Molecular characterization reveals genomic and transcriptomic subtypes of metastatic urothelial carcinoma

J. Alberto Nakauma-González[1,2,3,^], Maud Rijnders[3,^], Job van Riet[1,2,3], 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^]

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.

Corresponding author

Joost Boormans, MD, PhD

Dr. Molewaterplein 40, 3015 GD Rotterdam, The Netherlands

j.boormans@erasmusmc.nl

Direct dial: 06 144 269 75

Word count: 10.500 words including methods section

5470 words excluding methods section

Number of figures: 7 main figures, 20 supplementary figures

Number of tables: 1 supplementary table file

Abstract

Background: Molecular characterization of primary urothelial carcinoma (UC) revealed molecular subtypes with different genomic, transcriptomic, and clinicopathological characteristics, which might guide therapeutic decision making. A comprehensive molecular characterization of metastatic UC (mUC), however, is currently lacking in the literature. Because of the lethality of mUC, with few therapeutic options available for patients, a multi-omics characterization of mUC could aid to improve patient selection for new and existing therapies.

Methods: To define the molecular landscape of mUC and to identify potential targets for therapy, we performed whole genome DNA sequencing on fresh-frozen metastatic tumor biopsies of 116 mUC patients, and mRNA sequencing on 90 matched biopsies.

Results: Hierarchical clustering based on mutational signatures revealed two major genomic subtypes. The most prevalent subtype (67%) consisted almost exclusively of tumors with high APOBEC mutagenesis. APOBEC mutagenesis was detected in 91% of the samples, and appeared to be an ongoing process in mUC based on analysis of eight patients from whom serial biopsies were obtained during treatment. Contrary to the overall distribution of mutations, APOBEC associated mutations occurred throughout the genome, and independently of predicted accessible or transcribed genomic regions, suggesting that these mutations were generated during replication. Transcriptomic analysis revealed five mRNA-based subtypes: two luminal subtypes (40%), a stroma-rich (24%), basal/squamous (23%), and non-specified subtype (12%). The transcriptomic subtypes were different regarding driver gene alterations (e.g. *ELF3* and *TSC1*), gene amplifications (*NECTIN4* and *PPARG*), pathway activity, and immune cell infiltration. By integrating the genomic and transcriptomic data, potential therapeutic options per transcriptomic subtype and individual patient were proposed.

Conclusions: This study expands our knowledge on the molecular landscape of mUC, and serves as a reference for subtype-oriented and patient-specific research on the etiology of mUC, and for novel drug development.

Trial registration: The mUC cohort studied here is part of the Netherlands nationwide study of the center for personalized cancer treatment consortium (CPCT-02 Biopsy Protocol, NCT01855477), and the Drug Rediscovery Protocol (DRUP Trial, NCT02925234).

Keywords

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

Background

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 [1]. Comprehensive molecular profiling of UC has been restricted to NMIBC [2] and localized MIBC [3]. At the genomic level, NMIBC is characterized by frequent *FGFR3* and *PIK3CA* mutations, whereas *TP53* mutations are uncommon [1]. In MIBC, *TP53* is the most commonly mutated gene [4]. *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% [3]. 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 [5]. 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 [6]. Expanding the 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.

82 ***Methods***

83 **Patient cohort and study procedures**

84 Between 07 June 2012 up to and including 28 February 2019, patients with advanced or mUC (n = 210) from 23
85 Dutch hospitals (Fig. 1a, Fig. S1) who were scheduled for 1st or 2nd line palliative systemic treatment were
86 included in the Netherlands nationwide study of the Center for Personalized Cancer Treatment (CPCT)
87 consortium (CPCT-02 Biopsy Protocol, NCT01855477 [7]) and the Drug Rediscovery Protocol (DRUP Trial,
88 NCT02925234), which aimed to analyze the cancer genome and transcriptome of patients with advanced
89 cancer. The CPCT-02 and DRUP study protocols were approved by the medical ethics review board of the
90 University Medical Center Utrecht and the Netherlands Cancer Institute, respectively. Patients eligible for
91 inclusion were those aged ≥ 18 years old, with locally advanced or mUC, from whom a histological tumor
92 biopsy could be safely obtained, and whom had an indication for initiation of a new line of systemic treatment
93 with anti-cancer agents. Written informed consent was obtained from all participants prior to inclusion in the
94 trial; the studies comply with all relevant ethical regulations. Tumor biopsies and matched normal blood
95 samples were collected following a standardized procedure described by the Hartwig Medical Foundation
96 (HMF; https://www.hartwigmedicalfoundation.nl; [7]). Whole genome sequencing (WGS) was successfully
97 performed on DNA from freshly obtained biopsies from metastatic sites in 116 mUC patients (124 samples),
98 and matched RNA-sequencing (RNA-seq) data was available for 90 patients (97 samples; Fig. 1a). Patient
99 characteristics are described in Table S1.1. Biopsies were obtained from a safely accessible site, including
100 lymph nodes, liver, bone and other organs (Fig. 1b). In five patients, a tumor biopsy was obtained from the
101 primary bladder or upper urinary tract tumor as no safely accessible metastatic lesion was present. In two
102 patients, a biopsy was obtained from a local recurrence after cystectomy and nephrectomy, respectively (Table
103 S1.2). Sequential biopsies of a metastatic lesion taken at the time of clinical or radiological disease progression
104 from eight patients were additionally sequenced. This study extends the pan-cancer analysis of Priestley *et al.*,
105 2019, in which WGS (but not RNA-seq) data of 72 mUC patients included in the current cohort were initially
106 analyzed (Table S1.2).

107

108

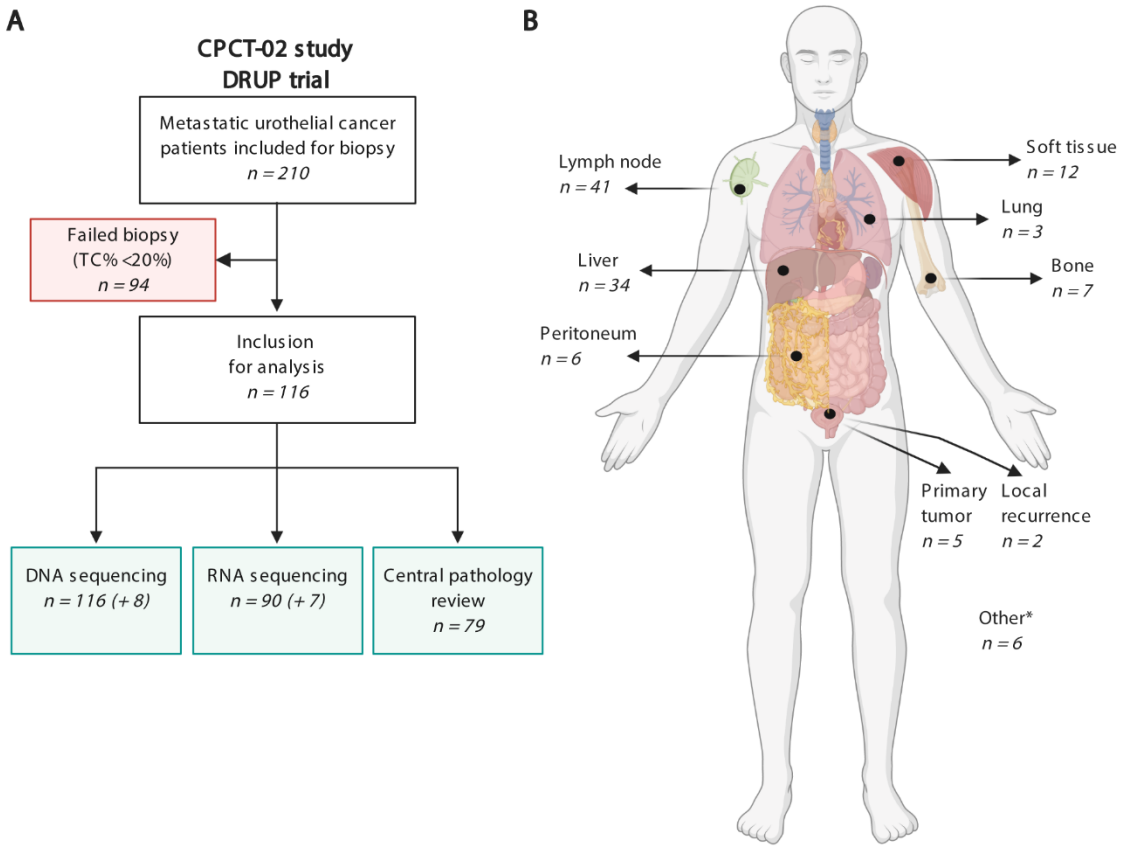


Figure 1 – Overview of the study design and biopsy sites of 116 patients with metastatic urothelial cancer

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 and the Drug Rediscovery Protocol (DRUP Trial (n = 210). Patients were excluded if the tumor cell percentage in the biopsy was <20%, resulting in WGS data and RNA-seq 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).

122 **Central pathology review**

123 Tumor tissue slides for central pathological revision of the diagnosis of UC was available for 79/116 patients.
124 Hematoxylin and eosin (H&E) stained slides from primary tumor tissues (cystectomy and transurethral
125 resection specimens of the bladder, n = 23 patients), metastatic tumor biopsies (n = 15 patients), or both (n =
126 41 patients) were requested from the Nationwide Network and Registry of Histo- and Cytopathology in the
127 Netherlands (PALGA) [8]. Tissue slides and corresponding pathological reports were provided anonymously. All
128 patient materials used for central pathology review were obtained within the CPCT-02 biopsy protocol, the
129 DRUP trial, or during routine patient care, and the use of these materials for research purposes was approved
130 by the medical ethics review board of the Erasmus University Medical Center, Rotterdam, the Netherlands
131 (MEC-2019-0188). H&E slides were reviewed by an expert genitourinary pathologist (LLB), and used for re-
132 evaluation of the diagnosis and description of aberrant histology (Tables S1.3 and S1.4). Tumors were classified
133 as pure UC (n = 66), or predominant UC with variant histology (n = 9 squamous, n = 3 neuro-endocrine, n = 1
134 micropapillary UC), and pure squamous cell bladder carcinomas (n = 3). In patients for whom both the primary
135 and the metastatic tumor biopsy was available for review, the highest grade (WHO 1973 classification) was
136 assigned, and presence of aberrant histology in one of the tissue samples was considered as positive.

137

138 **Whole-genome sequencing and analysis**

139 ***Whole-genome DNA sequencing, alignment and data processing***

140 Sufficient amount of DNA (50-200 ng) was extracted from fresh-frozen tumor tissue and blood samples
141 following standard protocols from Qiagen. DNA was fragmented by sonication for NGS Truseq library
142 preparation and sequenced paired-end reads of 2x150 bases with the Illumina HiSeqX platform. Alignment,
143 somatic alterations, ploidy, sample purity and copy numbers estimations were performed as previously
144 described [7]. WGS was aligned to the human reference genome GRCH37 with BWA-mem v.0.7.5a [9], and
145 duplicate reads were marked for filtering. Indels were realigned using GATK IndelRealigner v3.4.46 [10].
146 Recalibration of base qualities for single nucleotide variants (SNVs) and small insertions and deletions (Indels)
147 was performed with GATK BQSR [11], and SNV and Indel variants were evaluated with Strelka v.1.0.14 [12]
148 using matched blood WGS as normal reference (Table S1.5). Somatic mutations were further annotated with

149 Ensembl Variant Effect Predictor (VEP, v99, cache 99_GRCh37) [13] using GENCODE v33 in combinations with
150 the dbNSFP plugin v3.5 hg19 [14] for gnomAD [15] population frequencies. SNVs, Indels and multiple
151 nucleotide variants (MNVs) variants were removed if the following filters were not passed: default Strelka
152 filters (PASS-only), gnomAD exome (ALL) allele frequency < 0.001, gnomAD genome (ALL) < 0.005 and number
153 of reads < 3. In addition, structural variants (SVs) and copy number changes were estimated using GRIDSS,
154 PURPLE and LINX suit v2.25 [16]. SVs that passed the default QC filters (PASS-only) and Tumor Allele Frequency
155 (TAF) ≥ 0.1 were annotated as "somatic SVs" if there was overlap with coding region. Mean read coverages of
156 tumor and reference samples were estimated using Picard Tools v1.141 (CollectWgsMetrics) based on GRCh37
157 (https://broadinstitute.github.io/picard/). Genomic and coding tumor mutational burden (TMB; mutations per
158 megabase pair (Mbp)) were calculated considering SNVs, Indels and MNVs (Table S1.5). The total number of
159 somatic mutations in coding region was divided by 28.71 Mbp (protein-coding region size) and in the whole
160 genome by 2,858.67 Mbp (genomic alignment size).

161

162 ***Detection of driver genes using dN/dS ratios***

163 Cancer driver genes under strong positive selection were detected with dNdScv v0.0.0.9 [17]. This method
164 uses 192 mutation rates representing all combinations in trinucleotide context. Mutation rates of each gene
165 were corrected by the global mutation rate. The ratio of non-synonymous over synonymous mutations was
166 calculated with maximum-likelihood methods, and statistical significance was estimated. Genes with either
167 qglobal_cv ≤ 0.05 or qallsubs_cv ≤ 0.05 were considered drivers of mUC (Table S1.6 and 1.7).

168

169 ***Detection and characterization of recurrent copy number alterations***

170 Ploidy and copy number alterations (CNAs) were estimated as described by Priestley *et al.* 2019; and following
171 the pipeline described by van Dessel *et al.*, 2019. Recurrent focal and broad CNAs were estimated with
172 GISTIC2.0 v2.0.23 [19]. CNAs were classified as shallow or deep according to the threshold in GISTIC2 calls.

173 Significant recurrent focal CNAs were identified when q ≤ 0.05 and annotated with genes overlapping these

174 regions, which were considered drivers (Table S1.7 and 1.8).

175

176 ***APOBEC enrichment and mutagenesis***

177 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

178 these mutations, the total number of APOBEC mutations was estimated by counting all mutations in TCW

179 (WGA) context (TCW^{Cmut}), where W = A or T. The total number of TCW (WGA) motifs and total C (G)

180 nucleotides in the hg19 reference genome were also estimated ($TCW^{context}$ and $C^{context}$, respectively). Using this

181 information and following Roberts *et al.*, 2013 , a contingency table was constructed; one-sided Fisher's exact

182 test was applied to calculate the overrepresentation of APOBEC mutations. P-values were Benjamini-Hochberg

183 corrected. Tumors with adjusted p-values lower than 0.01 were considered APOBEC enriched.

184 The magnitude of APOBEC enrichment *E* was estimated as [20]

185 $$E = \frac{\mathrm{TCW}^{Cmut} \cdot \mathrm{C}^{context}}{\mathrm{TCW}^{context} \cdot \mathrm{C}^{mut(C>T,C>G)}}. \quad (1)$$

186 APOBEC enriched tumors (always $E > 1$) were classified as high APOBEC mutagenesis when $E \geq 2$, and as low

187 APOBEC mutagenesis when $E < 2$. Tumors without APOBEC enrichment were considered tumors with no

188 APOBEC mutagenesis (Table S1.9).

189 It has been shown that mutations caused by APOBEC3A and APOBEC3B are distinguishable at the tetra-

190 nucleotide context [21]. Mutations in the YTCA (Y = T or C) context have been related to APOBEC3A, while

191 mutations in the RTCA (R = G or A) context are attributed to APOBEC3B. Fold enrichment of C>T and C>G

192 mutations in the tetra YTCA and RTCA context was calculated with equation (1), using the corresponding tetra-

193 nucleotide context (Table S1.9).

194

195 ***Clonality fraction estimation***

196 Mutations start as a single copy in the DNA, and multiple copies of the mutated nucleotide may appear if
197 affected by CNAs events. Correcting for tumor purity and CNA, the number of copies n_{SNV} of each SNV was
198 calculated as follows [22]

199 $$n_{SNV} = \frac{f_m}{p}[pC_t + (1-p)C_h], \quad (2)$$

200 where f_m is the relative frequency of the mutant variant reads, p is the tumor purity, C_t is the local copy
201 number affecting a particular SNV, and C_h is the healthy copy number (two for autosomes and one for
202 allosomes).

203 Equation (2) is equivalent to the cancer cell fraction (CCF) with $n_{SNV} \approx 1$ in haploid and heterozygous-diploid
204 regions; i.e., the fraction of tumor cells carrying a particular mutation. For regions with CNAs, $n_{SNV} > 1$, we must
205 estimate the fraction of cancer cells carrying a particular SNV. As described previously [23], we assume that all
206 SNVs are present in the major copy number C_M; hence $n_{SNV} \leq C_M$ will include mutations that were acquired
207 after copy number change events or present only in the minor copy number. Given the number of reference
208 and mutant reads, and assuming binomial distribution, we estimated the expected number of allelic copies
209 (n_{chr}) carrying the observed SNV resulting from f_m values when the mutation is present in 1, 2, 3,…, N_{chr} allelic
210 copies. In some cases (sequencing noise) $n_{SNV} > C_M$, which was corrected with $N_{chr} = max(C_M, n_{SNV})$. We also
211 corrected each f_m value with normal cell contamination – multiplying it by p. The resulting estimated n_{chr} with
212 the maximum likelihood serves to calculate the CCF as n_{SNV}/n_{chr}.

