Whole-genome mapping of APOBEC mutagenesis in metastatic urothelial carcinoma identifies driver hotspot mutations and a novel mutational signature

J. Alberto Nakauma-González[1,2,3], Maud Rijnders[3], Minouk T. W. Noordsij[1], John W. M. Martens[3], Astrid A.M. van der Veldt[3,5], Martijn P. J. Lolkema[3,4], Joost L. Boormans[2], Harmen J. G. van de Werken[1,2,6*]

1. Cancer Computational Biology Center, Erasmus MC Cancer Institute, University Medical Center Rotterdam, Rotterdam 3015 GD, the Netherlands
2. Department of Urology, Erasmus MC Cancer Institute, University Medical Center Rotterdam, Rotterdam 3015 GD, the Netherlands
3. Department of Medical Oncology, Erasmus MC Cancer Institute, University Medical Center Rotterdam, Rotterdam 3015 GD, the Netherlands
4. Currently employed by Amgen Europe BV
5. Department of Radiology & Nuclear Medicine, Erasmus MC, University Medical Center Rotterdam, Rotterdam 3015 GD, the Netherlands
6. Department of Immunology, Erasmus MC Cancer Institute, University Medical Center Rotterdam, Rotterdam 3015 GD, the Netherlands

*Correspondence: h.vandewerken@erasmusmc.nl, Cancer Computational Biology Center & Department of Immunology, Department of Urology, Erasmus MC Cancer Institute, University Medical Center; Rotterdam, the Netherlands. PO Box 2040, 3000 CA, NA-1218.

Abstract

Background

In urothelial carcinoma (UC), apolipoprotein B mRNA-editing enzyme catalytic polypeptide-like (APOBEC) enzymes are major contributors to the mutational burden. The APOBEC family mutates specific DNA sequences and hairpin-loop structures, creating Hotspot Mutations (ApoHM). The vast majority of ApoHM are passengers and only a few appear more frequently than expected (drivers). In this study, we aimed to characterize 115 whole-genomes of metastatic UC (mUC) to identify APOBEC mutagenic hotspot drivers of mUC.

Results

APOBEC-associated mutations (TpC context) were detected in 92% of mUC samples and were equally present across the genome. However, ApoHM were enriched in open chromatin and DNA hairpin-loop structures. The hairpin-loops were frequent targets of two hotspot mutations we called didymi (twins in Greek). Didymi were characterized by the APOBEC mutational signature SBS2, in conjunction with an uncharacterized mutational context (C>T mutations in ApC context), which was associated with DNA mismatch. Next, we developed a statistical framework that identified 0.40% of ApoHM as drivers of mUC, which affected the protein-coding regions of known driver genes *FGFR3*, *PIK3CA, RXRA,* and *TP53,* and non-coding regions near exons of potential novel driver genes. Our results and statistical framework were validated in independent cohorts of 23 non-metastatic UC and 442 metastatic breast cancers (mBC), showing that drivers are cancer-type-specific and only 0.07% of ApoHM were drivers of mBC.

Conclusion

By whole-genome DNA-sequencing analysis of mUC, we identified a novel mutational signature associated with APOBEC-targeted hairpin-loops and found that 0.07%-0.40% of ApoHM were drivers. Our study highlights the consequences of APOBEC in cancer development and may serve to develop novel targeted therapy options for APOBEC-driven mUC.

Keywords

APOBEC; breast cancer; didymi; driver mutations; hairpin-loops; hotspot mutations; metastasis; mutational signature; twin mutations; urothelial carcinoma

52 ## Background

53 Cancer genomes accumulate somatic mutations via different mutagenic processes and one of the most
54 common is attributed to the apolipoprotein B mRNA-editing enzyme catalytic polypeptide-like (APOBEC)
55 family [1]. APOBEC has a specific mutational signature, which is characterized by C>T/G mutations in
56 the TpC context and is captured in the SBS2 and SBS13 signatures, as defined by the *Catalogue Of*
57 *Somatic Mutations In Cancer* (COSMIC) [2]. In some tumor types with high APOBEC activity, the
58 contribution to the tumor mutational burden is substantial, which increases the neoantigen load favoring
59 response to immune checkpoint inhibitors [3, 4]. However, APOBEC is also responsible for the
60 emergence of driver mutations that contribute to cancer development as shown in mouse models [5].
61 Discriminating driver events from passenger events is essential to reconstruct the evolutionary history
62 of cancers and identify effective novel drug targets in APOBEC-driven tumors.

