasis? *Clin Cancer Res* **2019**;25:3476–3478.

28. E Mathew, Y Zhang, AM Holtz, KT Kane, JY Song, BL Allen, *et al.*, Dosage-dependent regulation of pancreatic cancer growth and angiogenesis by Hedgehog signaling. *Cell Rep* **2014**;9:484–494.

29. P Eriksson, M Aine, S Veerla, F Liedberg, G Sjödahl, M Höglund, Molecular subtypes of urothelial carcinoma are defined by specific gene regulatory systems. *BMC Med Genomics* **2015**;8:25.

30. BL Fridley, J Dai, R Raghavan, Q Li, SJ Winham, X Hou, *et al.*, Transcriptomic characterization of endometrioid, clear cell, and high-grade serous epithelial ovarian carcinoma. *Cancer Epidemiol Biomarkers Prev* **2018**;27:1101–1109.

31. N Hauptman, E Boštjančič, M Žlajpah, B Ranković, N Zidar, Bioinformatics Analysis Reveals Most Prominent Gene Candidates to Distinguish Colorectal Adenoma from Adenocarcinoma. *Biomed Res Int* **2018**;2018:9416515.

32.   DZ Dieters-Castator, PF Rambau, LE Kelemen, GM Siegers, GA Lajoie, LM Postovit, *et al.*, Proteomics-derived biomarker panel improves diagnostic precision to classify endometrioid and high-grade serous ovarian carcinoma. *Clin Cancer Res* **2019**;25:4309–4319.

33.   F Sanchez-Vega, M Mina, J Armenia, WK Chatila, A Luna, KC La, *et al.*, Oncogenic Signaling Pathways in The Cancer Genome Atlas. *Cell* **2018**;173:321-337.e10.

34.   G Yu, QY He, ReactomePA: An R/Bioconductor package for reactome pathway analysis and visualization. *Mol Biosyst* **2016**;12:477–479.

35.   Y Haupt, R Maya, A Kazaz, M Oren, Mdm2 promotes the rapid degradation of p53. *Nature* **1997**;387:296–299.

36.   NI Nissen, M Karsdal, N Willumsen, Collagens and Cancer associated fibroblasts in the reactive stroma and its relation to Cancer biology. *J Exp Clin Cancer Res* **2019**;38:115.

37.   A Taber, E Christensen, P Lamy, I Nordentoft, F Prip, SV Lindskrog, *et al.*, Molecular correlates of cisplatin-based chemotherapy response in muscle invasive bladder cancer by integrated multi-omics analysis. *Nat Commun* **2020**;11:1–15.

38.   P Priestley, J Baber, MP Lolkema, N Steeghs, E de Bruijn, C Shale, *et al.*, Pan-cancer whole-genome analyses of metastatic solid tumours. *Nature* **2019**;575:210–216.

39.   W Traut, D Weichenhan, H Himmelbauer, H Winking, New members of the neurexin superfamily: Multiple rodent homologues of the human CASPR5 gene. *Mamm Genome* **2006**;17:723–731.

40.   Y Zou, W Zhang, H Liu, X Li, X Zhang, X Ma, *et al.*, Structure and function of the contactin-associated protein family in myelinated axons and their relationship with nerve diseases. *Neural Regen Res* **2017**;12:1551.

41.   MR Conserva, I Redavid, L Anelli, A Zagaria, G Specchia, F Albano, RARG Gene Dysregulation in Acute Myeloid Leukemia. *Front Mol Biosci* **2019**;6:114.

42.   SR Gheorghe, AM Craciun, Matrix Gla protein in tumoral pathology. *Clujul Med* **2016**;89:319–321.

43.   X Li, R Wei, M Wang, L Ma, Z Zhang, L Chen, *et al.*, MGP Promotes Colon Cancer Proliferation by

Activating the NF-κB Pathway through Upregulation of the Calcium Signaling Pathway. *Mol Ther - Oncolytics* **2020**;17:371–383.