213 Dirichlet process from DPClust v2.2.8 (https://github.com/Wedge-lab/dpclust) with 250 iterations and 125
214 burn in iterations was applied to the CCF distribution to estimate the fraction of clonal and subclonal SNVs per
215 tumor. Multiple distributions (clusters) were obtained, representing different cancer cell populations. The
216 mean of the distributions was used to classify clusters of SNVs as clonal or subclonal. Clusters of SNVs with
217 mean distribution > 0.8 were considered clonal (Table S1.5).

218

219 ***Mutational signatures and genomic subtypes***

220 The mutational pattern of each sample was established by categorizing SNVs according to their 96-
221 trinucleotide context. The contribution of each of the 67 mutational signatures from COSMIC v3 (as deposited
222 on May 2019) [24] was subsequently estimated with MutationalPatterns v1.4.2 (Table S1.10) [25]. To reduce
223 the noise attributed to mutational signatures with very low contribution, mutational signatures were grouped
224 into 26 proposed etiology categories (Table S1.11) derived from Alexandrov *et al.,* 2020, Petljak *et al.,* 2019,
225 Angus *et al.,* 2019 and Christensen *et al.,* 2019. All 26 proposed etiology contributions were used, and
226 hierarchical clustering was applied on 1-Pearson's correlation coefficient, 80% resampling and 1,000 iterations
227 using ConsensusClusterPlus v1.48.0 [30]. Considering average stability of each cluster and the cluster size
228 (favoring large clusters) after each partition, samples were grouped into five distinct clusters.

229 Independently, mutational patterns were deconvoluted to estimate *de novo* mutational signatures. Non-
230 negative Matrix Factorization from the NMF R package v0.21.0 was used with 1000 iterations [31]. Evaluating
231 different metrics provided by the NMF R package (high cophenetic correlation coefficient, high dispersion
232 coefficient, high silhouette consensus, high sparseness basis and low sparseness coefficients), seven *de novo*
233 signatures were recovered from the mutational patterns. Cosine similarity was applied to compare the *de novo*
234 signatures with mutational signatures from COSMIC v3.

235

236 ***Detection of chromothripsis***

237 Genomic catastrophic-like events, such as chromothripsis, were detected with Shatterseek v0.4 [32] using
238 default parameters. Absolute copy numbers (as derived by PURPLE) were rounded to the nearest integer; only
239 structural variants with TAF ≥ 0.1 at either end of the breakpoint were considered, and chrY was excluded.

240 Applying the filters suggested by Cortés-Ciriano, et al. [32], we identified 220 chromothripsis events suggesting
241 an enrichment in mUC compared to primary UC from the PCAWG dataset [31] (76% vs 48%, p = 0.011).
242 Inspecting manually these events, we concluded that a more stringent filter should be used to reduce the rate
243 of false positive events. The following filters were applied and chromothripsis was considered when: a) the
244 number of intra-chromosomal structural variants ≥ 25; b) the maximum number of oscillating CN segments

245 with two states ≥ 7 or with three states ≥ 14; c) the size of the chromothripsis event ≥ 20 Mbp; d) random

246 distribution of breakpoints $p \leq 0.05$; and e) *chromosomal breakpoint enrichment* $p \leq 0.05$. All 220

247 chromothripsis events are summarized in Table S1.12, indicating those detected by the stringent filtering,

248 which we refer to as chromothripsis throughout the text.

249

250 ***MicroSatellite Instabilty (MSI) status***

251 As previously described [7], MSI status was determined by estimating the MSI score as the number of indels

252 (length < 50 bp) per Mbp occurring in homopolymers of five or more bases, dinucleotide, trinucleotide and

253 tetranucleotide sequences of repeat count above five. Tumors with MSI score > 4 were considered MSI

254 positive (Table S1.13).

255

256 ***Detection of homologous recombination (HR) deficiency***

257 The Classifier for Homologues Recombination Deficiency (CHORD; v2.0) with default parameters was used to

258 identify tumors with HR proficiency and deficiency [33]. Four samples had very high number of indels

259 corresponding with MSI samples and were discarded for the HR deficiency analysis (Table S1.14).

260

261 ***Detection of kataegis***

262 Following the method described by van Dessel *et al.* 2019, kataegis events were estimated using all SNVs. Each

263 chromosome was divided into segments (maximum 5000 segments) of five or more consecutive SNVs.

264 Segments were considered a kataegis event when the mean intermutational distance was ≤ 2000 bp (Table

265 S1.15). Events were considered APOBEC-driven when >60% of mutations were C>T or C>G mutations in TCW

266 context.

267

268 ***Mutational load across genomic regions***

269 The genome was divided in regions (bins) of one Mbp size. The number of SNVs was counted in each bin, and
270 the mean number of SNVs was estimated from the entire cohort. These values represented the average
271 SNVs/Mbp reflecting the mutational load in each genomic region. The average SNVs/Mbp was smoothed by
272 applying a moving average with k = 9 bins. This approach was used per sample and for mean values from the
273 entire cohort.

274

275 ***Genomic alteration of oncogenic pathways***

276 Eleven oncogenic pathways were analyzed for somatic alterations. The list of genes was modified from
277 Sanchez-Vega *et al.,* 2018 and Leonard, 2001 (Table S1.16). Altered pathways were defined when at least one
278 of the pathway-genes was affected by any somatic mutation (SNV, Indel, MNV, SV or deep CNA; excluding
279 synonymous mutations).

280

281 ***Inventory of clinically-actionable somatic alterations and putative therapeutic targets***

282 Current clinical relevance of somatic alterations in relation to putative treatment options or resistance
283 mechanisms and trial eligibility was determined based upon the following databases: CiViC [36] (Nov. 2018),
284 OncoKB [37] (Nov. 2018), CGI [38] (Nov. 2018) and the iClusion (Dutch) clinical trial database (Sept. 2019,
285 Rotterdam, the Netherlands). The databases were aggregated and harmonized using the HMF knowledgebase-
286 importer (v1.7; https://github.com/hartwigmedical/hmftools/tree/master/knowledgebase-importer).
287 Subsequently, we curated the linked putative treatments and selected treatments for which level A (biomarker
288 for approved therapy or in guidelines) or level B (biomarker on strong biological evidence or used in clinical
289 trials) evidence was available. Genomic alterations that confer resistance to certain therapies were excluded
290 from the analysis. Treatment strategies including anti-hormonal therapy (as used for breast and prostate
291 cancer), surgical resection, or radioiodine uptake therapy were excluded. Furthermore, closed trials (according
292 to www.clinicaltrials.gov), and trials with only pediatric patients or patients with hematological malignancies
293 were excluded (Table S1.17). The data base was complemented with *FGFR3* and *NTRK2* gene fusions (at RNA

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

300

301 ***DNA accessibility estimation (ChIPseq)***

302 All available ChIPseq data for healthy urinary bladder (H3K4me1, H3K4me3, H3K36me3 and H3K27ac) were
303 downloaded from the ENCODE portal (https://www.encodeproject.org) to our local server. The *bed.gz* files
304 were imported with *narrowPeak* format for analysis. The signal of each experiment was divided in regions of
305 one Mbp, and a moving average with k = 9 bins was applied. The scale of the signal was normalized; hence the
306 sum of all regions in a chromosome is one. This step was taken to compensate for the bias observed in peak
307 intensity signals across different chromosomes, possible due to technical issues in the ChIPseq technology, e.g.
308 hyper-ChIPable regions or mappability [39].

309 High DNA accessible regions (open chromatin) were determined as such if the ChIPseq signal value of the
310 region was above the median. Otherwise, the region was considered as low DNA accessible (condensed
311 chromatin). This procedure was applied on each chromosome.

312

313 **Whole-transcriptome sequencing and analysis**

314 ***RNA-sequencing, alignment and data pre-processing***

315 Total RNA was extracted using the QIAGEN QIAsymphony kit (Qiagen, FRITSCH GmbH, Idar-Oberstein,
316 Germany). Samples with a minimum of 100 ng total RNA were sequenced according to the manufacturer's
317 protocols. Paired-end sequencing of RNA was performed on the Illumina NextSeq 550 platform (2x75bp) and
318 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 [40] 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 [41] with genomic annotations from GENCODE hg19 release 30 [42]. 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 [43]. For each genome-aligned sample, the uniformity of read distributions across transcript lengths was assessed using tin.py v2.6.6 [44] from the RseQC library v3.0.0 [45]. FeatureCounts v1.6.3 [46] was applied to count the number of overlapping reads per gene using genomic annotations from GENCODE (hg19) release 30 [42]; only primary (uniquely mapped) reads were counted per exon and summarized per gene:

```
featureCounts -T 50 -t exon -g gene_id --primary -p -s 2 -a <gencode> -o <output> <genomic BAMs>
```

RSEM v1.3.1 [47] was applied to quantify RNA expression into transcripts per million (TPM) values using transcript annotations from GENCODE (hg19) release 30 [42]:

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

343 ***Transcriptome expression data mapped to genomic regions***

344 MultiBamSummary from deepTools v1.30.0 [48] was used to read BAM files and estimate number of reads in
345 genomic regions with a size of one Mbp. The average raw read count per Mbp was calculated, and a moving
346 average with k = 9 bins was applied. The scale of the read counts was normalized following the method for
347 DNA accessibility regions, and high transcriptional regions were defined as such when the expression value of
348 one region was above the median. This procedure was applied on each chromosome.

349

350 ***Transcriptomic subtypes: clustering samples by RNA-seq data***

351 Several methods have been proposed to classify bladder cancer into transcriptomic subtypes. In an attempt to
352 standardize the molecular profiling of bladder cancer, a consensus molecular classification was proposed for
353 MIBC based on RNA-seq data from 1750 patients [5]. This classifier was developed strictly for MIBC and is not
354 directly applicable to mUC [49]. Furthermore, this classifier was developed for samples derived from the same
355 organ carrying transcriptomic contamination of normal urothelial cells. In this study, biopsies were obtained
356 from metastatic sites leading to contamination with normal cells from multiple different organs for which no
357 correction was applied in the consensus classifier. Therefore, it was mandatory to perform *de novo* subtyping
358 in this study, which is described below.

359 Multiple methods were explored to correct for the bias of biopsy site, including batch-correction with DESeq2
360 [50], and a tissue-aware correction method developed by the Genotype-Tissue Expression (GTEx) project [51].
361 In both cases, transcripts from liver tissue were very dominant and clustered together in one stable cluster.
362 The tissue-specific transcript removal method described above was successfully able to correct for organ-
363 specific transcripts, and as a result samples were clustered based on transcriptomic features rather than
364 biopsy site.

365 Transcripts were normalized using DESeq2 v1.24.0 [50] with variance stabilizing transformation. Only highly
366 expressed mRNA with base mean above 100 was kept. The top 50% most variably expressed genes (6,410
367 transcripts) were used for clustering. To reduce the 'transcriptomic noise' introduced by normal cells of the
368 tissue from which the biopsy was taken, these transcripts were identified and excluded. Samples were grouped

369 according to their biopsy site: liver (n = 31), lymph node (n = 30), bone (n = 5), other (n = 23) and unknown
370 (n=1). Differential expression analysis was performed to compare tumors from a specific biopsy site (liver,
371 lymph node and bone) against all other tumors using DESeq2 with Wald test p-value estimation. Tissue-specific
372 transcripts with $\log_2$ Fold Change ($\log_2$FC) > 1.0 and Benjamini-Hochberg corrected p-value < 0.05 were
373 considered differentially expressed and identified as tissue-specific (Table S1.18). A total of 689 transcripts
374 were tissue-specific, and were removed from the data set.

375 The remaining 5,721 transcripts were grouped by hierarchical clustering with 1-Pearson's correlation
376 coefficient, 80% resampling and 1,000 iterations using ConsensusClusterPlus v1.48.0 [30]. The mean cluster
377 consensus value was obtained as a measure of cluster stability. Increasing the number of clusters will increase
378 the stability by creating smaller clusters. Taking this into account, the criteria for selecting five clusters was
379 based on cluster stability and cluster size by not allowing clusters with <5 samples (Table S1.19). Patients with
380 primary upper tract tumors did not cluster together as was observed for biopsy sites, instead they were
381 distributed across all different transcriptomic clusters (Fig. 5), suggesting that their influence on the clustering
382 was negligible.

383 To identify transcripts that contribute most to each cluster, we followed the same strategy used to identify
384 tissue-specific transcripts. The top five transcripts with the highest $\log_2$FC and with Benjamini-Hochberg
385 adjusted p-values lower than $1x10^{-5}$ were identified as the most overexpressed genes per cluster. Other
386 differentially expressed genes were included for their clinical relevance (*TGFB3*, *DDR2*, *PDGFRA*, *CD274* and
387 *TGFBR1*). All differentially expressed genes per cluster with adjusted $p < 1x10^{-5}$ and $\log_2$FC > 1 are listed in
388 Table S1.20.

389 To compare our classification system developed for mUC with the consensus classifier, all samples were
390 classified into one of six molecular classes identified in MIBC. All normalized transcripts (excluding biopsy
391 specific transcripts) were used as input for the consensus classifier of primary MIBC (v1.1.0) [5].

392

393 ***Phenotypic markers and signature score***

394 Marker genes for basal (*CD44, CDH3, KRT1, KRT14, KRT16, KRT5, KRT6A, KRT6B, KRT6C*), squamous (*DSC1,*
395 *DSC2, DSC3, DSG1, DSG2, DSG3, S100A7, S100A8*), luminal (*CYP2J2, ERBB2, ERBB3, FGFR3, FOXA1, GATA3,*
396 *GPX2, KRT18, KRT19, KRT20, KRT7, KRT8, PPARG, XBP1, UPK1A, UPK2*), neuroendocrine (*CHGA, CHGB, SCG2,*
397 *ENO2, SYP, NCAM1*), cancer-stem cell (*CD44, KRT5, RPSA, ALDH1A1*), EMT (*ZEB1, ZEB2, VIM, SNAI1, TWIST1,*
398 *FOXC2, CDH2*) and claudin (*CLDN3, CLDN7, CLDN4, CDH1, SNAI2, VIM, TWIST1, ZEB1, ZEB2*) were used for
399 signature scores [3]. Stroma (*FAP*), interferon, and CD8+ effector T cell (*IFNG, CXCL9, CD8A, GZMA, GZMB,*
400 *CXCL10, PRF1, TBX21*) markers were also included [52]. All normalized expression values were median
401 centered, and the mean expression of each group of genes was defined as signature score.

402

403 ***Pathway activity score***

404 Transcriptionally activated genes by the eleven canonical pathways analyzed in this study were used to
405 estimate pathway activity score. All normalized expression values were median centered, and the mean
406 expression of each group of genes was defined as activity score. Activity score was estimated for the TGFβ
407 pathway (*ACTA2, ACTG2, ADAM12, ADAM19, CNN1, COL4A1, CCN2, CTPS1, RFLNB, FSTL3, HSPB1, IGFBP3,*
408 *PXDC1, SEMA7A, SH3PXD2A, TAGLN, TGFBI, TNS1, TPM1*) [53], cell cycle pathway (*MKI67, CCNE1, BUB1,*
409 *BUB1B, CCNB2, CDC25C, CDK2, MCM4, MCM6, MCM2*) [52], WNT pathway (*EFNB3, MYC, TCF12, VEGFA*) [53],
410 Notch pathway (*HES1, HES5, HEY1*) [54], PI3K pathway (*AGRP, BCL2L11, BCL6, BNIP3, BTG1, CAT, CAV1, CCND1,*
411 *CCND2, CCNG2, CDKN1A, CDKN1B, ESR1, FASLG, FBXO32, GADD45A, INSR, MXI1, NOS3, PCK1, POMC,*
412 *PPARGC1A, PRDX3, RBL2, SOD2, TNFSF10*) [55], hippo pathway (*TAZ, YAP1*) [56], p53 pathway (*CDKN1A,*
413 *RRM2B, GDF15, SUSD6, BTG2, DDB2, GADD45A, PLK3, TIGAR, RPS27L, TNFRSF10B, TRIAP1, ZMAT3, BAX,*
414 *BLOC1S2, PGF, POLH, PPM1D, PSTPIP2, SULF2, XPC*) [57], Nrf2 pathway (*GCLM, NQO1, PHGDH, PSAT1, SHMT2*)
415 [58], MYC pathway (*TFAP4, BMP7, CCNB1, CCND2, CCNE1, CDC25A, CDK4, CDT1, E2F1, GATA4, HMGA1,*
416 *HSP90AA1, JAG2, CDCA7, LDHA, MCL1, NDUFAF2, MTA1, MYCT1, NPM1, ODC1, SPP1, PIN1, PTMA, PRDX3,*
417 *PRMT5, DNPH1, TFRC, EMP1, PMEL, C1QBP*) [59], RTK-RAS pathway (*SPRY2, SPRY4, ETV4, ETV5, DUSP4,*
418 *DUSP6, CCND1, EPHA2, EPHA4*) [60] and JAK-STAT pathway (*IRGM, ISG15, GATA3, FCER2, THY1, NFIL3, ARG1,*
419 *RETNLB, CLEC7A, CHIA, OSM, BCL2L1, CISH, PIM1, SOCS2, GRB10*) [61].