63 The mutational process of APOBEC has been extensively studied, revealing its preference for single-
64 stranded DNA structures that form hairpin-loops [6]. This characteristic of APOBEC could result in
65 identical somatic mutations in tumors from multiple patients, so-called hotspot mutations or hotspots.
66 Due to their high prevalence, these hotspots can erroneously be assigned as driver mutations, especially
67 in the non-coding area of the genome. However, the vast majority of mutations are passengers and do
68 not contribute to cancer development, and the same principle may also apply to hotspot mutations [7].
69 Although bioinformatic strategies to identify driver hotspot mutations have been developed [8, 9], the
70 unique characteristics of the APOBEC mutagenic process require specific considerations to account for
71 all co-variables accurately.

72 APOBEC-derived mutations are a dominant contributor to the mutational landscape in urothelial
73 carcinoma (UC). Therefore, we analyzed whole-genome DNA-sequencing data of 115 metastatic UC
74 (mUC) and matched blood samples [10] to identify driver hotspot mutations in the context of APOBEC
75 mutagenesis. The comprehensive characterization of APOBEC-enriched tumors identified a novel
76 mutational signature associated with DNA mismatch repair as well as genomic co-variates associated
77 with APOBEC-derived hotspot mutations (ApoHM), which we used to develop a statistical framework
78 and identify driver ApoHM. Furthermore, our findings were validated in whole-genomes of two
79 independent cohorts of 23 non-metastatic UC and 442 metastatic breast cancer (mBC) [11].

Results

APOBEC mutagenesis dominates the mutational landscape of urothelial carcinoma

The analysis of whole-genome sequencing (WGS) data of mUC and matched blood samples revealed a median of 20,667 (Q1=14,304, Q3=31,411) single nucleotide variants (SNVs) per tumor. mUC with a significant enrichment (E) for C>T mutations in TCW (W = A or T) context were considered APOBEC positive (92%). These tumors were further stratified according to APOBEC enrichment as APOBEC-high (41%; E>3), APOBEC-medium (33%; 2<E<3), and APOBEC-low (18%; E>1) (**Fig. 1a**). The median contribution of APOBEC COSMIC signatures (SBS2+SBS13) in APOBEC-high, -medium and -low tumors was 61%, 37% and 15%, respectively. For the remaining 8% of tumors lacking APOBEC mutations, the median APOBEC signature was <2%, potentially reflecting the noise of the mutational signature calling. We associated the APOBEC stratification with multiple factors to better assess the different APOBEC subtypes. Tumor purity, for instance, declined with increasing APOBEC mutagenesis (Additional file 1: **Fig. S1**). Moreover, age was associated with the enrichment of APOBEC mutations (Additional file 1: **Fig. S1**). The median clonal fraction of SNVs was lower in tumors with APOBEC mutations than in non-APOBEC tumors, suggesting higher tumor heterogeneity in APOBEC-enriched tumors (Additional file 1: **Fig. S1**). Localized hypermutation events (kataegis) strongly correlated with APOBEC enrichment (spearman r=0.80, p<0.001). Homologous recombination (HR) deficiency (n=3) was only present in APOBEC-low tumors (Fisher's exact p=0.005), while none of the patients with microsatellite instability (MSI; n=4) had evidence of APOBEC mutagenesis (Fisher's exact p<0.001). Structural variants were more frequent in APOBEC tumors than in non-APOBEC tumors (Additional file 1: **Fig. S1**). Additionally, APOBEC tumors had a higher ploidy (median ploidy = 3) and a higher number of genes affected by copy number alterations (CNA) than non-APOBEC tumors (Additional file 1: **Fig. S1**), suggesting genomic instability in APOBEC-driven mUC tumors. APOBEC mutagenesis was not associated with sex, the primary origin of mUC (upper tract versus bladder) or chromothripsis.