44. M Wang, L Chen, Y Chen, R Wei, Q Guo, S Zhu, *et al.*, Intracellular matrix Gla protein promotes tumor progression by activating JAK2/STAT5 signaling in gastric cancer. *Mol Oncol* **2020**;14:1045–1058.

45. S Wu, T Ou, N Xing, J Lu, S Wan, C Wang, *et al.*, Whole-genome sequencing identifies ADGRG6 enhancer mutations and FRS2 duplications as angiogenesis-related drivers in bladder cancer. *Nat Commun* **2019**;10:1–12.

46. RJA Bell, HT Rube, A Xavier-Magalhães, BM Costa, A Mancini, JS Song, *et al.*, Understanding TERT promoter mutations: A common path to immortality. *Mol Cancer Res* **2016**;14:315–323.

47. ARJ Lawson, F Abascal, THH Coorens, Y Hooks, L O'Neill, C Latimer, *et al.*, Extensive heterogeneity in somatic mutation and selection in the human bladder. *Science* **2020**;370:75–82.

48. MJ Shi, XY Meng, J Fontugne, CL Chen, F Radvanyi, I Bernard-Pierrot, Identification of new driver and passenger mutations within APOBEC-induced hotspot mutations in bladder cancer. *Genome Med* **2020**;12:85.

49. LHG Matheus, SV Dalmazzo, RBO Brito, LA Pereira, RJ De Almeida, CP Camacho, *et al.*, 1-Methyl-D-tryptophan activates aryl hydrocarbon receptor, a pathway associated with bladder cancer progression. *BMC Cancer* **2020**;20:869.

50. J YU, Y LU, S MUTO, H IDE, S HORIE, The Dual Function of Aryl Hydrocarbon Receptor in Bladder Carcinogenesis. *Anticancer Res* **2020**;40:1345–1357.

51. SU Kang, JT Park, Functional evaluation of alternative splicing in the FAM190A gene. *Genes and Genomics* **2019**;41:193–199.

52. K Patel, F Scrimieri, S Ghosh, J Zhong, MS Kim, YR Ren, *et al.*, FAM190A deficiency creates a cell division defect. *Am J Pathol* **2013**;183:296–303.

53. JP Stein, G Lieskovsky, R Cote, S Groshen, AC Feng, S Boyd, *et al.*, Radical cystectomy in the treatment of invasive bladder cancer: Long-term results in 1,054 patients. *J Clin Oncol* **2001**;19:666–675.

54. J-C Marine, S-J Dawson, MA Dawson, Non-genetic mechanisms of therapeutic resistance in cancer. *Nat Rev Cancer* **2020**;20:1–14.

55. XB Pivot, I Bondarenko, M Dvorkin, E Trishkina, J-H Ahn, S-A Im, *et al.*, A randomized, double-blind, phase III study comparing SB3 (trastuzumab biosimilar) with originator trastuzumab in patients treated by neoadjuvant therapy for HER2-positive early breast cancer. *J Clin Oncol* **2017**;35:509–509.

56. JE Rosenberg, PH O'Donnell, A V. Balar, BA McGregor, EI Heath, EY Yu, *et al.*, Pivotal Trial of Enfortumab Vedotin in Urothelial Carcinoma After Platinum and Anti-Programmed Death 1/Programmed Death Ligand 1 Therapy. *J Clin Oncol* **2019**;37:2592–2600.

57. V Boni, MR Sharma, A Patnaik, The Resurgence of Antibody Drug Conjugates in Cancer Therapeutics: Novel Targets and Payloads. *Am Soc Clin Oncol Educ B* **2020**;40:e58–e74.

58. X Sheng, X Yan, L Wang, Y Shi, X Yao, H Luo, *et al.*, Open-label, Multicenter, Phase II Study of RC48-ADC, a HER2-Targeting Antibody–Drug Conjugate, in Patients with Locally Advanced or Metastatic Urothelial Carcinoma. *Clin Cancer Res* **2021**;27:43–51.