420 ***Pathway enrichment analysis***

421 All differentially expressed genes with Benjamini-Hochberg adjusted $p < 0.05$ and absolute $\log_2 FC > 1$ in each
422 transcriptomic subtype were used for pathway enrichment analysis. Using reactomePA v1.34 [62], the top ten
423 (sorted by Benjamini-Hochberg adjusted p-value) up- and down-regulated pathways were selected.

424

425 ***Immune cell infiltration***

426 To quantify immune cell fractions in each sample, we analyzed RSEM read counts of all transcripts with
427 immunedeconv v2.0.3 [63] using the quanTIseq method [64].

428

429 ***Detection of gene fusions***

430 In-frame gene fusions were detected at DNA level by the GRIDSS, PURPLE, LINX suite v2.25 [16], and reported
431 relevant if they appear in the ChimerDB 4.0 (Table S1.21) [65]. At RNA level, Arriba v2.0.0
432 (https://github.com/suhrig/arriba/) was used to infer gene fusion events with the option to discard known
433 false positives from a list provided by Arriba. High confidence fusions were retained, and only events where at
434 least one transcript is protein coding were kept (Table S1.22). Gene fusions previously identified by other
435 studies, mostly from the TCGA data, were identified with ChimerDB 4.0 [65]. All "deletion/read-through"
436 events were discarded as possible false positives unless they were supported by the ChimerDB 4.0 database.
437 Medium confidence fusions were included in the final list if one of the fused genes appeared in a high
438 confidence fusion event.

439

440 ***mRNA editing***

441 Jalili*, et al.* 2017 [67] identified hotspot mutations in the mRNA of *DDOST* and *CYFIP1* that are targeted by
442 APOBEC3A. The genomic position of these hotspot mutations reveals hairpin loop structures that are the ideal
443 substrate for APOBEC3A. Due to the short life-time of mRNA molecules, the presence of these hotspot

444 mutations reflects ongoing APOBEC mutagenesis. The proportion of C>U mutations in chr1:20981977 and
445 chr15:22999350 were estimated to identify the RNA-editing activity of APOEBC3A.

446

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

448 A second metastatic tumor biopsy was taken in eight patients from the same (n = 5) or a different (n = 3)
449 metastatic lesion, and analyzed by WGS (n = 8) and RNA-seq (n = 7). Patients with high APOBEC mutagenesis (n
450 = 5) tumors were all treated with pembrolizumab (one patient received consecutive lines of chemotherapy and
451 pembrolizumab). Two out of three patients with low APOBEC mutagenesis tumors were treated with
452 chemotherapy. All patients, except two, received systemic pre-treatment.

453 Each patient's first and second biopsies shared a high proportion of mutations (SNVs, Indels and MNVs),
454 confirming the clonal relation of the sampled sites. Dirichlet process from the DPClust v2.2.8 R package
455 (https://github.com/Wedge-lab/dpclust) with 250 iterations and 125 burn in iterations was applied to the CCF
456 distribution of paired-biopsies to estimate the subclonal (clusters) composition of each tumor. All unique
457 mutations in each biopsy were considered a subclone; only subclones with >5% of SNVs were considered
458 relevant. Small populations of subclones (<5% of SNVs) were merged to the nearest subclone. The
459 evolutionary tree was reconstructed following the *sum rule* [66].

460 The CCF of somatic mutations in the branches was lower than that in the trunk, suggesting that these
461 mutations are recently acquired mutations. To compare APOBEC mutagenesis between patients, the rate of
462 novel APOBEC associated mutations was estimated. Only unique SNVs from the second biopsy were kept, as
463 these somatic alterations probably correspond to new mutations acquired during the time frame between the
464 biopsies. As the time elapsed between the first and the second biopsy varied between tumors, we normalized
465 the number of recent APOBEC associated mutations by dividing the total over the number of days elapsed
466 between the biopsies. The value estimated is proportional to the mutation rate of APOBEC associated
467 mutations (mutations per day).

468 For seven tumors, RNA-seq data was available. Expression of *APOBEC3A* and *APOBEC3B* per patient represents
469 the mean normalized expression of the paired biopsies.

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

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

481 Transcript counts were normalized with DESeq2 [50] following the same procedure used for the mUC cohort.
482 All tumors were from primary UC, and organ-specific transcripts were not discarded. The consensus classifier
483 of primary MIBC [5] was applied to infer the transcriptomic subtype of each tumor.

484

485 **Code availability**

486 All custom code and scripts are available at https://bitbucket.org/ccbc/dr31_hmf_muc/ and
487 https://github.com/hartwigmedical/.

488

489 **Statistical analysis**

490 Several statistical tests were used in this study: Fisher's exact test, T-test, binomial test, Wald test for
491 differential expression analysis and logistic regression analysis, Wilcoxon signed-rank test, Wilcoxon rank-sum
492 test, Kruskal-Wallis test, and tests performed by dNdScv [17] and GISTIC2 [19]. In cases of multiple testing, p-
493 values were Benjamini-Hochberg corrected. The appropriate statistical test is mentioned in the text when
494 describing significance values. All statistical analyses were performed using the statistical computing and
495 graphics platform R v3.6.1 [69].

496 ***Results***

497 **Genomic landscape of mUC**

498 Analysis of WGS (mean coverage 106 X) and matched blood samples (mean coverage 38 X) identified a median
499 of 20,634 SNVs, 1,018 Indels and 175 MNVs (Fig. S2a). SNVs were more frequent in coding regions (7.63 SNVs
500 per Megabase pair; SNVs/Mbp) than in the whole genome (7.22 SNVs/Mbp; Wilcoxon signed-rank test $p =$
501 0.0024; Supplementary Fig. 2b). However, Indels and MNVs were less frequent in coding regions (Wilcoxon
502 signed-rank test $p < 0.001$ and $p = 0.0072$, respectively). Analysis of all SNVs revealed that 68% of all SNVs were
503 clonal with a median of 74% per tumor, and that 91% of the tumors were enriched for APOBEC associated
504 mutations (73% high and 18% low enrichment of APOBEC mutagenesis; Fig. 2a). The mean contribution of
505 APOBEC COSMIC signatures (SBS2 and SBS13) in tumors with high APOBEC mutagenesis enrichment was 52%
506 *versus* 15% in tumors with low APOBEC mutagenesis.

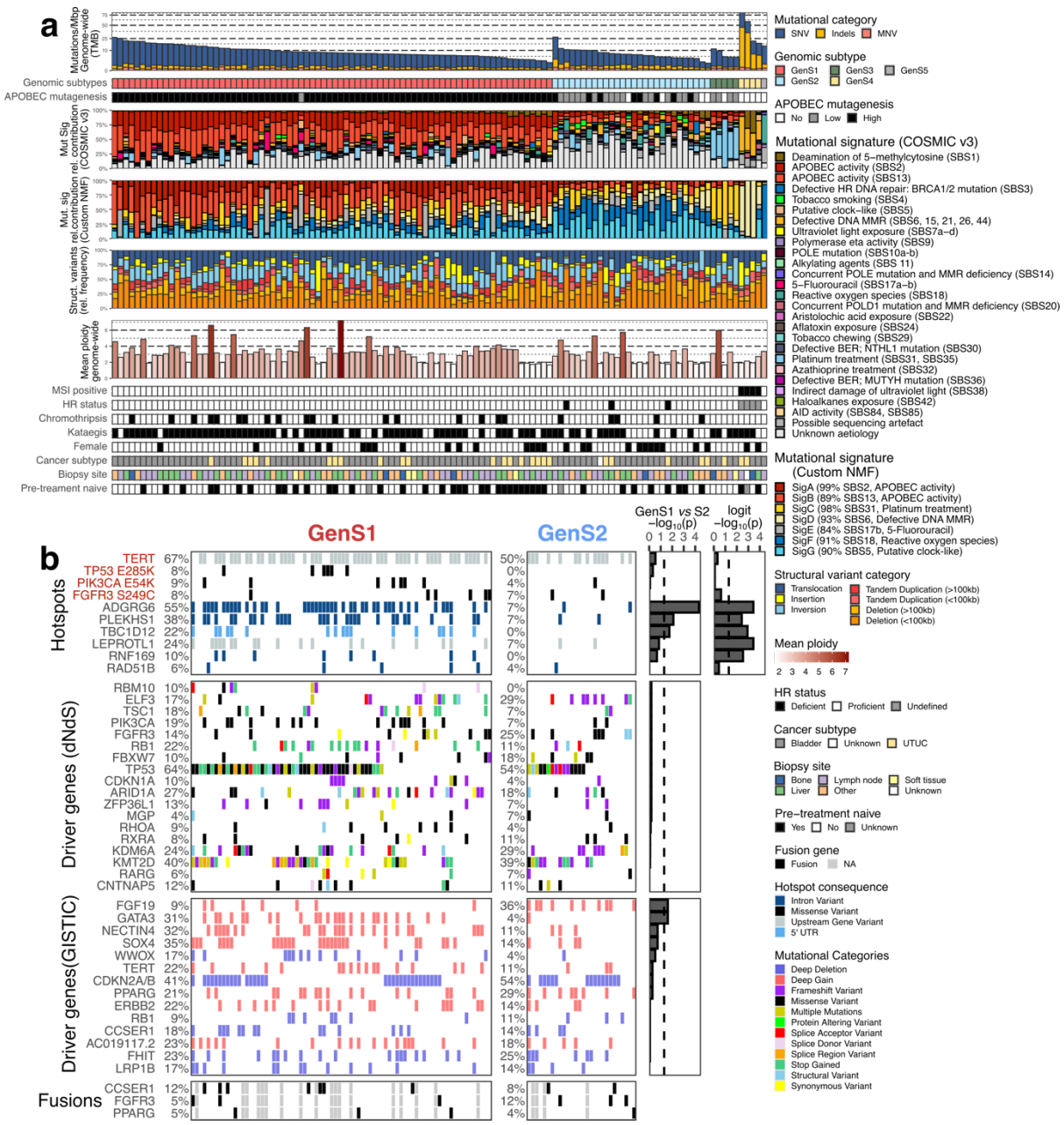


Figure 2 – Genomic landscape of metastatic urothelial carcinoma stratified by genomic subtypes

a) Whole-genome sequencing data from biopsy samples of metastatic urothelial carcinoma were classified into genomic subtypes by hierarchical consensus clustering of the relative contribution of COSMIC v3 mutational signatures [24] 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

516 tri-nucleotide context with NMF; Relative frequency of different types of structural variants; Mean
517 ploidy; Tumors with MSI; HR deficiency status; Samples with at least one chromothripsis event;
518 Samples with at least one kataegis event; Female patients; Origin of primary tumor; Metastatic site
519 from which a biopsy was obtained; Treatment-naïve patients.
520 b) Overview of recurrent hotspot mutations, driver genes and gene fusions for the genomic subtypes
521 GenS1 and GenS2. Name of genes affected by hotspot mutations in >5% of samples are displayed in
522 red when the hotspot had a COSMIC id. Significantly mutated genes were estimated by dNdScv [17];
523 all genes with $q < 0.05$ were considered driver genes. Recurrent focal copy number changes were
524 estimated by GISTIC2 [19]; genes in genomic regions with $q < 0.05$ were considered significant. Only
525 affected genes present in >10% of the samples are shown. Gene fusions were detected from RNA-seq
526 data. Benjamini-Hochberg adjusted p-values of Fisher's exact test (for hotspot mutations and GISTIC2)
527 and of logistic regression analysis corrected by mutational load (driver genes by dNdScv) are shown
528 on the right to reflect the significance of the difference between GenS1 and GenS2. An additional
529 logistic (logit) regression analysis was performed on hotspot mutations to show the linear relation
530 with the number of APOBEC associated mutations. Bars beyond the dashed line ($-\log_{10}(0.05)$) are
531 statistically significant.
532 Abbreviations: TMB = tumor mutational burden; Mbp = mutations per mega base pair; NMF = non-
533 Negative Matrix Factorization; MSI = microsatellite instability; HR = Homologous Recombination;
534 UTUC = upper tract urothelial carcinoma.

535

536 Genes harboring more mutations in their coding sequence than expected by random chance were analyzed
537 with dNdScv; the analysis revealed 18 significantly mutated genes (Table S1.6). These genes resembled those
538 reported for primary UC [3], although mUC lesions did harbor more somatic mutations in *TP53* than numbers
539 reported in TCGA (non-synonymous mutations and indels; 60% *vs* 49%, Fisher's exact test $p = 0.021$, Fig. S3).

540 SVs were common with a median of 259 (40,297 in total) per tumor. Deletion was the most frequent type of
541 SV with a median of 92 per tumor (Fig. S2d). The genes most frequently affected by SVs were *CCSER1* (13%)
542 and *AHR* (12%; Fig. S3). Chromothripsis, a complex event that produces SVs in which chromosomes are
543 shattered and rearranged, was detected in 20% of the tumors (Fig. S4).

544 Chromosomal arm and focal CNA were analyzed with GISTIC2. This revealed frequent deletion of chromosome
545 9 and amplification of chromosome 20 (Fig. S5a). In total, 49 genomic regions were significantly altered by
546 focal CNAs which included several oncogenic genes (Table S1.8). The most frequently amplified genes were
547 *SOX4* (28%), *GATA3* (22%), *PPARG* (22%), and ERBB2 (19%); the most frequently deleted genes were *CDKN2A/B*
548 (43%), *FHIT* (24%), *CCSER1* (17%) and *LRP1B* (17%), and also resembled those reported in primary UC [3] (Fig.
549 S3).

550 Hotspot mutations in driver genes concerned *FGFR3* S249C (8%), *PIK3CA* E54K (7%), and *TP53* E285K (5%).
551 Hotspot mutations in the *TERT* promoter were present in 62% of the tumors (Fig. S3; Table S1.23). Still, *TERT*
552 expression did not differ between tumors with and without hotspot mutations (Fig. S6b), in line with a
553 previous report [68]. However, differential gene expression analysis showed that tumors with hotspot
554 mutations in the *TERT* promoter had downregulation of genes related to the muscle contraction pathway (Fig.
555 S6a, Table S1.24). Furthermore, hotspot mutations were identified in non-coding regions of *ADGRG6 (40%)*,
556 *PLEKHS1 (28%)*, *LEPROTL1* (18%), and *TBC1D12* (15%; Fig. S3) with no apparent association with gene
557 expression and minimal transcriptomic effect (Fig. S6a-b). The hotspot areas of *ADGRG6*, *PLEKHS1* and
558 *TBC1D12* form hairpin loop structures in the DNA with specific tri-nucleotide sequences frequently mutated by
559 APOBEC enzymes (Fig. S6c). Unlike other known driver genes affected by hotspot mutations (*TERT*, *FGFR3*,
560 *PIK3CA* and *TP53*), these genes were not significantly affected by other somatic mutations in the coding region
561 or by CNAs, suggesting that hotspot mutations in *ADGRG6*, *PLEKHS1* and *TBC1D12* are likely passenger
562 hotspots attributed to APOBEC activity as theoretically predicted [69].

563 Gene fusion analysis performed at the transcriptomic level (Table S1.22) detected 1394 gene fusions, of which
564 10% were also reported in the TCGA cohort [65]. Seventy-six percent of all individual genes found involved in
565 fusion events have previously been implicated in fusions [65]. *FGFR3* gene fusions were present in seven out of
566 90 samples with only one *FGFR3-TACC3* fusion. Four *PPARG* fusions were detected, of which two *PPARG-TSEN2*
567 fusions were confirmed at DNA level (Table S1.21). Other putative fusion events in cancer-related genes were
568 found in *CCSER1* (n = 9), *ERBB4* (n = 5), *RB1* (n = 4), *MDM2* (n = 4), *TERT* (n = 3) and *STAG2* (n = 3).