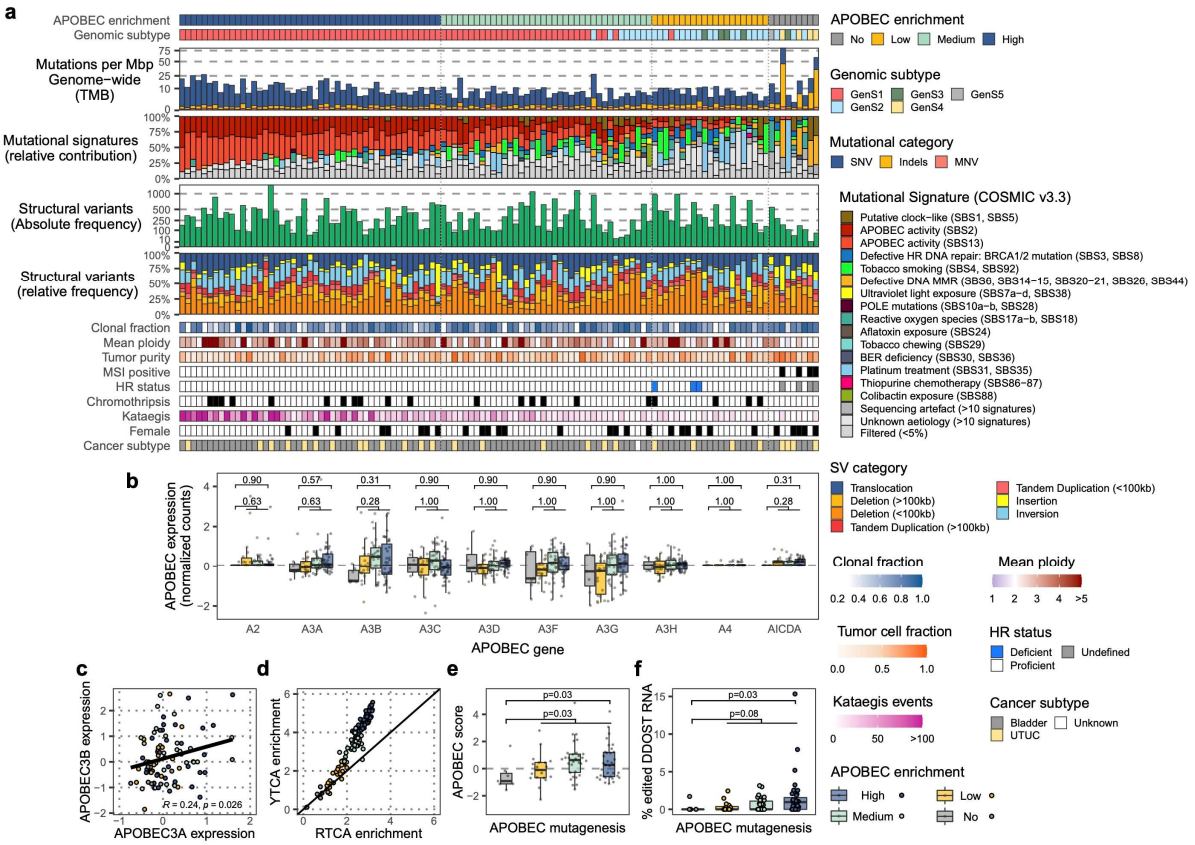


Fig. 1: Genomic landscape and APOBEC activity of metastatic urothelial carcinoma (n=115) stratified by APOBEC-enrichment

Whole-genome sequencing data of metastatic urothelial carcinoma (mUC) were classified according to the enrichment of APOBEC-associated mutations as having high, medium, low or no APOBEC enrichment. **(a)** The genomic features are displayed from top to bottom as follows: APOBEC mutagenesis; genomic subtype (GenS1-5) as previously described [10]; genome-wide tumor mutational burden (TMB); mutational signatures grouped by etiology, while both APOBEC signatures are shown separately; absolute frequency of structural variants (SV); relative frequency of SV categories; clonal fraction; ploidy; tumor purity; microsatellite instability (MSI) status; homologous recombination (HR) deficiency status; samples with at least one chromothripsis event; frequency of kataegis events; female patients and; primary origin of mUC (upper tract versus bladder). **(b)** Expression of *APOBEC* and *AICDA* genes in 90 samples with available RNA-sequencing data. **(c)** Pearson correlation of RNA expression of *APOBEC3A* and *APOBEC3B*. **(d)** Fold enrichment of C>T and C>G alterations in YTCA (related to APOBEC3A) and RTCA (related to APOBEC3B) context. **(e)** APOBEC score (normalized expression of *APOBEC3A*+*APOBEC3B*). **(f)** Percentage of mRNA C>U mutations in *DDOST* at position chr1:20981977. In (b), (e) and (f), the Wilcoxon rank-sum test was

applied to compare APOBEC tumors *vs.* non-APOBEC tumors. P-values were Benjamini-Hochberg corrected in (b).