59. JE Delmore, GC Issa, ME Lemieux, PB Rahl, J Shi, HM Jacobs, *et al.*, BET bromodomain inhibition as a therapeutic strategy to target c-Myc. *Cell* **2011**;146:904–917.

60. JT Goldstein, AC Berger, J Shih, FF Duke, L Furst, DJ Kwiatkowski, *et al.*, Genomic activation of PPARG reveals a candidate therapeutic axis in bladder cancer. *Cancer Res* **2017**;77:6987–6998.

61. H Peng, E Wisse, Z Tian, Liver natural killer cells: Subsets and roles in liver immunity. *Cell Mol Immunol* **2016**;13:328–336.

62. MC Ochoa, L Minute, I Rodriguez, S Garasa, E Perez-Ruiz, S Inogés, *et al.*, Antibody-dependent cell cytotoxicity: immunotherapy strategies enhancing effector NK cells. *Immunol Cell Biol* **2017**;95:347–355.

63. AR Moore, SC Rosenberg, F McCormick, S Malek, RAS-targeted therapies: is the undruggable drugged? *Nat Rev Drug Discov* **2020**;19:533–552.

64. D Chakravarty, J Gao, S Phillips, R Kundra, H Zhang, J Wang, *et al.*, OncoKB: A Precision Oncology

Knowledge Base. *JCO Precis Oncol* **2017**;1:1–16.

65. D Tamborero, C Rubio-Perez, J Deu-Pons, MP Schroeder, A Vivancos, A Rovira, *et al.*, Cancer Genome Interpreter annotates the biological and clinical relevance of tumor alterations. *Genome Med* **2018**;10:25.

66. XM Meng, DJ Nikolic-Paterson, HY Lan, TGF-β: The master regulator of fibrosis. *Nat Rev Nephrol* **2016**;12:325–338.

67. LA Borthwick, TA Wynn, AJ Fisher, Cytokine mediated tissue fibrosis. *Biochim Biophys Acta - Mol Basis Dis* **2013**;1832:1049–1060.

68. T Powles, M Kockx, A Rodriguez-Vida, I Duran, SJ Crabb, MS Van Der Heijden, *et al.*, Clinical efficacy and biomarker analysis of neoadjuvant atezolizumab in operable urothelial carcinoma in the ABACUS trial. *Nat Med* **2019**;25:1706–1714.

69. N van Dijk, A Gil-Jimenez, K Silina, K Hendricksen, LA Smit, JM de Feijter, *et al.*, Preoperative ipilimumab plus nivolumab in locoregionally advanced urothelial cancer: the NABUCCO trial. *Nat Med* **2020**;26:1839–1844.

70. S Mariathasan, SJ Turley, D Nickles, A Castiglioni, K Yuen, Y Wang, *et al.*, TGFβ attenuates tumour response to PD-L1 blockade by contributing to exclusion of T cells. *Nature* **2018**;554:544–548.

71. J Lv, P Li, Mesothelin as a biomarker for targeted therapy. *Biomark Res* **2019**;7:18.

References to methods

1. P Priestley, J Baber, MP Lolkema, N Steeghs, E de Bruijn, C Shale, *et al.*, Pan-cancer whole-genome analyses of metastatic solid tumours. *Nature* **2019**;575:210–216.

2. M Casparie, ATMG Tiebosch, G Burger, H Blauwgeers, A van de Pol, JHJM van Krieken, *et al.*, Pathology Databanking and Biobanking in The Netherlands, a Central Role for PALGA, the Nationwide Histopathology and Cytopathology Data Network and Archive. *Cell Oncol* **2007**;29:19.

3. H Li, R Durbin, Fast and accurate short read alignment with Burrows-Wheeler transform. *Bioinformatics* **2009**;25:1754–1760.

4. A McKenna, M Hanna, E Banks, A Sivachenko, K Cibulskis, A Kernytsky, *et al.*, The genome analysis toolkit: A MapReduce framework for analyzing next-generation DNA sequencing data. *Genome Res* **2010**;20:1297–1303.