569 A stratification based on the proposed etiology of SNV COSMIC signatures using unsupervised consensus
570 clustering [30] revealed two major genomic subtypes (GenS; Fig. 2; Fig. S7). GenS1 (67%) contained almost
571 exclusively tumors with high APOBEC mutagenesis, which was reflected by a large contribution of APOBEC

572 signatures SBS2 and SBS13. In addition, we performed deconvolution of SNV patterns by non-negative matrix
573 factorization (NMF; Fig. S8), which confirmed APOBEC signatures as the main source of mutations in GenS1,
574 with high contribution of *de novo* mutational signatures SigA (0.99 cosine similarity with APOBEC signature
575 SBS2) and SigB (0.89 cosine similarity with APOBEC signature SBS13). GenS2 (24%) aggregated predominantly
576 tumors with low APOBEC mutagenesis (16 out of 28), and was characterized by signatures of unknown
577 etiology. *De novo* mutational signatures SigF (0.91 cosine similarity with SBS18 COSMIC signature) and SigG
578 (0.90 cosine similarity with SBS5 COSMIC signature) were dominant in GenS2. Analysis of the TCGA cohort
579 (WES data) showed that GenS1 and GenS2 were also the two major genomic subtypes in localized UC (Fig. S9a-
580 b). The other three smaller subtypes (9% of the present cohort) were related to the platinum treatment
581 signature (GenS3), the defective DNA mismatch repair (MMR) signature and microsatellite instability (MSI,
582 GenS4), and the reactive oxygen species signature (GenS5), which was characterized by *de novo* sigF in > 95%.

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

589 Other genomic differences between GenS1 and GenS2 (Fig. S3, S10 and S11) included higher SNVs/Mbp in
590 GenS1 and higher Indels/Mbp in GenS2, which was also observed in the TCGA cohort (Fig. S9c). All three
591 tumors with homologous recombination (HR) deficiency identified were of subtype GenS2. Clinical
592 characteristics such as sex, cancer subtype (bladder or upper tract UC), and pre-treatment status did not differ
593 between GenS1 and GenS2. Thus, despite that two very different etiologies lead to UC development, the two
594 mutagenic processes lead to similar profiles of somatically affected driver genes.

595

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

597 In tumors with high APOBEC mutagenesis, the mean ploidy and the number of genes affected by CNA were
598 higher than in tumors without APOBEC mutagenesis (Wilcoxon rank-sum test $p = 0.01$ and $p < 0.001$,
599 respectively; Fig. S12e-f). This phenomenon may indicate genomic instability in APOBEC-driven mUC tumors.

600 APOBEC enzyme expression analysis revealed neither significant differences between GenS1 and GenS2 (Fig.
601 S12a), nor between tumors with and without APOBEC mutagenesis (Fig. S12c). To further investigate the role
602 of APOBEC mutagenesis in mUC, we analyzed WGS data of eight tumors from patients who had undergone
603 serial biopsies, and reconstructed their evolutionary paths (Fig. 3a). Significantly mutated genes (dNdScv),
604 kataegis events and hotspot mutations displayed in the trunk represent clonal alterations that are fixed in the
605 tumor and are present in all cancer cells of both biopsies. Alternatively, mutations displayed in the branches
606 represent subclonal alterations that are present exclusively in one biopsy, or present in both biopsies but are
607 found only in a fraction of cancer cells (e.g. patient 1 in Fig. 3a). A lower cancer cell fraction in the branches of
608 than the trunk of the evolutionary trees (Fig. 3b) suggests that mutations from branches corresponding to the
609 second biopsy might be novel and not widely spread in the cancer cell population. The rate of novel APOBEC
610 mutations (number of APOBEC mutations divided by the number of days elapsed between serial biopsies) was
611 calculated using only mutations from branches corresponding to the second biopsy. This analysis showed that
612 tumors with high APOBEC mutagenesis accumulated more novel APOBEC mutations than other tumors (Fig.
613 3d, Wilcoxon rank-sum test $p = 0.036$). In line with this, we observed that the APOBEC mutational signature
614 was still present in the second biopsy of these patients, together suggesting that APOBEC mutagenesis is
615 ongoing in these samples (Fig. 3c). We further confirmed ongoing APOBEC mutagenesis by analyzing hotspot
616 mutations in mRNA. The frequency of hotspot mutations in *DDOST* (mRNA) was enriched in tumors with high
617 APOBEC mutagenesis and in GenS1 compared to GenS2, with up to 15% of mRNA molecules mutated in one
618 single sample (Fig. S13a).

619 Several studies have linked the expression of APOBEC3A/3B to APOBEC mutagenesis in UC [3,70]. When
620 comparing the estimated relative expression of APOBEC3A and 3B in the same tumor, we observed differential
621 expression of these enzymes (Fig. 3e). Some tumors had high levels of *APOBEC3A* expression while the
622 expression of *APOBEC3B* was low – or *vice versa*. The correlation between the expression of APOBEC3A and
623 APOBEC3B was poor (Fig S13b), which may explain the lack of differential expression of individual APOBEC

624 enzymes between tumors with different levels of APOBEC mutagenesis (Fig. S12c). To further investigate the
625 link between the occurrence of APOBEC mutations and APOBEC gene expression, we considered the
626 expression of both APOBEC enzymes and calculated an APOBEC score (sum of the median centered normalized
627 expression of *APOBEC3A* and *3B*). It appeared that tumors with high APOBEC mutagenesis had a higher
628 APOBEC score than other tumors (Wilcoxon rank-sum test $p = 0.012$; Fig. S12d). The APOBEC score was also
629 higher in GenS1 compared to GenS2 (Wilcoxon rank-sum test $p < 0.001$; Fig. S12b). This analysis confirmed a
630 link between APOBEC gene expression and APOBEC mutations in mUC. To further validate this result, the fold
631 enrichment of C>T and C>G mutations in the tetra YTCA (related to APOBEC3A) and RTCA (related to
632 APOBEC3B) context was calculated for the entire cohort (Fig. S13c). We found that both APOBEC3A and
633 APOBEC3B contribute to APOBEC associated mutations (fold enrichment is above 1.0), but APOBEC3A appears
634 to be the main contributor in mUC, as was reported previously [21,71].

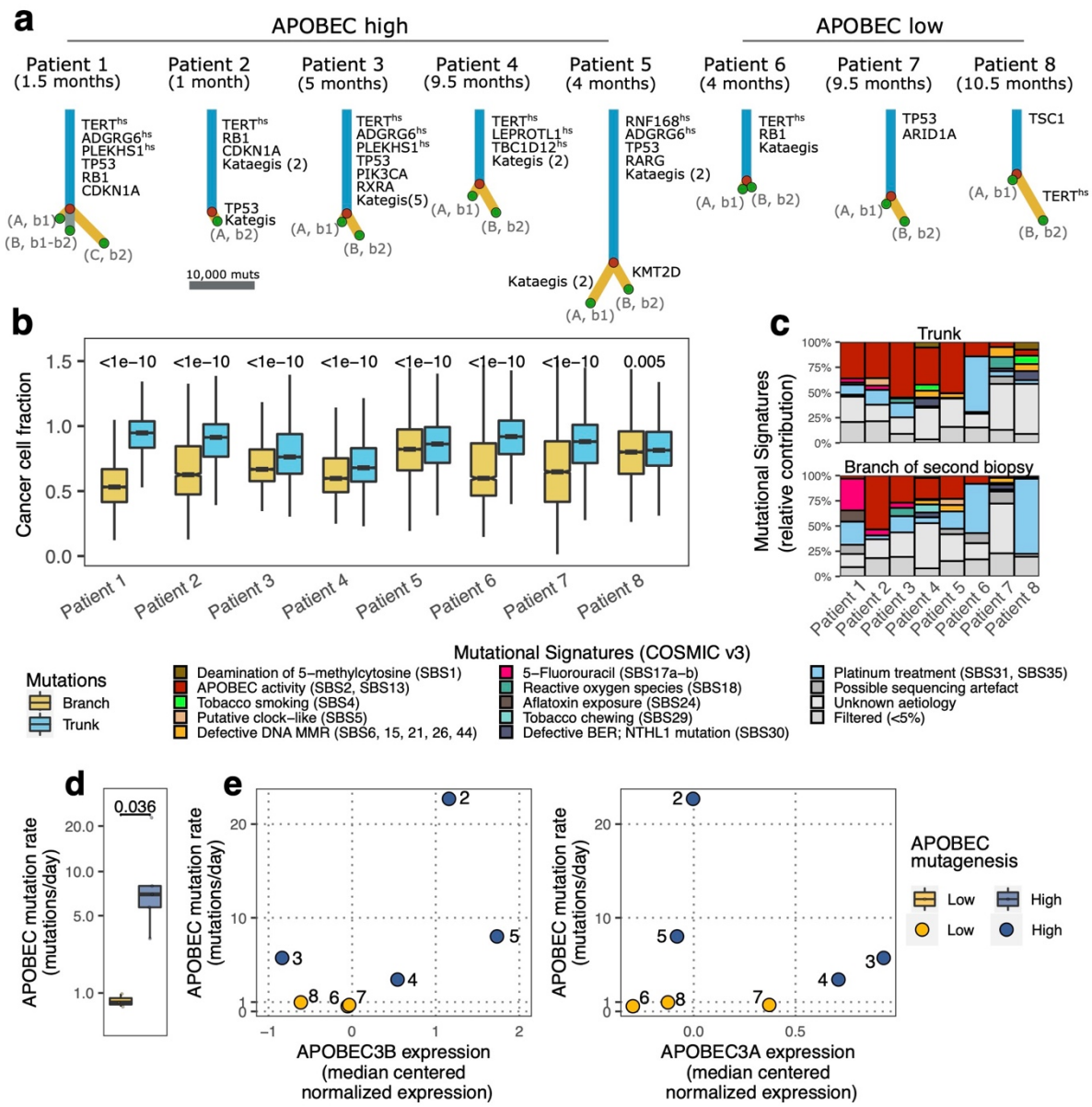


Figure 3 – Cancer evolution of eight tumors from metastatic urothelial carcinoma patients with serial biopsies of metastatic lesions

a) Evolutionary trees from eight tumors (five with high APOBEC and three with low APOBEC mutagenesis) from which two biopsies were obtained (interval depicted in months) were reconstructed from single nucleotide variants. Significantly mutated genes (dNdScv), kataegis (number of events in parenthesis) 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.

644 The cancer cell fraction of each single nucleotide variant was calculated and clustered using DPClust
645 for paired-biopsies. The evolutionary tree was reconstructed using the *sum rule* [66].
646 b) Boxplots comparing the cancer cell fraction of somatic mutations from the trunk and branches.
647 Wilcoxon rank-sum test was applied and p-values were Benjamini-Hochberg corrected.
648 c) COSMIC v3 mutational signatures calculated from the trunk and from the branch exclusive to the
649 second biopsy.
650 d) The APOBEC mutation rate from novel (recent) mutations in the second biopsy was compared
651 between low and high APOBEC mutagenesis tumors. Wilcoxon rank-sum test was applied.
652 e) APOBEC mutation rate is displayed as a function of *APOBEC3A* and *APOBEC3B* median centered
653 normalized expression. APOBEC expression was estimated as the mean expression of both biopsies
654 per tumor. Numbers indicate patient number. RNA-seq was not available for patient 1.

655

656 **APOBEC associated mutations are randomly distributed across the genome in mUC**

657 The substrate of APOBEC enzymes is single-stranded DNA (ssDNA) [20], this has led to the hypothesis that
658 APOBEC enzymes are mainly active during replication or in open chromatin and transcriptionally active
659 genomic regions [72,73]. As our cohort contained predominantly tumors with APOBEC mutagenesis, and WGS
660 data of these tumors was available, we had the unique opportunity to explore the mutational consequences of
661 APOBEC mutagenesis across the genome.

662 The total number of SNVs/Mbp varied across the genome, and non-APOBEC mutations followed the same
663 pattern (Fig. 4a). The frequency of non-APOBEC mutations decreased as the predicted DNA accessibility and
664 overall gene expression level increased (Fig. 4b). In contrast, the frequency of APOBEC mutations was constant
665 across the genome, demonstrating that APOBEC mutagenesis was likely independent of genomic regions (Fig.
666 4a-c). The ratio of APOBEC mutations between high and low transcriptional regions also suggests an equal
667 distribution of APOBEC associated mutations across the genome (Fig S14). This ratio was close to 0.5 (0.45-
668 0.55) in tumors with high APOBEC mutagenesis. The equal distribution of APOBEC mutations across genomic
669 regions supported the hypothesis that these mutations had been generated during replication, when APOBEC
670 enzymes have equal access to ssDNA across the genome [73].

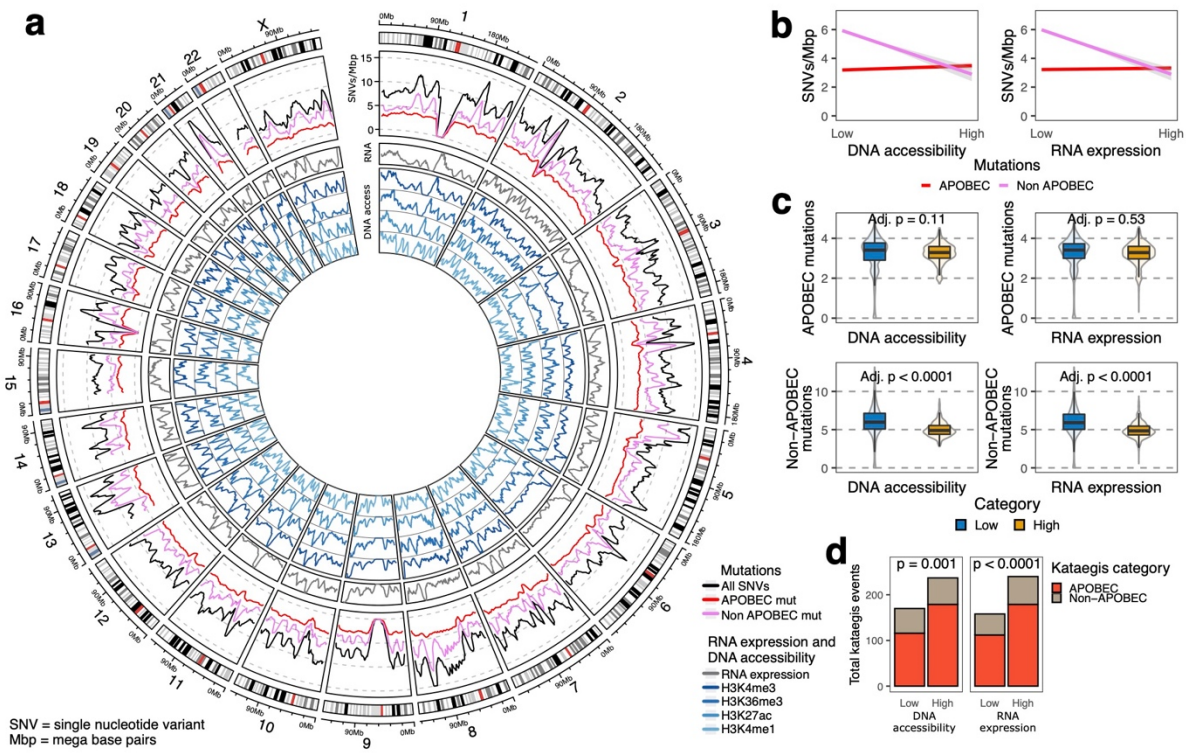


Figure 4 – Differences in the load of APOBEC associated mutations between high and low DNA accessibility regions in metastatic urothelial carcinoma genomes

a) WGS data (n = 116) was analyzed to estimate the mean number of single nucleotide variants in windows of one mega base pair 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; Average RNA counts (expression) from 90 tumors with RNA-seq data; DNA accessibility estimation from different ChIPseq experiments in normal urothelial samples derived from the ENCODE [74]. Peaks represent highly accessible DNA.

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

c) The number of APOBEC and non-APOBEC associated mutations were compared between high and low RNA expression and DNA accessibility regions. The distributions are shown as boxplots and as violin plots. T-test was applied and p-values were Benjamini-Hochberg corrected for multiple testing. Here,

685 in b) and in c), results from H3K4me1 ChIPseq were used. Using other ChIPseq experiments showed
686 similar results.
687 d) Frequency of kataegis events (n = 116) in high and low DNA accessibility or in high and low RNA
688 expression regions. P-values of binomial test are shown for each comparison.
689

690 Localized hypermutation events (kataegis) were present in 70% of the samples (Fig. 2a); which was more
691 frequent when APOBEC mutagenesis was high (Fig. S14). This higher frequency confirmed a link between
692 kataegis and APOBEC activity [75], and we therefore expected to find kataegis events scattered across the
693 genome. However, our data suggested that kataegis was more likely to happen in regions with high DNA
694 accessibility and high transcriptional activity (Fig. 4d). Thus, while general APOBEC mutagenesis seemed to
695 occur primarily during replication, kataegis-like APOBEC events seemed to occur more frequently at
696 transcribed loci.

697 In summary, APOBEC mutagenesis was an ongoing process in mUC that equally affected the whole genome,
698 and seemed to be triggered by both APOBEC3A and APOBEC3B. Tumors with APOBEC mutagenesis were
699 genomically less stable and displayed more kataegis events.

700

701 **Transcriptomic subtypes of mUC**

702 The consensus classifier of primary MIBC stratifies organ-confined UC of the bladder into six transcriptomic
703 subtypes [5]. Unlike primary bladder tumor samples, metastatic biopsies are derived from different organs
704 with some degree of normal non-urothelial cell contamination. Using the consensus classifier would lead to
705 misclassification of samples when no correction for organ-specific transcripts is applied. It would also limit the
706 detection of new phenotypic subtypes if they existed in mUC, which is crucial in this study as the
707 transcriptomic phenotypes of mUC are unknown. Therefore, we performed *de novo* subtyping for mUC
708 samples (see Methods section for details).

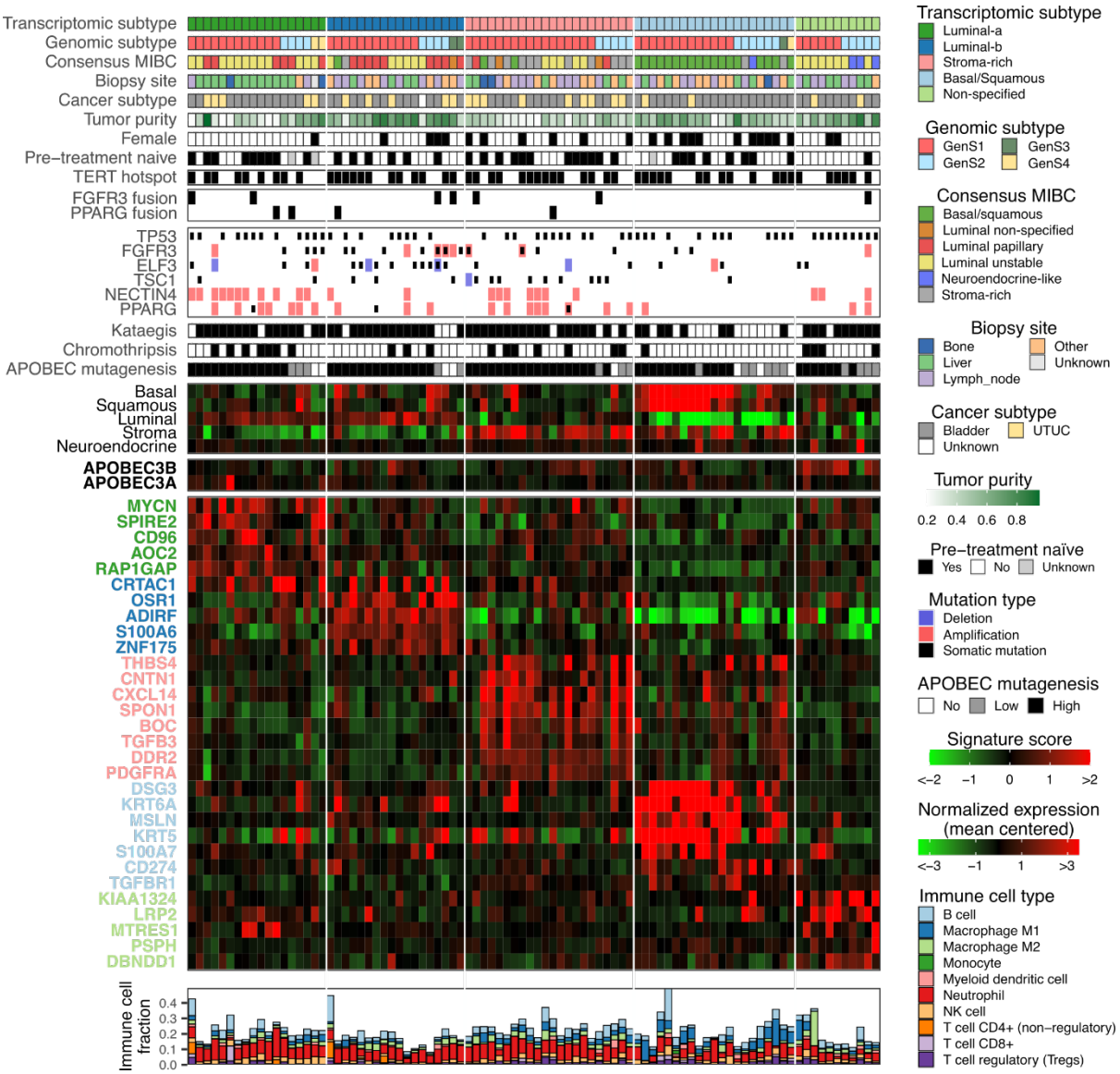
709 High-quality RNA-seq data was available for 90 (97 samples) out of 116 patients (Fig. S15). To reduce the bias
710 introduced by biopsy location, we filtered the organ-specific transcripts prior to hierarchical consensus
711 clustering (Fig. S16). Five transcriptomic subtypes were identified (Fig. 5). We did not observe a

712 neuroendocrine subtype in our cohort, and noticed that the neuroendocrine signature score was equally low

713 in all mUC subtypes (Fig. S17a). Several phenotypic markers were used to establish the phenotype of each

714 subtype (Fig. S17a), and are described below.

715



716

717 **Figure 5 – Genomic and transcriptomic characteristics of patients with metastatic urothelial carcinoma**

718 **stratified by mRNA subtypes**

719 Transcriptomic profiles of 90 metastatic urothelial carcinoma samples were clustered using

720 ConsensusClusterPlus [30]. Five transcriptomic subtypes were identified: luminal-a, luminal-b, stroma,

721 basal/squamous and non-specified phenotype. From top to bottom: Transcriptomic subtypes;

722 Genomic subtypes (GenS1-4); Transcriptional subtypes based on the consensus classifier of primary
723 MIBC [5]; Metastatic site from which a biopsy was obtained; Site of origin of primary tumor (UTUC =
724 upper tract urinary cancer); Estimated tumor cell percentage; Female patients; Pre-treatment naïve
725 patients; Tumors with hotspot mutations in *TERT* promoter; Tumors with gene fusions detected by
726 RNA-seq; Tumors with alterations in selected genes; Tumors with one or more kataegis events;
727 Tumors with one or more chromothripsis events; APOBEC enrichment analysis showing tumors with
728 no-, low- and high-APOBEC mutagenesis; *APOBEC3B* and *APOBEC3A* expression; Signature score
729 (mean expression of genes related to each phenotype) of basal, squamous, luminal, stroma and
730 neuroendocrine markers; Top overexpressed genes in each mRNA subtype; Immune cell fractions
731 estimated with immunedeconv [63], using the quanTIseq method [64].
732

733 The luminal subtypes represented 51% of the tumors in the TCGA cohort *versus* 40% in the present cohort (p =
734 0.061, Fig. S9d). In contrast to the consensus classifier of primary MIBC, we identified two and not three
735 luminal subtypes. The luminal mUC tumors were mostly identified as luminal-papillary and luminal-unstable
736 according to the consensus classifier of primary MIBC. All luminal tumors together exhibited similarities with
737 the consensus-based luminal subtypes [5] with respect to high luminal signature scores, high expression of
738 *PPARG* and *GATA3* (Fig. S16d), and frequent alterations in *FGFR3* and *KDM6A* (Fig. S17). In contrast, the
739 individual luminal mUC subtypes lacked a high TMB; high APOBEC mutation load, as described for the luminal
740 unstable subtype; high stromal signature score, as described for the luminal non-specified subtype; and
741 frequent CDKN2A alterations, as described for the luminal papillary and unstable subtypes. Looking into the
742 characteristics of the two luminal mUC subtypes, we noted that the luminal-a subtype had high expression of
743 *MYCN*, one of the MYC oncogene family members that regulates different species of RNA [76], and high
744 expression of *CD96* (Fig. 5). This subtype had low tumor purity, a high fraction of NK cells, a low clonal fraction
745 (interpreted as high heterogeneity), and relatively high expression of *FGFR3* and *NECTIN4* (Fig. S17). *NECTIN4*
746 was amplified in 61% of these tumors. The luminal-b subtype had high tumor purity, a low number of SVs, a
747 low fraction of NK cells, high expression of *FGFR3* and *S100A6* (Fig. S17), and a higher proportion of *ELF3* (56%)
748 and *FGFR3* (50%) DNA alterations compared to all other subtypes (Fig. S18; Fisher's exact test p = 0.0023 and p
749 = 0.0053, respectively).

750 The stroma-rich subtype, had the highest level of stroma signature score. Genes known to be associated with
751 stromal content and cancer-associated fibroblasts (*THBS4*, *CNTN1*, *CXCL14* and *BOC*) [77–80] were
752 differentially expressed (Fig. 5). This subtype was highly concordant with the stratification of the consensus
753 classifier of primary MIBC: 79% of tumors identified as stroma-rich by the consensus classifier of MIBC were in
754 the stroma-rich subtype of mUC. However, the stroma-rich subtype was more prevalent in the present mUC
755 cohort than in the TCGA cohort (24% *vs* 9%, Fisher's exact test $p < 0.001$, Fig. S9d). Tumors of the stroma-rich
756 subtype showed high expression of *TGFB3*, a ligand of the TGF-β pathway (Fig. 5), low tumor purity, a high
757 signature score for epithelial to mesenchymal transition (Fig. S17), high expression of various collagens (Table
758 S1.20), and a higher rate of *TSC1* DNA alterations (45% of the tumors, Fig. 5) compared to the rest of the
759 cohort (Fisher's exact test $p < 0.001$).

760 The basal/squamous subtype was also highly concordant to the similarly named cluster of the consensus
761 classifier for MIBC; 86% of tumors identified as basal/squamous by the consensus MIBC were in this group.
762 Yet, the prevalence of this subtype was lower in the present cohort than in the TCGA cohort (23% *vs* 37%,
763 Fisher's exact test $p = 0.013$; Fig. S9d). This subtype was characterized by high expression of basal/squamous
764 markers (*DSG3*, *KRT5, KRT6A* and *S100A7*), highest levels of basal and squamous signature scores, and
765 enrichment in females (52%, Fisher's exact test $p = 0.0043$). *TGFBR1,* a receptor of the TGF-β pathway; *CD274*,
766 the gene that encodes PD-L1; and *MSLN*, a tumor-associated antigen, were highly expressed in this subtype
767 (Fig. 5). *NECTIN4* amplifications were not found, and *NECTIN4* expression level was low (Fig. S17). In line with a
768 previous study [81], the expression of adipogenesis regulatory factor (*ADIRF*) was low (Fig. 5). The immune cell
769 compartment consisted of a large fraction of M1 macrophages and a low fraction of neutrophils (Fig. S17b).
770 This subtype was less affected by kataegis and chromothripsis events than the other subtypes (Fig. 5, Fisher's
771 exact test $p = 0.0006$ and $p = 0.019$, respectively).

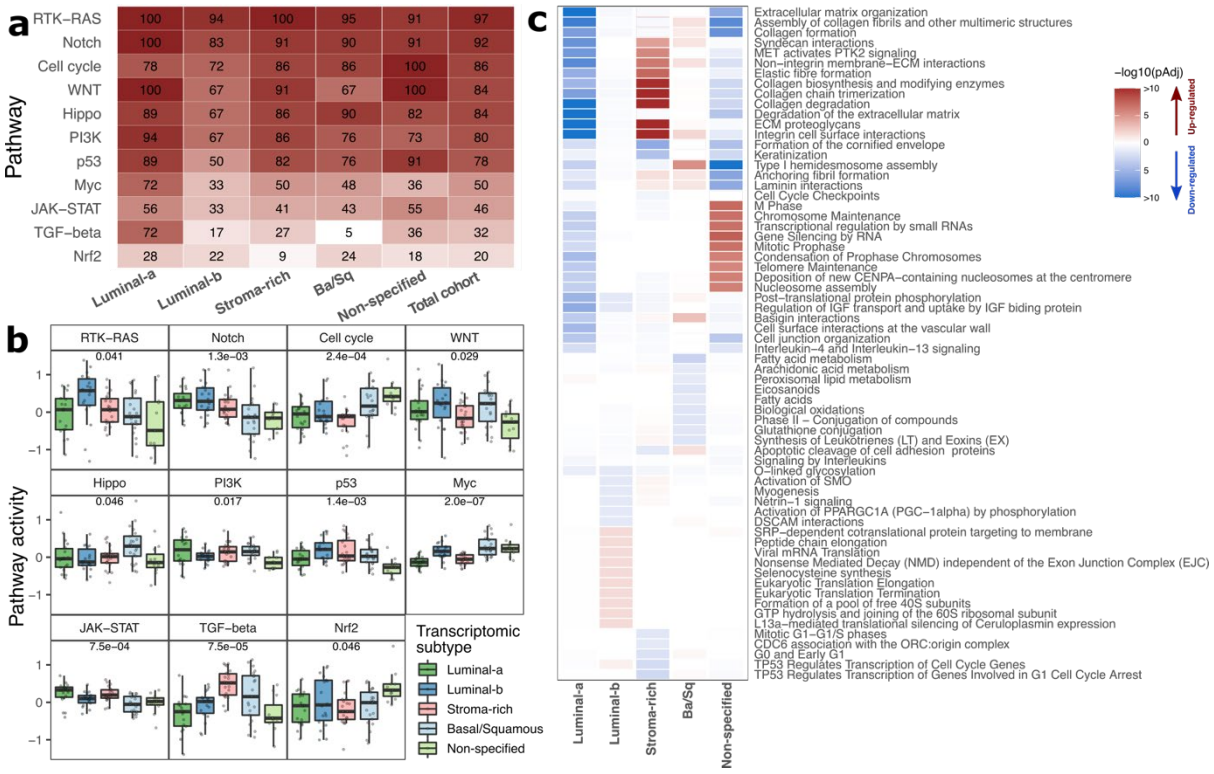
772 The majority of samples in the non-specified subtype was identified as luminal unstable according to the
773 consensus classifier of primary MIBC. However, key markers of luminal phenotypes such as the luminal
774 signature score, and *PPARG* and *GATA3* expression were relatively low, and genomic instability (high TMB, high
775 APOBEC mutagenesis) was not observed in this subtype (Fig. 5 and S17). This subtype had overexpression of
776 *KIAA1324;* a diagnostic biomarker in different cancer subtypes [82], and of *LRP2* (Fig. 5),. Furthermore, it had a
777 high score of claudin markers, a low fraction of neutrophils, high numbers of Indels, high numbers of SVs, high

levels of *APOBEC3B* expression (Fig. S16), and this subtype was enriched for patients who were previously treated with chemotherapy (Fisher's exact test p = 0.023, Fig. 5).

In summary, transcriptomic profiling revealed that mUC can be stratified into five transcriptomic subtypes, of which the stroma-rich and basal/squamous subtypes are highly concordant to primary MIBC subtypes. Both luminal subtypes showed some concordance with the luminal MIBC subtypes, however the individual luminal subtypes in mUC and MIBC differed. The phenotype of the non-specified mUC subtype did not match any of the phenotypes of the consensus classifier established for MIBC. A complete overview of driver genes, fusion genes and hotspot mutations per transcriptomic subtype is presented in Fig. S18.

Altered canonical signaling pathways in different transcriptomic subtypes

Several canonical pathways involved in cell growth, proliferation and survival [34] were altered at the DNA level (Fig. 6a). Of all subtypes, the luminal-a subtype showed most alterations in the Myc (72%) and TGF-β (72%) pathways. In contrast, only 5% of basal/squamous tumors had TGF-β pathway 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 Fig. S19.



794 **Figure 6 – Pathway alterations at genomic and transcriptomic level across mRNA-based subtypes of**
795 **metastatic urothelial carcinoma**

796 a) The percentage of samples with DNA alterations in 11 canonical pathways is shown for each
797 transcriptomic subtype (90 tumors in total) and for the entire cohort (n = 116). A patient was
798 considered to have an altered pathway when at least one of the pathway-genes was altered either by
799 non-synonymous mutations, structural variants or by deep copy number changes.
800 b) Pathway activity was estimated as the mean expression of downstream genes targeted by each
801 pathway. Only genes that were transcriptionally activated by these pathways were considered.
802 Kruskal-Wallis test p-values were Benjamini-Hochberg corrected.
803 c) Pathways up- (red) or down-regulated (blue) were estimated with reactomePA [62] from RNA-seq
804 data. Only the top ten up- and top ten down-regulated pathways per subtype are shown.
805

806 The luminal-b subtype was characterized by fewer alterations in Notch, Cell cycle, Hippo, PI3K, p53, Myc and
807 JAK-STAT pathways. Alterations in the p53 pathway were common in the other subtypes, and most of them
808 were the result of somatic mutations in *TP53* or amplification of *MDM2* (mutually exclusive p = 0.024).
809 Amplification of *MDM2* has been previously reported in a pan-cancer study [34] as an alternative to *TP53*
810 alterations to inactivate the p53 pathway through direct inhibition of p53 protein [83].