5. GA Van der Auwera, MO Carneiro, C Hartl, R Poplin, G del Angel, A Levy-Moonshine, *et al.*, From fastQ data to high-confidence variant calls: The genome analysis toolkit best practices pipeline. *Curr Protoc Bioinforma* **2013**;43:11.10.1-11.10.33.

6. CT Saunders, WSW Wong, S Swamy, J Becq, LJ Murray, RK Cheetham, Strelka: Accurate somatic small-variant calling from sequenced tumor-normal sample pairs. *Bioinformatics* **2012**;28:1811–1817.

7. LF van Dessel, J van Riet, M Smits, Y Zhu, P Hamberg, MS van der Heijden, *et al.*, The genomic landscape of metastatic castration-resistant prostate cancers reveals multiple distinct genotypes with potential clinical impact. *Nat Commun* **2019**;10:1–13.

8. W McLaren, L Gil, SE Hunt, HS Riat, GRS Ritchie, A Thormann, *et al.*, The Ensembl Variant Effect Predictor. *Genome Biol* **2016**;17:122.

9. X Liu, C Wu, C Li, E Boerwinkle, dbNSFP v3.0: A One-Stop Database of Functional Predictions and Annotations for Human Nonsynonymous and Splice-Site SNVs. *Hum Mutat* **2016**;37:235–241.

10. M Lek, KJ Karczewski, E V. Minikel, KE Samocha, E Banks, T Fennell, *et al.*, Analysis of protein-coding genetic variation in 60,706 humans. *Nature* **2016**;536:285–291.

11. D Cameron, J Baber, C Shale, A Papenfuss, JE Valle-Inclan, N Besselink, *et al.*, GRIDSS, PURPLE, LINX:

Unscrambling the tumor genome via integrated analysis of structural variation and copy number. *bioRxiv* **2019**;1:781013.

12. CH Mermel, SE Schumacher, B Hill, ML Meyerson, R Beroukhim, G Getz, GISTIC2.0 facilitates sensitive and confident localization of the targets of focal somatic copy-number alteration in human cancers. *Genome Biol* **2011**;12:R41.

13. I Martincorena, KM Raine, M Gerstung, KJ Dawson, K Haase, P Van Loo, *et al.*, Universal Patterns of Selection in Cancer and Somatic Tissues. *Cell* **2017**;171:1029-1041.e21.

14. SA Roberts, MS Lawrence, LJ Klimczak, SA Grimm, D Fargo, P Stojanov, *et al.*, An APOBEC cytidine deaminase mutagenesis pattern is widespread in human cancers. *Nat Genet* **2013**;45:970–976.

15. PJ Stephens, PS Tarpey, H Davies, P Van Loo, C Greenman, DC Wedge, *et al.*, The landscape of cancer genes and mutational processes in breast cancer. *Nature* **2012**;486:400–404.

16. N Bolli, H Avet-Loiseau, DC Wedge, P Van Loo, LB Alexandrov, I Martincorena, *et al.*, Heterogeneity of genomic evolution and mutational profiles in multiple myeloma. *Nat Commun* **2014**;5:2997.

17. G Gundem, P Van Loo, B Kremeyer, LB Alexandrov, JMC Tubio, E Papaemmanuil, *et al.*, The evolutionary history of lethal metastatic prostate cancer. *Nature* **2015**;520:353–357.

18. SC Dentro, DC Wedge, P Van Loo, Principles of Reconstructing the Subclonal Architecture of Cancers. *Cold Spring Harb Perspect Med* **2017**;7:a026625.

19. JG Tate, S Bamford, HC Jubb, Z Sondka, DM Beare, N Bindal, *et al.*, COSMIC: The Catalogue Of Somatic Mutations In Cancer. *Nucleic Acids Res* **2019**;47:D941–D947.

20. F Blokzijl, R Janssen, R van Boxtel, E Cuppen, MutationalPatterns: Comprehensive genome-wide analysis of mutational processes. *Genome Med* **2018**;10:33.