811 To assess pathway activity across different transcriptomic subtypes, we calculated the mean expression of
812 genes targeted by each pathway as *a proxy* of activity (Fig. 6b). Myc and TGF-β pathway activities were low in
813 the luminal-a subtype, corresponding with high frequencies of pathway alterations at the genomic level (Fig.
814 6a). The luminal-b subtype showed the highest RTK-RAS and high WNT pathway activity. The stroma-rich
815 subtype had high TGF-β pathway activity. The basal/squamous subtype had high activity of the Hippo, Myc and
816 TGF-β pathways. The non-specified subtype had very low p53 pathway activity and very active cell cycle
817 pathway signaling, two pathways that are usually co-altered [34].

818 In addition to the 11 oncogenic pathways described above, any pathway up- or down-regulated was analyzed
819 by enriched pathway analysis with ReactomePA (Fig. 6c). Up-regulation of pathways involved in collagen
820 metabolism and extracellular matrix in the stroma-rich subtype corresponded with the stromal phenotype of
821 this subtype [84]. In the non-specified subtype, pathways related to cell cycle and chromosome integrity were

822 up-regulated. Considering as well the high cell cycle pathway activity (Fig. 6b), and high frequency of mutations
823 in the cell cycle pathway (Fig. 6a), this up-regulation may suggest that the non-specified subtype is highly
824 proliferative.

825 In summary, signaling pathway analysis showed the extent of heterogeneity between the transcriptomic
826 subtypes, reflecting phenotypic characteristics of each group. The most striking difference was observed for
827 the TGF-β pathway, in which genomic alterations greatly affected the luminal-a subtype and pathway activity
828 was heavily reduced.

829 ***Discussion***

830 To the best of our knowledge, we defined, for the first time, molecular subtypes of mUC based on whole
831 genome and transcriptome characteristics of metastatic biopsies of 116 mUC patients. In line with findings
832 reported for primary UC, we identified a central role for APOBEC mutagenesis in mUC. Furthermore, we
833 showed that mUC is a heterogeneous disease with various genomic and transcriptomic subtypes revealing the
834 main mutational processes and phenotypes of this cancer.

835 The genomic landscape of mUC showed important similarities to that of primary UC. We validated our mUC
836 findings in the TCGA cohort of primary MIBC, showing that aggregating mutational signatures by etiology is a
837 robust approach to identify genomic subtypes. A recent study analyzed archived paraffin-embedded primary
838 or metastatic tumor samples from UC patients who received palliative chemotherapy. By WES analysis, two
839 major genomic subtypes were identified [85]. The GenS2 subtype, enriched with SigG that correlates with
840 COSMIC SBS5, in the present study largely overlapped with the SBS5 subtype reported by Taber *et al.*, 2020.
841 Furthermore, an APOBEC high signature was identified that was similar to GenS1 in our study.

842 We identified significantly mutated genes similar to those reported for primary UC. Frequent hotspot
843 mutations in the non-coding region of *TERT*, *ADGRG6*, *PLEKHS1*, *LEPROTL1* and *TBC1D12* occurred similarly in
844 NMIBC and MIBC [86,87]. Evidence suggests that clones with known driver genes emerge early during bladder
845 cancer development and colonize distant areas of the bladder, which may explain the genomic similarity
846 between mUC and primary UC [88].

847 WGS analysis revealed frequent SVs affecting *AHR* (aryl hydrocarbon receptor) and *CCSER1* (coiled-coil serine
848 rich protein). SVs in *AHR* have not been described in cancer, but other molecular alterations in this gene have
849 been associated with bladder cancer progression [89–91]. *CCSER1* is located in a common fragile site region;
850 thus, it is exposed to chromosomal rearrangements [92]. Altered transcripts created through the deletion of
851 specific exons in *CCSER1* have been associated with oncogenesis [92,93]; it is unclear, however, if SVs may
852 have similar oncogenic effects in UC.

853 Previous studies that performed RNA-based subtyping showed that NMIBC is a homogeneous disease primarily
854 of luminal origin (> 90%) and that MIBC is highly heterogeneous with multiple subtypes [5,94]. Here, we
855 performed *de novo* transcriptomic subtyping of mUC, as the consensus classifier of primary MIBC is not
856 suitable for mUC, and it does not correct for organ-specific transcriptomic contamination. Although this

857 method may also remove transcription related to urothelial adaptation, or to a specific metastatic site, it was
858 mandatory in order to prevent subtyping of tumors to be predominantly based on biopsy site rather than
859 biological differences. We showed that mUC is a heterogeneous disease, with similarities to subtypes
860 described for primary MIBC, specifically regarding the stroma-rich and basal/squamous subtypes. The luminal
861 subtypes overall showed some concordance, although there was not a one to one match with the primary
862 MIBC luminal subtypes. The phenotypic similarity of primary MIBC and mUC suggests that despite ongoing
863 mutagenesis, UC cell behavior does not change significantly during the metastatic process – all subtypes have
864 metastatic potential. However, as matched data from the primary tumor were lacking in this study, we were
865 unable to draw conclusions on specific factors which may drive some clones towards metastasis whilst other
866 remain tissue confined. Furthermore, some patients with primary non-metastatic MIBC, as assessed by cross-
867 sectional imaging, actually have systemic rather than localized disease. In a previous study, lymph node
868 metastases were present in the cystectomy resection specimen of 25% of clinically node-negative patients. In
869 addition, patients with locally advanced bladder tumors (pathological stage T3) have a poor 5-year overall
870 survival rate of only 35% due to rapid onset of metastatic disease, despite radical surgery [95,96].

871 In the present cohort, we did not identify a NE-like subtype at the transcriptional level. The prevalence of this
872 subtype in UC is, however low; in the TCGA cohort only 2% of tumors were of the NE-like subtype. Central
873 pathology revision of the metastatic biopsies identified only three NE-like tumors in our cohort (Table S3-4).
874 The non-specified subtype we identified did not express any of the markers used to identify the consensus
875 subtypes of primary MIBC (luminal, basal, squamous, stroma or NE), suggesting rewiring of its transcriptomic
876 profile for adaptation. Studies in various cancers have shown that therapeutic pressure may trigger a
877 phenotype-switching event [97], which could have happened in the non-specified phenotype as it was
878 enriched for patients who had received systemic therapy prior to biopsy. Studies with larger numbers of paired
879 biopsies, obtained before and after treatment, and obtained from the primary and metastatic site would be
880 needed to explore this phenomenon in mUC.

881 APOBEC mutagenesis was widespread in mUC; the reconstruction of evolutionary paths from sequential
882 biopsies of eight patients indicated that it was an ongoing process, which was confirmed in the entire cohort
883 by RNA-editing of *DDOST* attributed to APOBEC3A. This suggests that mUC is in continuous adaptation by
884 generating novel mutations. A previous study indeed reported accumulating mutations in six patients whose

885 primary tumor and metachronous metastases were analyzed by WES [85]. In our study, the accumulation of
886 new mutations in the sequential biopsy specimen of one of eight patients led to the identification of new
887 therapeutic targets (Fig. S20).

888 In a previous study [7], the genomic landscape of 85 (72 re-analyzed here) mUC patients was compared with
889 that of other metastatic tumor types. This pan-cancer study concluded that mUC was characterized by high
890 tumor mutational load, with no difference between mUC and primary UC, high CNAs, the highest number of
891 driver genes among all cancer types analyzed, and actionable targets in 75% of the patients. In our study, we
892 identified a potential targetable alteration in the genome of 98% of the patients (Fig. 7a). In line with Priestley
893 *et al.*, 2019, we found that 41% of patients could benefit from on-label therapies, and 63% from therapies
894 approved by the US Food and Drug Administration for other tumor types. Additionally, we identified targets
895 for therapies under investigation in clinical trials including basket trials in 109 of 116 patients. We identified
896 four patients with MSI-high tumors that are potentially sensitive to immune checkpoint inhibitors [98]. HR
897 deficiency, observed in three patients, is a potential target for treatment with poly-ADP ribose polymerase
898 inhibitors and/or double-stranded DNA break-inducing chemotherapy. At the RNA level, targetable *FGFR3* and
899 *NTRK2* gene fusions were identified in eight patients.

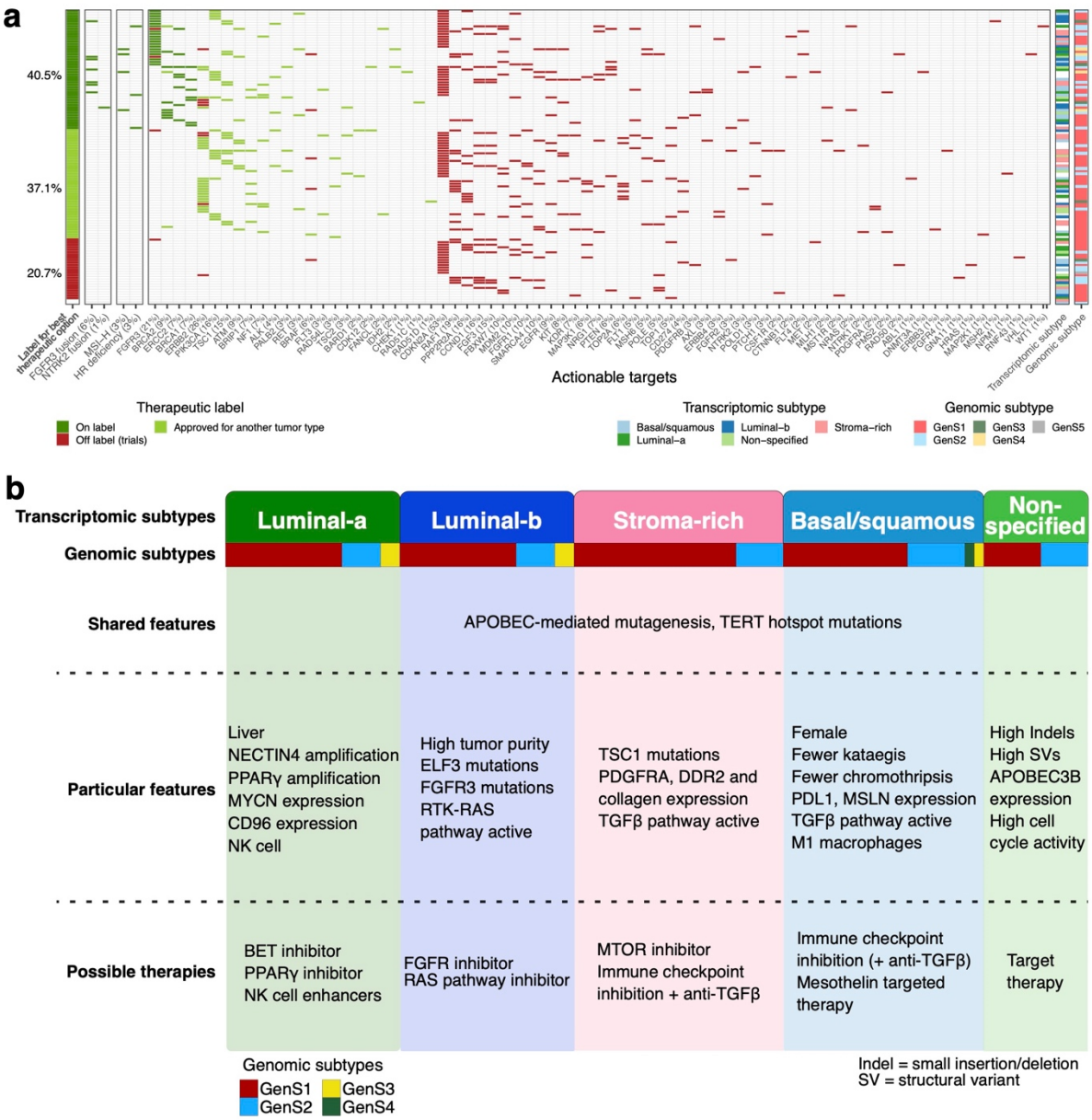


Figure 7 – Overview of actionable targets and possible treatments per transcriptomic subtype for metastatic urothelial carcinoma

a) Per patient overview of therapeutic targets based on gene fusions at RNA level (first column), tumors with microsatellite instability high (MSI-H) tumors, or homologous recombination (HR) deficiency (second column), and clinically-actionable genomic alterations for on- and off-label therapies for urothelial carcinoma (third column). On the left side, the therapeutic label for the best treatment

907 option per patient is shown. Bars on the right depict the genomic and the transcriptomic subtype per
908 patient.
909 b) Summary of molecular characteristics found in the present study, and potential therapeutic
910 implications for the treatment of metastatic urothelial carcinoma per transcriptomic subtype. From
911 top to bottom: transcriptomic mUC subtypes, genomic mUC subtypes, shared genomic features
912 among transcriptomic subtypes; unique characteristics per transcriptomic subtype; suggested
913 therapeutic strategies per transcriptomic subtype.
914

915 In a previous study, the antibody-drug conjugate enfortumab vedotin targeting NECTIN4 induced objective
916 clinical responses in 44% of mUC patients who experienced disease progression following chemotherapy and
917 anti-PD1/L1 therapy [99]. Currently, preselection for this treatment is not required. However, we found
918 significant variation in the expression of *NECTIN4*, suggesting that patients with tumors of the basal/squamous
919 subtype may be less likely to experience clinical benefit, as no *NECTIN4* amplifications were detected, and
920 *NECTIN4* expression levels were low. Thus, subtype-specific treatment with enfortumab vedotin might result in
921 better risk-benefit ratios. The 23 patients with *HER2* aberrations may be sensitive to HER2 targeting agents;
922 especially some of the newer antibody-drug conjugates with DNA damaging payloads could represent an
923 effective treatment [100,101].

924 Based on the identified transcriptomic subtypes we suggested potential therapeutic targets per subtype (Fig.
925 7b). The luminal-a subtype was characterized by *MYCN* and *PPARGC1B* overexpression. In pre-clinical studies,
926 treatment with a BET- or PPARγ-inhibitor downregulated the expression levels of both genes, and had an
927 antiproliferative effect on tumor cells [102,103]. The immune cell compartment of tumors of the luminal-a
928 subtype was found rich in NK cells, which could be explained by the large fraction of liver biopsies, which are
929 known to be enriched for NK cells [104]. Thus, other potential treatment strategies comprise of cytokine-
930 mediated stimulation of NK cells and TLR agonists [105].

931 The luminal-b subtype was enriched for *FGFR3* mutations and had high expression of *FGFR3*, suggesting that
932 this subtype may be susceptible to FGFR inhibitors. This subtype may also be sensitive to RAS pathway
933 inhibitors as the RTK-RAS pathway activity was high [106].

934 The stroma-rich subtype was characterized by *TSC1* alterations that confer sensitivity to MTOR inhibitors,
935 which have been approved for treatment of several tumor types [37,38]. Compared with the other subtypes,
936 the stroma-rich subtype displayed the highest TGF-β pathway activity and overexpression of different
937 collagens. Previous studies have shown that TGF-β can stimulate cancer-associated fibroblasts to produce
938 collagens [107,108]. Other studies found that TGF-β expression was associated with resistance to immune
939 checkpoint inhibition in bladder cancer [52,109]. Results from pre-clinical studies suggest that addition of a
940 TGF-β inhibitor may improve anti-PD1 efficacy [53].

941 The basal/squamous subtype has been found associated with high immune cell infiltration (significantly more
942 M1 macrophages) and overexpression of PD-L1, which suggests that patients with tumors of this subtype are
943 likely to benefit from immunotherapy [3]. Since TGF-β pathway activity was also high in this subtype,
944 combination therapy with a TGF-β inhibitor could be of added value. Furthermore, this subtype was
945 characterized by overexpression of mesothelin, a known tumor antigen that is being investigated as a target
946 for antibody-based, vaccine and CAR-T cell therapies in several tumor types [110].

947 A limitation of this study is the lack of matched primary tumor samples. Despite this, we showed striking
948 genomic and transcriptomic similarities between mUC and what has been reported for primary MIBC. Our
949 results, however, require validation in other independent mUC cohorts. Also, as our studied cohort was
950 heterogeneous regarding pre-treatment history and type of treatment initiated after biopsy collection, we
951 were unable to correlate the characteristics of the molecular subtypes to clinical endpoints such as overall
952 survival. Additional studies in which biopsies are collected from uniformly treated mUC patients would be
953 crucial to be able to properly correlate large scale genomic and transcriptomic data with clinical outcomes.

954 **Conclusions**

955 By performing WGS and RNA-seq analysis of metastatic sites of 116 mUC patients who participated in a clinical
956 trial, this study contributed to the knowledge on the molecular landscape of mUC, which has important
957 similarities to the molecular landscape of primary UC. The findings reported here serve as a reference for
958 subtype-oriented and patient-specific research on the etiology of mUC and for novel drug development – with
959 the ultimate aim to improve the management of mUC patients.