21. LB Alexandrov, J Kim, NJ Haradhvala, MN Huang, AW Tian Ng, Y Wu, *et al.*, The repertoire of mutational signatures in human cancer. *Nature* **2020**;578:94–101.

22. M Petljak, LB Alexandrov, JS Brammeld, S Price, DC Wedge, S Grossmann, *et al.*, Characterizing Mutational Signatures in Human Cancer Cell Lines Reveals Episodic APOBEC Mutagenesis. *Cell*

2019;176:1282-1294.e20.

23. L Angus, M Smid, SM Wilting, J van Riet, A Van Hoeck, L Nguyen, *et al.*, The genomic landscape of metastatic breast cancer highlights changes in mutation and signature frequencies. *Nat Genet* **2019**;51:1450–1458.

24. S Christensen, B Van der Roest, N Besselink, R Janssen, S Boymans, JWM Martens, *et al.*, 5-Fluorouracil treatment induces characteristic T>G mutations in human cancer. *Nat Commun* **2019**;10:1–11.

25. MD Wilkerson, DN Hayes, ConsensusClusterPlus: A class discovery tool with confidence assessments and item tracking. *Bioinformatics* **2010**;26:1572–1573.

26. R Gaujoux, C Seoighe, A flexible R package for nonnegative matrix factorization. *BMC Bioinformatics* **2010**;11:367.

27. I Cortés-Ciriano, JJK Lee, R Xi, D Jain, YL Jung, L Yang, *et al.*, Comprehensive analysis of chromothripsis in 2,658 human cancers using whole-genome sequencing. *Nat Genet* **2020**;52:331–341.

28. L Nguyen, J Martens, A van Hoeck, E Cuppen, Pan-cancer landscape of homologous recombination deficiency. *bioRxiv* **2020**; https:/doi.org/10.1101/2020.01.13.905026.

29. F Sanchez-Vega, M Mina, J Armenia, WK Chatila, A Luna, KC La, *et al.*, Oncogenic Signaling Pathways in The Cancer Genome Atlas. *Cell* **2018**;173:321-337.e10.

30. WJ Leonard, Role of JAK kinases and stats in cytokine signal transduction. *Int J Hematol* **2001**;73:271–277.

31. M Griffith, NC Spies, K Krysiak, JF McMichael, AC Coffman, AM Danos, *et al.*, CIViC is a community knowledgebase for expert crowdsourcing the clinical interpretation of variants in cancer. *Nat Genet* **2017**;49:170–174.

32. D Chakravarty, J Gao, S Phillips, R Kundra, H Zhang, J Wang, *et al.*, OncoKB: A Precision Oncology Knowledge Base. *JCO Precis Oncol* **2017**;1:1–16.

33. D Tamborero, C Rubio-Perez, J Deu-Pons, MP Schroeder, A Vivancos, A Rovira, *et al.*, Cancer Genome Interpreter annotates the biological and clinical relevance of tumor alterations. *Genome Med*

2018;10:25.

34.   R Nakato, K Shirahige, Recent advances in ChIP-seq analysis: from quality management to whole-genome annotation. *Brief Bioinform* **2017**;18:279–290.

35.   AM Bolger, M Lohse, B Usadel, Trimmomatic: a flexible trimmer for Illumina sequence data. *Bioinformatics* **2014**;30:2114–2120.

36.   A Dobin, CA Davis, F Schlesinger, J Drenkow, C Zaleski, S Jha, *et al.*, STAR: Ultrafast universal RNA-seq aligner. *Bioinformatics* **2013**;29:15–21.

37.   J Harrow, A Frankish, JM Gonzalez, E Tapanari, M Diekhans, F Kokocinski, *et al.*, GENCODE: The reference human genome annotation for the ENCODE project. *Genome Res* **2012**;22:1760–1774.