960

961 ***List of abbreviations***

962 APOBEC: Apolipoprotein B mRNA Editing Catalytic Polypeptide-like

963 CCF: Cancer cell fraction

964 CNA: Copy number alterations

965 CPCT: The Center for Personalized Cancer Treatment

966 FC: Fold change

967 H&E: Hematoxylin and eosin

968 HMF: Hartwig Medical Foundation

969 HR: Homologues recombination

970 Indel: insertion/deletion

971 Mbp: Megabase pair

972 MIBC: Muscle invasive bladder cancer

973 MMR: Mismatch repair

974 MNV: Multiple nucleotide variants

975 MSI: Microsatellite instability

976 mUC: Metastatic urothelial carcinoma

977 NE-like: Neuroendrocrine-like

978 NMF: Non-negative matrix factorization

979 NMIBC: Non-muscle invasive bladder cancer

980 RNA-seq: RNA sequencing

981 SNV: Single nucleotide variant

982 SV: Structural variant

983 TAF: Tumor allele frequency

984 TCGA: The Cancer Genome Atlas

985 TMB: Tumor mutational burden

986 UC: Urothelial carcinoma

987 UTUC: Upper tract urothelial carcinoma

988 VEP: Variant effect predictor

989 WES: whole-exome sequencing

990 WGS: Whole-genome sequencing

991 **Declarations**

992 ***Ethics approval and consent to participate***

993 Patients studied here were included in the CPCT-02 Biopsy Protocol (ClinicalTrial.gov no. NCT01855477) and
994 the DRUP Trial (ClinicalTrial.gov no. NCT02925234). The study protocols were approved by the medical ethics
995 review board of the University Medical Center Utrecht and the Netherlands Cancer Institute. Written informed
996 consent was obtained from all participants prior to inclusion in the trials; the studies comply with all relevant
997 ethical regulations.

998 ***Consent for publication***

999 All patients studied here provided consent to report individual (anonymised) patient data.

1000 ***Availability of data and materials***

1001 WGS data, RNA-seq data and corresponding clinical data have been requested from the HMF and were
1002 provided under data request number DR-031. All data are freely available for academic use from the HMF
1003 through standardized procedures. Request forms can be found at https://www.hartwigmedicalfoundation.nl
1004 [7].

1005 ChIPseq data experiments are freely available through The ENCODE Project Consortium [111] and the
1006 Roadmap Epigenomics Consortium [112] on the ENCODE portal (https://www.encodeproject.org) [74]. Files
1007 were downloaded with the following identifiers: ENCSR065IQH, ENCSR054BKO, ENCSR632OWD and
1008 ENCSR449TNC.

1009 TCGA data for the bladder cancer cohort was downloaded through the portal: https://www.cancer.gov/tcga.

1010 ***Competing interests***

1011 Michiel S. van der Heijden has received research support from Bristol-Myers Squibb, AstraZeneca and Roche,
1012 and consultancy fees from Bristol-Myers Squibb, Merck Sharp & Dohme, Roche, AstraZeneca, Seattle Genetics
1013 and Janssen (all paid to the Netherlands Cancer Institute). Niven Mehra has received research support from
1014 Astellas, Janssen, Pfizer, Roche and Sanofi Genzyme, and consultancy fees from Roche, MSD, BMS, Bayer,
1015 Astellas and Janssen (all paid to the Radboud University Medical Center). Sjoukje F. Oosting has received
1016 research support from Celldex and Novartis (both paid to the University Medical Center Groningen). Hans M.

Westgeest has received consultancy fees from Roche and Astellas (all paid to the Amphia hospital, Breda), Ronald de Wit has received research support from Sanofi and Bayer, and consultancy fees from Sanofi, Merck, Astellas, Bayer, Janssen, and Roche (all paid to the Erasmus MC Cancer Institute). Astrid A.M. van der Veldt has received consultancy fees from for BMS, MSD, Merck, Novartis, Roche, Sanofi, Pierre Fabre, Ipsen, Eisai, Pfizer (all paid to the Erasmus MC Cancer Institute). Martijn P. J. Lolkema has received research support from JnJ, Sanofi, Astella and MSD, and personal fees from Incyte, Amgen, JnJ, Bayer, Servier, Roche, INCa, Pfizer, Sanofi, Astellas, AstraZeneca, MSD, Novartis and Julius Clinical (all paid to the Erasmus MC Cancer Institute). Joost L. Boormans has received research support from Decipher Biosciences and Merck Sharp & Dohme, and consultancy fees from Merck Sharp & Dohme, Eight Medical, Ambu, APIM therapeutics and Janssen (all paid to the Erasmus MC Cancer Institute). J. Alberto Nakauma-González, Maud Rijnders, Job van Riet, Jens Voortman, Edwin Cuppen, Sandra van Wilpe, L. Lucia Rijstenberg, Ellen C. Zwarthoff, and Harmen J. G. van de Werken declare no competing interests.

Funding

This research was funded by the Barcode for Life foundation through M.P. J. L. H.J.G.W., J.A.N. and the Erasmus MC Cancer Computational Biology Center were financed through a grant from the Daniel den Hoed Foundation and the Dutch Uro-Oncology Study group (DUOS). Revision of pathological diagnosis was funded by a DUOS research grant. This study was also partially financed by a grant from Merck Sharpe & Dome, Kenilworth, N.J., U.S.A., through M.P.J.L.

Author's 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. All authors read and approved the final manuscript.

Acknowledgements

The Hartwig Medical Foundation and the Center of Personalized Cancer Treatment are acknowledged for making the clinical and genomic data available to the study. We thank all local principal investigators and the nurses of all contributing centers for their help with patient recruitment. We are particularly grateful to all participating patients and their families.

1050 ***References***

1051 1. Knowles MA, Hurst CD. Molecular biology of bladder cancer: New insights into pathogenesis and
1052 clinical diversity. Nat Rev Cancer. 2015;15(1):25–41.

1053 2. Lindskrog SV, Prip F, Lamy P, Taber A, Groeneveld CS, Birkenkamp-Demtröder K, et al. An integrated
1054 multi-omics analysis identifies prognostic molecular subtypes of non-muscle-invasive bladder cancer.
1055 Nat Commun. 2021;12(1):2301.

1056 3. Robertson AG, Kim J, Al-Ahmadie H, Bellmunt J, Guo G, Cherniack AD, et al. Comprehensive Molecular
1057 Characterization of Muscle-Invasive Bladder Cancer. Cell. 2017;171(3):540-556.e25.

1058 4. Giannopoulou A, Velentzas A, Konstantakou E, Avgeris M, Katarachia S, Papandreou N, et al. Revisiting
1059 Histone Deacetylases in Human Tumorigenesis: The Paradigm of Urothelial Bladder Cancer. Int J Mol
1060 Sci. 2019;20(6):1291.

1061 5. Kamoun A, Reyniès A de, Allory Y, Sjödahl G, Gordon Robertson A, Seiler R, et al. A Consensus
1062 Molecular Classification of Muscle-invasive Bladder Cancer. Eur Urol. 2019;77(4):420–33.

1063 6. Faltas BM, Prandi D, Tagawa ST, Molina AM, Nanus DM, Sternberg C, et al. Clonal evolution of
1064 chemotherapy-resistant urothelial carcinoma. Nat Genet. 2016;48(12):1490–9.

1065 7. Priestley P, Baber J, Lolkema MP, Steeghs N, de Bruijn E, Shale C, et al. Pan-cancer whole-genome
1066 analyses of metastatic solid tumours. Nature. 2019;575(7781):210–6.

1067 8. Casparie M, Tiebosch ATMG, Burger G, Blauwgeers H, Pol A van de, Krieken JHJM van, et al. Pathology
1068 Databanking and Biobanking in The Netherlands, a Central Role for PALGA, the Nationwide
1069 Histopathology and Cytopathology Data Network and Archive. Cell Oncol. 2007;29(1):19.

1070 9. Li H, Durbin R. Fast and accurate short read alignment with Burrows-Wheeler transform.
1071 Bioinformatics. 2009;25(14):1754–60.

1072 10. McKenna A, Hanna M, Banks E, Sivachenko A, Cibulskis K, Kernytsky A, et al. The genome analysis
1073 toolkit: A MapReduce framework for analyzing next-generation DNA sequencing data. Genome Res.
1074 2010;20(9):1297–303.

1075 11. Van der Auwera GA, Carneiro MO, Hartl C, Poplin R, del Angel G, Levy-Moonshine A, et al. From fastQ
1076 data to high-confidence variant calls: The genome analysis toolkit best practices pipeline. Curr Protoc
1077 Bioinforma. 2013;43(SUPL.43):11.10.1-11.10.33.

1078 12. Saunders CT, Wong WSW, Swamy S, Becq J, Murray LJ, Cheetham RK. Strelka: Accurate somatic small-
1079 variant calling from sequenced tumor-normal sample pairs. Bioinformatics. 2012;28(14):1811–7.

1080 13. McLaren W, Gil L, Hunt SE, Riat HS, Ritchie GRS, Thormann A, et al. The Ensembl Variant Effect
1081 Predictor. Genome Biol. 2016;17(1):122.

1082 14. Liu X, Wu C, Li C, Boerwinkle E. dbNSFP v3.0: A One-Stop Database of Functional Predictions and
1083 Annotations for Human Nonsynonymous and Splice-Site SNVs. Hum Mutat. 2016;37(3):235–41.

1084 15. Lek M, Karczewski KJ, Minikel E V., Samocha KE, Banks E, Fennell T, et al. Analysis of protein-coding
1085 genetic variation in 60,706 humans. Nature. 2016;536(7616):285–91.

1086 16. Cameron D, Baber J, Shale C, Papenfuss A, Valle-Inclan JE, Besselink N, et al. GRIDSS, PURPLE, LINX:
1087 Unscrambling the tumor genome via integrated analysis of structural variation and copy number.
1088 bioRxiv preprint. 2019; https://doi.org/10.1101/781013.

1089 17. Martincorena I, Raine KM, Gerstung M, Dawson KJ, Haase K, Van Loo P, et al. Universal Patterns of
1090 Selection in Cancer and Somatic Tissues. Cell. 2017;171(5):1029-1041.e21.

1091 18. van Dessel LF, van Riet J, Smits M, Zhu Y, Hamberg P, van der Heijden MS, et al. The genomic landscape
1092 of metastatic castration-resistant prostate cancers reveals multiple distinct genotypes with potential
1093 clinical impact. Nat Commun. 2019;10(1):1–13.

1094 19. Mermel CH, Schumacher SE, Hill B, Meyerson ML, Beroukhim R, Getz G. GISTIC2.0 facilitates sensitive
1095 and confident localization of the targets of focal somatic copy-number alteration in human cancers.
1096 Genome Biol. 2011;12(4):R41.

1097 20. Roberts SA, Lawrence MS, Klimczak LJ, Grimm SA, Fargo D, Stojanov P, et al. An APOBEC cytidine
1098 deaminase mutagenesis pattern is widespread in human cancers. Nat Genet. 2013;45(9):970–6.

1099 21. Chan K, Roberts SA, Klimczak LJ, Sterling JF, Saini N, Malc EP, et al. An APOBEC3A hypermutation

1100 signature is distinguishable from the signature of background mutagenesis by APOBEC3B in human
1101 cancers. Nat Genet. 2015;47(9):1067–72.

1102 22. Stephens PJ, Tarpey PS, Davies H, Van Loo P, Greenman C, Wedge DC, et al. The landscape of cancer
1103 genes and mutational processes in breast cancer. Nature. 2012;486(7403):400–4.

1104 23. Gundem G, Van Loo P, Kremeyer B, Alexandrov LB, Tubio JMC, Papaemmanuil E, et al. The evolutionary
1105 history of lethal metastatic prostate cancer. Nature. 2015;520(7547):353–7.

1106 24. Tate JG, Bamford S, Jubb HC, Sondka Z, Beare DM, Bindal N, et al. COSMIC: The Catalogue Of Somatic
1107 Mutations In Cancer. Nucleic Acids Res. 2019;47(D1):D941–7.

1108 25. Blokzijl F, Janssen R, van Boxtel R, Cuppen E. MutationalPatterns: Comprehensive genome-wide
1109 analysis of mutational processes. Genome Med. 2018;10(1):33.

1110 26. Alexandrov LB, Kim J, Haradhvala NJ, Huang MN, Tian Ng AW, Wu Y, et al. The repertoire of mutational
1111 signatures in human cancer. Nature. 2020;578(7793):94–101.

1112 27. Petljak M, Alexandrov LB, Brammeld JS, Price S, Wedge DC, Grossmann S, et al. Characterizing
1113 Mutational Signatures in Human Cancer Cell Lines Reveals Episodic APOBEC Mutagenesis. Cell.
1114 2019;176(6):1282-1294.e20.

1115 28. Angus L, Smid M, Wilting SM, van Riet J, Van Hoeck A, Nguyen L, et al. The genomic landscape of
1116 metastatic breast cancer highlights changes in mutation and signature frequencies. Nat Genet.
1117 2019;51(10):1450–8.

1118 29. Christensen S, Van der Roest B, Besselink N, Janssen R, Boymans S, Martens JWM, et al. 5-Fluorouracil
1119 treatment induces characteristic T>G mutations in human cancer. Nat Commun. 2019;10(1):1–11.

1120 30. Wilkerson MD, Hayes DN. ConsensusClusterPlus: A class discovery tool with confidence assessments
1121 and item tracking. Bioinformatics. 2010;26(12):1572–3.

1122 31. Gaujoux R, Seoighe C. A flexible R package for nonnegative matrix factorization. BMC Bioinformatics.
1123 2010;11(1):367.

1124 32. Cortés-Ciriano I, Lee JJK, Xi R, Jain D, Jung YL, Yang L, et al. Comprehensive analysis of chromothripsis in

1125 2,658 human cancers using whole-genome sequencing. Nat Genet. 2020;52(3):331–41.

1126 33. Nguyen L, Martens J, Hoeck A van, Cuppen E. Pan-cancer landscape of homologous recombination
1127 deficiency. Nat Commun. 2020;11:5584.

1128 34. Sanchez-Vega F, Mina M, Armenia J, Chatila WK, Luna A, La KC, et al. Oncogenic Signaling Pathways in
1129 The Cancer Genome Atlas. Cell. 2018;173(2):321-337.e10.

1130 35. Leonard WJ. Role of JAK kinases and stats in cytokine signal transduction. Int J Hematol.
1131 2001;73(3):271–7.

1132 36. Griffith M, Spies NC, Krysiak K, McMichael JF, Coffman AC, Danos AM, et al. CIViC is a community
1133 knowledgebase for expert crowdsourcing the clinical interpretation of variants in cancer. Nat Genet.
1134 2017;49(2):170–4.

1135 37. Chakravarty D, Gao J, Phillips S, Kundra R, Zhang H, Wang J, et al. OncoKB: A Precision Oncology
1136 Knowledge Base. JCO Precis Oncol. 2017;1:1–16.

1137 38. Tamborero D, Rubio-Perez C, Deu-Pons J, Schroeder MP, Vivancos A, Rovira A, et al. Cancer Genome
1138 Interpreter annotates the biological and clinical relevance of tumor alterations. Genome Med.
1139 2018;10(1):25.

1140 39. Nakato R, Shirahige K. Recent advances in ChIP-seq analysis: from quality management to whole-
1141 genome annotation. Brief Bioinform. 2017;18(2):279–90.

1142 40. Bolger AM, Lohse M, Usadel B. Trimmomatic: a flexible trimmer for Illumina sequence data.
1143 Bioinformatics. 2014;30(15):2114–20.

1144 41. Dobin A, Davis CA, Schlesinger F, Drenkow J, Zaleski C, Jha S, et al. STAR: Ultrafast universal RNA-seq
1145 aligner. Bioinformatics. 2013;29:15–21.

1146 42. Harrow J, Frankish A, Gonzalez JM, Tapanari E, Diekhans M, Kokocinski F, et al. GENCODE: The
1147 reference human genome annotation for the ENCODE project. Genome Res. 2012;22(9):1760–74.

1148 43. Tarasov A, Vilella AJ, Cuppen E, Nijman IJ, Prins P. Sambamba: Fast processing of NGS alignment
1149 formats. Bioinformatics. 2015;31(12):2032–4.

1150 44. Wang L, Nie J, Sicotte H, Li Y, Eckel-Passow JE, Dasari S, et al. Measure transcript integrity using RNA-
1151 seq data. BMC Bioinformatics. 2016;17:58.

1152 45. Wang L, Wang S, Li W. RSeQC: quality control of RNA-seq experiments. Bioinformatics.
1153 2012;28(16):2184–5.