38.   A Tarasov, AJ Vilella, E Cuppen, IJ Nijman, P Prins, Sambamba: Fast processing of NGS alignment formats. *Bioinformatics* **2015**;31:2032–2034.

39.   L Wang, J Nie, H Sicotte, Y Li, JE Eckel-Passow, S Dasari, *et al.*, Measure transcript integrity using RNA-seq data. *BMC Bioinformatics* **2016**;17:58.

40.   L Wang, S Wang, W Li, RSeQC: quality control of RNA-seq experiments. *Bioinforma Oxford Engl* **2012**;28:2184–2185.

41.   Y Liao, GK Smyth, W Shi, FeatureCounts: An efficient general purpose program for assigning sequence reads to genomic features. *Bioinformatics* **2014**;30:923–930.

42.   B Li, CN Dewey, "RSEM: Accurate transcript quantification from RNA-seq data with or without a reference genome" in *Bioinformatics: The Impact of Accurate Quantification on Proteomic and Genetic Analysis and Research*, p. 323.

43.   F Ramírez, DP Ryan, B Grüning, V Bhardwaj, F Kilpert, AS Richter, *et al.*, deepTools2: a next generation web server for deep-sequencing data analysis. *Nucleic Acids Res* **2016**;44:W160–W165.

44.   MI Love, W Huber, S Anders, Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome Biol* **2014**;15:550.

45.   A Kamoun, A de Reyniès, Y Allory, G Sjödahl, A Gordon Robertson, R Seiler, *et al.*, A Consensus

Molecular Classification of Muscle-invasive Bladder Cancer. *Eur Urol* **2019**;77:420–433.

46.   AG Robertson, J Kim, H Al-Ahmadie, J Bellmunt, G Guo, AD Cherniack, *et al.*, Comprehensive Molecular Characterization of Muscle-Invasive Bladder Cancer. *Cell* **2017**;171:540-556.e25.

47.   T Powles, M Kockx, A Rodriguez-Vida, I Duran, SJ Crabb, MS Van Der Heijden, *et al.*, Clinical efficacy and biomarker analysis of neoadjuvant atezolizumab in operable urothelial carcinoma in the ABACUS trial. *Nat Med* **2019**;25:1706–1714.

48.   S Mariathasan, SJ Turley, D Nickles, A Castiglioni, K Yuen, Y Wang, *et al.*, TGFβ attenuates tumour response to PD-L1 blockade by contributing to exclusion of T cells. *Nature* **2018**;554:544–548.

49.   T Borggrefe, F Oswald, The Notch signaling pathway: Transcriptional regulation at Notch target genes. *Cell Mol Life Sci* **2009**;66:1631–1646.

50.   H van Ooijen, M Hornsveld, C Dam-de Veen, R Velter, M Dou, W Verhaegh, *et al.*, Assessment of Functional Phosphatidylinositol 3-Kinase Pathway Activity in Cancer Tissue Using Forkhead Box-O Target Gene Expression in a Knowledge-Based Computational Model. *Am J Pathol* **2018**;188:1956–1972.

51.   X Varelas, The hippo pathway effectors TAZ and YAP in development, homeostasis and disease. *Dev* **2014**;141:1614–1626.

52.   M Fischer, Census and evaluation of p53 target genes. *Oncogene* **2017**;36:3943–3956.

53.   H Kitamura, H Motohashi, NRF2 addiction in cancer cells. *Cancer Sci* **2018**;109:900–911.

54.   M Hartl, The quest for targets executing MYC-dependent cell transformation. *Front Oncol* **2016**;6:132.

55.   M-C Wagle, D Kirouac, C Klijn, B Liu, S Mahajan, M Junttila, *et al.*, A transcriptional MAPK Pathway Activity Score (MPAS) is a clinically relevant biomarker in multiple cancer types. *npj Precis Oncol* **2018**;2:1–12.

56.   PJ Murray, The JAK-STAT Signaling Pathway: Input and Output Integration. *J Immunol* **2007**;178:2623–2629.

57.   G Yu, QY He, ReactomePA: An R/Bioconductor package for reactome pathway analysis and

visualization. *Mol Biosyst* **2016**;12:477–479.

58.	G Sturm, F Finotello, F Petitprez, JD Zhang, J Baumbach, WH Fridman, *et al.*, Comprehensive evaluation of transcriptome-based cell-type quantification methods for immuno-oncology in *Bioinformatics*, (Oxford University Press), pp. i436–i445.

59.	F Finotello, C Mayer, C Plattner, G Laschober, D Rieder, H Hackl, *et al.*, Molecular and pharmacological modulators of the tumor immune contexture revealed by deconvolution of RNA-seq data. *Genome Med* **2019**;11:34.

60.	YE Jang, I Jang, S Kim, S Cho, D Kim, K Kim, *et al.*, ChimerDB 4.0: an updated and expanded database of fusion genes. *Nucleic Acids Res* **2019**;48:D817–D824.

61.	HX Dang, BS White, SM Foltz, CA Miller, J Luo, RC Fields, *et al.*, ClonEvol: clonal ordering and visualization in cancer sequencing. *Ann Oncol* **2017**;28:3076–3082.

62.	I Dunham, A Kundaje, SF Aldred, PJ Collins, CA Davis, F Doyle, *et al.*, An integrated encyclopedia of DNA elements in the human genome. *Nature* **2012**;489:57–74.

63.	Roadmap Epigenomics Consortium, A Kundaje, W Meuleman, J Ernst, M Bilenky, A Yen, *et al.*, Integrative analysis of 111 reference human epigenomes. *Nature* **2015**;518:317–329.

64.	CA Davis, BC Hitz, CA Sloan, ET Chan, JM Davidson, I Gabdank, *et al.*, The Encyclopedia of DNA elements (ENCODE): data portal update. *Nucleic Acids Res* **2018**;46:D794–D801.

65.	R Core Team, R Core Team (2017). R: A language and environment for statistical computing. *R Found Stat Comput Vienna, Austria URL http//wwwR-project.org/* **2017**;R Foundation for Statistical Computing.

Author contributions

Conceptualization: JAN, MR, HJGvdW, MPJL, JLB; Methodology: JAN, MR, HJGvdW, MPJL, and JLB; Software: JAN, HJGvdW, and JvR; Validation: MPJL, JLB, JvR ; Formal Analysis: JAN and HJGvdW; Investigation: MR, MSvdH, JV, NM, SvW, SO, **HMW,** ECZ, RdW, AAMvdV; Resources: HJGvdW, MPJL, JLB, MSvdH, JV, EC, NM, SvW, SO, **HMW,** ECZ, RdW, AAMvdW, MPJL and JLB; Data Curation: JAN, MR, HJGvdW, JvR, and EC ; Writing – Original Draft: JAN, MR, HJGvdW, MPJL and JLB; Writing – Review & Editing: JAN, MR, JvR, MSvdH, JV, EC, NM, SvW, SO, **HMW,** ECZ, RdW, AAMvdW, HJGvdW, MPJL and JLB; Visualization: JAN and MR; Supervision: JLB, HJGvdW and MPJL; Project Administration: JAN, MR, HJGvdW, MPJL and JLB; Funding Acquisition: JLB , HJGvdW, and MPJL

Competing Interests statement

Michiel S. van der Heijden has received research support from Bristol-Myers Squibb, AstraZeneca and Roche, and consultancy fees from Bristol-Myers Squibb, Merck Sharp & Dohme, Roche, AstraZeneca, Seattle Genetics and Janssen (all paid to the Netherlands Cancer Institute). Niven Mehra has received research support from Astellas, Janssen, Pfizer, Roche and Sanofi Genzyme, and consultancy fees from Roche, MSD, BMS, Bayer, Astellas and Janssen (all paid to the Radboud University Medical Center). Sjoukje F. Oosting has received research support from Celldex and Novartis (both paid to the University Medical Center Groningen). Hans M.