1154 46. Liao Y, Smyth GK, Shi W. FeatureCounts: An efficient general purpose program for assigning sequence
1155 reads to genomic features. Bioinformatics. 2014;30(7):923–30.

1156 47. Li B, Dewey CN. RSEM: Accurate transcript quantification from RNA-seq data with or without a
1157 reference genome. BMC Bioinformatics. 2011;12:323.

1158 48. Ramírez F, Ryan DP, Grüning B, Bhardwaj V, Kilpert F, Richter AS, et al. deepTools2: a next generation
1159 web server for deep-sequencing data analysis. Nucleic Acids Res. 2016;44(W1):W160–5.

1160 49. Williams SB, Black PC, Dyrskjøt L, Seiler R, Schmitz-Dräger B, Nawroth R, et al. Re: Aurélie Kamoun,
1161 Aurélien de Reyniès, Yves Allory, et al. A Consensus Molecular Classification of Muscle-invasive Bladder
1162 Cancer. Eur Urol 2020;77:420–33: A Statement from the International Bladder Cancer Network. Vol.
1163 77, European Urology. Elsevier B.V.; 2020. p. e105–6.

1164 50. Love MI, Huber W, Anders S. Moderated estimation of fold change and dispersion for RNA-seq data
1165 with DESeq2. Genome Biol. 2014;15(12):550.

1166 51. Paulson JN, Chen CY, Lopes-Ramos CM, Kuijjer ML, Platig J, Sonawane AR, et al. Tissue-aware RNA-Seq
1167 processing and normalization for heterogeneous and sparse data. BMC Bioinformatics. 2017;18(1):1–
1168 10.

1169 52. Powles T, Kockx M, Rodriguez-Vida A, Duran I, Crabb SJ, Van Der Heijden MS, et al. Clinical efficacy and
1170 biomarker analysis of neoadjuvant atezolizumab in operable urothelial carcinoma in the ABACUS trial.
1171 Nat Med. 2019;25(11):1706–14.

1172 53. Mariathasan S, Turley SJ, Nickles D, Castiglioni A, Yuen K, Wang Y, et al. TGFβ attenuates tumour
1173 response to PD-L1 blockade by contributing to exclusion of T cells. Nature. 2018;554(7693):544–8.

1174 54. Borggrefe T, Oswald F. The Notch signaling pathway: Transcriptional regulation at Notch target genes.

1175 Cell Mol Life Sci. 2009;66(10):1631–46.

1176 55. van Ooijen H, Hornsveld M, Dam-de Veen C, Velter R, Dou M, Verhaegh W, et al. Assessment of
1177 Functional Phosphatidylinositol 3-Kinase Pathway Activity in Cancer Tissue Using Forkhead Box-O
1178 Target Gene Expression in a Knowledge-Based Computational Model. Am J Pathol. 2018;188(9):1956–
1179 72.

1180 56. Varelas X. The hippo pathway effectors TAZ and YAP in development, homeostasis and disease. Dev.
1181 2014;141(8):1614–26.

1182 57. Fischer M. Census and evaluation of p53 target genes. Oncogene. 2017;36(28):3943–56.

1183 58. Kitamura H, Motohashi H. NRF2 addiction in cancer cells. Cancer Sci. 2018;109(4):900–11.

1184 59. Hartl M. The quest for targets executing MYC-dependent cell transformation. Vol. 6, Frontiers in
1185 Oncology. Frontiers Research Foundation; 2016. p. 132.

1186 60. Wagle M-C, Kirouac D, Klijn C, Liu B, Mahajan S, Junttila M, et al. A transcriptional MAPK Pathway
1187 Activity Score (MPAS) is a clinically relevant biomarker in multiple cancer types. npj Precis Oncol.
1188 2018;2(1):1–12.

1189 61. Murray PJ. The JAK-STAT Signaling Pathway: Input and Output Integration. J Immunol.
1190 2007;178(5):2623–9.

1191 62. Yu G, He QY. ReactomePA: An R/Bioconductor package for reactome pathway analysis and
1192 visualization. Mol Biosyst. 2016;12(2):477–9.

1193 63. Sturm G, Finotello F, Petitprez F, Zhang JD, Baumbach J, Fridman WH, et al. Comprehensive evaluation
1194 of transcriptome-based cell-type quantification methods for immuno-oncology. Bioinformatics.
1195 2019;35(14):i436–45.

1196 64. Finotello F, Mayer C, Plattner C, Laschober G, Rieder D, Hackl H, et al. Molecular and pharmacological
1197 modulators of the tumor immune contexture revealed by deconvolution of RNA-seq data. Genome
1198 Med. 2019;11(1):34.

1199 65. Jang YE, Jang I, Kim S, Cho S, Kim D, Kim K, et al. ChimerDB 4.0: an updated and expanded database of

1200 fusion genes. Nucleic Acids Res. 2019;48(D1):D817–24.

1201 66. Dang HX, White BS, Foltz SM, Miller CA, Luo J, Fields RC, et al. ClonEvol: clonal ordering and
1202 visualization in cancer sequencing. Ann Oncol. 2017;28(12):3076–82.

1203 67. R Core Team. R Core Team (2017). R: A language and environment for statistical computing. R Found
1204 Stat Comput Vienna, Austria URL http//wwwR-project.org/. 2017;R Foundation for Statistical
1205 Computing.

1206 68. Allory Y, Beukers W, Sagrera A, Flández M, Marqués M, Márquez M, et al. Telomerase reverse
1207 transcriptase promoter mutations in bladder cancer: High frequency across stages, detection in urine,
1208 and lack of association with outcome. Eur Urol. 2014;65(2):360–6.

1209 69. Buisson R, Langenbucher A, Bowen D, Kwan EE, Benes CH, Zou L, et al. Passenger hotspot mutations in
1210 cancer driven by APOBEC3A and mesoscale genomic features. Science. 2019;364(6447):eaaw2872.

1211 70. Glaser AP, Fantini D, Wang Y, Yu Y, Rimar KJ, Podojil JR, et al. APOBEC-mediated mutagenesis in
1212 urothelial carcinoma is associated with improved survival, mutations in DNA damage response genes,
1213 and immune response. Oncotarget. 2018;9(4):4537–48.

1214 71. Cortez LM, Brown AL, Dennis MA, Collins CD, Brown AJ, Mitchell D, et al. APOBEC3A is a prominent
1215 cytidine deaminase in breast cancer. PLoS Genet. 2019;15(12):e1008545.

1216 72. Kazanov MD, Roberts SA, Polak P, Stamatoyannopoulos J, Klimczak LJ, Gordenin DA, et al. APOBEC-
1217 Induced Cancer Mutations Are Uniquely Enriched in Early-Replicating, Gene-Dense, and Active
1218 Chromatin Regions. Cell Rep. 2015;13(6):1103–9.

1219 73. Hoopes JII, Cortez LMM, Mertz TMM, Malc EPP, Mieczkowski PAA, Roberts SAA. APOBEC3A and
1220 APOBEC3B Preferentially Deaminate the Lagging Strand Template during DNA Replication. Cell Rep.
1221 2016;14(6):1273–82.

1222 74. Davis CA, Hitz BC, Sloan CA, Chan ET, Davidson JM, Gabdank I, et al. The Encyclopedia of DNA elements
1223 (ENCODE): data portal update. Nucleic Acids Res. 2018;46(D1):D794–801.

1224 75. Nik-Zainal S, Alexandrov LB, Wedge DC, Van Loo P, Greenman CD, Raine K, et al. Mutational processes

1225 molding the genomes of 21 breast cancers. Cell. 2012;149(5):979–93.

1226 76. Baluapuri A, Wolf E, Eilers M. Target gene-independent functions of MYC oncoproteins. Nat Rev Mol
1227 Cell Biol. 2020;21(5):255–67.

1228 77. Kuroda K, Yashiro M, Sera T, Yamamoto Y, Kushitani Y, Sugimoto A, et al. The clinicopathological
1229 significance of Thrombospondin-4 expression in the tumor microenvironment of gastric cancer. Katoh
1230 M, editor. PLoS One. 2019;14(11):e0224727.

1231 78. Gu Y, Li T, Kapoor A, Major P, Tang D. Contactin 1: An important and emerging oncogenic protein
1232 promoting cancer progression and metastasis. Genes (Basel). 2020;11(8):1–22.

1233 79. Zhang Q, Zhou N, Wang W, Zhou S. A novel autocrine CXCL14/ACKR2 Axis: The achilles' heel of cancer
1234 metastasis? Clin Cancer Res. 2019;25(12):3476–8.

1235 80. Mathew E, Zhang Y, Holtz AM, Kane KT, Song JY, Allen BL, et al. Dosage-dependent regulation of
1236 pancreatic cancer growth and angiogenesis by Hedgehog signaling. Cell Rep. 2014;9(2):484–94.

1237 81. Eriksson P, Aine M, Veerla S, Liedberg F, Sjödahl G, Höglund M. Molecular subtypes of urothelial
1238 carcinoma are defined by specific gene regulatory systems. BMC Med Genomics. 2015;8(1):25.

1239 82. Dieters-Castator DZ, Rambau PF, Kelemen LE, Siegers GM, Lajoie GA, Postovit LM, et al. Proteomics-
1240 derived biomarker panel improves diagnostic precision to classify endometrioid and high-grade serous
1241 ovarian carcinoma. Clin Cancer Res. 2019;25(14):4309–19.

1242 83. Haupt Y, Maya R, Kazaz A, Oren M. Mdm2 promotes the rapid degradation of p53. Nature.
1243 1997;387(6630):296–9.

1244 84. Nissen NI, Karsdal M, Willumsen N. Collagens and Cancer associated fibroblasts in the reactive stroma
1245 and its relation to Cancer biology. J Exp Clin Cancer Res. 2019;38(1):115.

1246 85. Taber A, Christensen E, Lamy P, Nordentoft I, Prip F, Lindskrog SV, et al. Molecular correlates of
1247 cisplatin-based chemotherapy response in muscle invasive bladder cancer by integrated multi-omics
1248 analysis. Nat Commun. 2020;11(1):1–15.

1249 86. Wu S, Ou T, Xing N, Lu J, Wan S, Wang C, et al. Whole-genome sequencing identifies ADGRG6 enhancer

1250 mutations and FRS2 duplications as angiogenesis-related drivers in bladder cancer. Nat Commun.
1251 2019;10(1):1–12.

1252 87. Bell RJA, Rube HT, Xavier-Magalhães A, Costa BM, Mancini A, Song JS, et al. Understanding TERT
1253 promoter mutations: A common path to immortality. Mol Cancer Res. 2016;14(4):315–23.

1254 88. Lawson ARJ, Abascal F, Coorens THH, Hooks Y, O'Neill L, Latimer C, et al. Extensive heterogeneity in
1255 somatic mutation and selection in the human bladder. Science. 2020;370(6512):75–82.

1256 89. Shi MJ, Meng XY, Fontugne J, Chen CL, Radvanyi F, Bernard-Pierrot I. Identification of new driver and
1257 passenger mutations within APOBEC-induced hotspot mutations in bladder cancer. Genome Med.
1258 2020;12(1):85.

1259 90. Matheus LHG, Dalmazzo SV, Brito RBO, Pereira LA, De Almeida RJ, Camacho CP, et al. 1-Methyl-D-
1260 tryptophan activates aryl hydrocarbon receptor, a pathway associated with bladder cancer
1261 progression. BMC Cancer. 2020;20(1):869.

1262 91. Yu J, Lu Y, Muto S, Ide H, Horie S. The Dual Function of Aryl Hydrocarbon Receptor in Bladder
1263 Carcinogenesis. Anticancer Res. 2020;40(3):1345–57.

1264 92. Kang SU, Park JT. Functional evaluation of alternative splicing in the FAM190A gene. Genes and
1265 Genomics. 2019;41(2):193–9.

1266 93. Patel K, Scrimieri F, Ghosh S, Zhong J, Kim MS, Ren YR, et al. FAM190A deficiency creates a cell division
1267 defect. Am J Pathol. 2013;183(1):296–303.

1268 94. Lindskrog SV, Prip FF, Lamy P, Taber A, Clarice S, Nordentoft I, et al. An integrated multi-omics analysis
1269 identifies clinically relevant molecular subtypes of non-muscle-invasive bladder cancer. 2020;

1270 95. Stein JP, Lieskovsky G, Cote R, Groshen S, Feng AC, Boyd S, et al. Radical cystectomy in the treatment of
1271 invasive bladder cancer: Long-term results in 1,054 patients. J Clin Oncol. 2001;19(3):666–75.

1272 96. Mari A, Campi R, Tellini R, Gandaglia G, Albisinni S, Abufaraj M, et al. Patterns and predictors of
1273 recurrence after open radical cystectomy for bladder cancer: a comprehensive review of the literature.
1274 Vol. 36, World Journal of Urology. Springer Verlag; 2018. p. 157–70.

1275 97. Marine J-C, Dawson S-J, Dawson MA. Non-genetic mechanisms of therapeutic resistance in cancer. Nat
1276 Rev Cancer. 2020;20:1–14.

1277 98. Pivot XB, Bondarenko I, Dvorkin M, Trishkina E, Ahn J-H, Im S-A, et al. A randomized, double-blind,
1278 phase III study comparing SB3 (trastuzumab biosimilar) with originator trastuzumab in patients treated
1279 by neoadjuvant therapy for HER2-positive early breast cancer. J Clin Oncol. 2017;35(15_suppl):509–
1280 509.

1281 99. Rosenberg JE, O'Donnell PH, Balar A V., McGregor BA, Heath EI, Yu EY, et al. Pivotal Trial of Enfortumab
1282 Vedotin in Urothelial Carcinoma After Platinum and Anti-Programmed Death 1/Programmed Death
1283 Ligand 1 Therapy. J Clin Oncol. 2019;37(29):2592–600.

1284 100. Boni V, Sharma MR, Patnaik A. The Resurgence of Antibody Drug Conjugates in Cancer Therapeutics:
1285 Novel Targets and Payloads. Am Soc Clin Oncol Educ B. 2020;40:e58–74.

1286 101. Sheng X, Yan X, Wang L, Shi Y, Yao X, Luo H, et al. Open-label, Multicenter, Phase II Study of RC48-ADC,
1287 a HER2-Targeting Antibody–Drug Conjugate, in Patients with Locally Advanced or Metastatic Urothelial
1288 Carcinoma. Clin Cancer Res. 2021;27(1):43–51.

1289 102. Delmore JE, Issa GC, Lemieux ME, Rahl PB, Shi J, Jacobs HM, et al. BET bromodomain inhibition as a
1290 therapeutic strategy to target c-Myc. Cell. 2011;146(6):904–17.

1291 103. Goldstein JT, Berger AC, Shih J, Duke FF, Furst L, Kwiatkowski DJ, et al. Genomic activation of PPARG
1292 reveals a candidate therapeutic axis in bladder cancer. Cancer Res. 2017;77(24):6987–98.

1293 104. Peng H, Wisse E, Tian Z. Liver natural killer cells: Subsets and roles in liver immunity. Cell Mol Immunol.
1294 2016;13(3):328–36.

1295 105. Ochoa MC, Minute L, Rodriguez I, Garasa S, Perez-Ruiz E, Inogés S, et al. Antibody-dependent cell
1296 cytotoxicity: immunotherapy strategies enhancing effector NK cells. Immunol Cell Biol.
1297 2017;95(4):347–55.

1298 106. Moore AR, Rosenberg SC, McCormick F, Malek S. RAS-targeted therapies: is the undruggable drugged?
1299 Nat Rev Drug Discov. 2020;19(8):533–52.

1300 107. Meng XM, Nikolic-Paterson DJ, Lan HY. TGF-β: The master regulator of fibrosis. Nat Rev Nephrol.
1301 2016;12(6):325–38.

1302 108. Borthwick LA, Wynn TA, Fisher AJ. Cytokine mediated tissue fibrosis. Biochim Biophys Acta - Mol Basis
1303 Dis. 2013;1832(7):1049–60.

1304 109. van Dijk N, Gil-Jimenez A, Silina K, Hendricksen K, Smit LA, de Feijter JM, et al. Preoperative ipilimumab
1305 plus nivolumab in locoregionally advanced urothelial cancer: the NABUCCO trial. Nat Med.
1306 2020;26(12):1839–1844.

1307 110. Lv J, Li P. Mesothelin as a biomarker for targeted therapy. Biomark Res. 2019;7(1):18.

1308 111. Dunham I, Kundaje A, Aldred SF, Collins PJ, Davis CA, Doyle F, et al. An integrated encyclopedia of DNA
1309 elements in the human genome. Nature. 2012;489(7414):57–74.

1310 112. Roadmap Epigenomics Consortium, Kundaje A, Meuleman W, Ernst J, Bilenky M, Yen A, et al.
1311 Integrative analysis of 111 reference human epigenomes. Nature. 2015;518(7539):317–29.

1312

1313