APOBEC mutagenesis is an ongoing process in metastatic lesions of urothelial carcinoma

Next, we analyzed RNA-sequencing data of 90 matched samples of mUC. Pathway activity based on downstream gene expression, such as cell cycle or p53 was similar between the APOBEC groups (Additional file 1: **Fig. S2**). Similarly, analysis of APOBEC expression of all genes of the APOBEC family (*APOBEC1* was not expressed) revealed no significant differences between APOBEC and non-APOBEC tumors (**Fig. 1b**). We detected a weak positive correlation between the expression of *APOBEC3A* and *APOBEC3B* (**Fig. 1c**). To further investigate the mutagenic activity of both enzymes, the fold enrichment of C>T and C>G mutations, at DNA level, in the tetra-base YTCA (related to APOBEC3A; Y are pyrimidine bases) and RTCA (related to APOBEC3B; R are purine bases) context was calculated [12] (**Fig. 1d**). In both cases, YTCA and RTCA mutations did not correlate with expression of *APOBEC3A* or *APOBEC3B* (Additional file 1: **Fig. S3**). The lack of correlation might be linked to the heterogeneous expression of APOBEC enzymes that oscillate throughout the cell cycle [13, 14]. Furthermore, we detected that both, APOBEC3A and APOBEC3B, contributed to APOBEC-associated mutations (fold enrichment is above 1.0). Nevertheless, APOBEC3A appeared to be the main contributor as suggested in primary cancers [12, 15]. Considering the mRNA expression of both *APOBEC3A* and *APOBEC3B* enzymes, we calculated the APOBEC expression score (sum of the normalized expressions of *APOBEC3A* and *APOBEC3B*). It appeared that the level of APOBEC enrichment correlated with the APOBEC expression score (**Fig. 1e**). This analysis confirmed the link between APOBEC RNA expression at the time of biopsy and the historical accumulation of APOBEC-associated mutations in mUC that others have reported in primary urothelial carcinoma [16, 17].

Recently, it was proposed that edited *DDOST* mRNA can be used to measure ongoing APOBEC mutagenesis [18]. We found that the frequency of C>U alterations in the *DDOST* mRNA at position chr1:20981977 was enriched in tumors with APOBEC mutagenesis, with up to 15% of mRNA reads edited in one single sample (**Fig. 1f**). Additionally, we analyzed ongoing APOBEC mutagenesis in mUC, by WGS of eight tumors from patients who had undergone serial biopsies of metastatic lesions (Additional file 1: **Fig. S4)**. We observed that the APOBEC mutational signature was present in private mutations of the second biopsy of these patients, suggesting that APOBEC mutagenesis could be active

153 in the period between the first and second biopsy (Additional file 1: **Fig. S4a).** A lower cancer cell fraction
154 (normalized allele frequency by copy number and purity, see Methods) in private SNVs of the second
155 biopsy compared to shared (trunk) mutations confirms that these mutations were acquired more recently
156 as they are only present in a subpopulation of cancer cells (Additional file 1: **Fig. S4b**). The presence of
157 subclonal populations due to APOBEC mutagenesis may contribute to the ongoing evolution of UC in
158 the metastatic setting.

159

160 **APOBEC-associated hotspot mutations are enriched in highly accessible genomic regions**

161 The high resolution achieved by WGS, allowed us to investigate the enrichment of APOBEC and non-
162 APOBEC mutations (non-TpC context) in specific genomic regions. We found that the number of non-
163 APOBEC-associated SNVs, for instance, varied across the genome (**Fig. 2a**). When this distribution
164 overlapped with DNA accessibility and overall gene expression level, the frequency of non-APOBEC
165 mutations decreased in open chromatin (highly accessible regions) and highly transcribed regions (**Fig.**
166 **2b**). In contrast, the frequency of APOBEC-associated mutations was nearly constant across the
167 genome. When restricting the analysis to APOBEC tetra-base mutations, we found that the difference
168 in the distribution of Y/RTCA mutations across genomic regions decreases with the level of APOBEC
169 enrichment (**Fig. 2c**). Interestingly, in APOBEC-high tumors, YTCA mutations were evenly distributed.
170 In contrast, RTCA mutations were enriched in low DNA accessible and low transcribed regions, although
171 this enrichment was considerably less compared to tumors with lower levels of APOBEC mutations

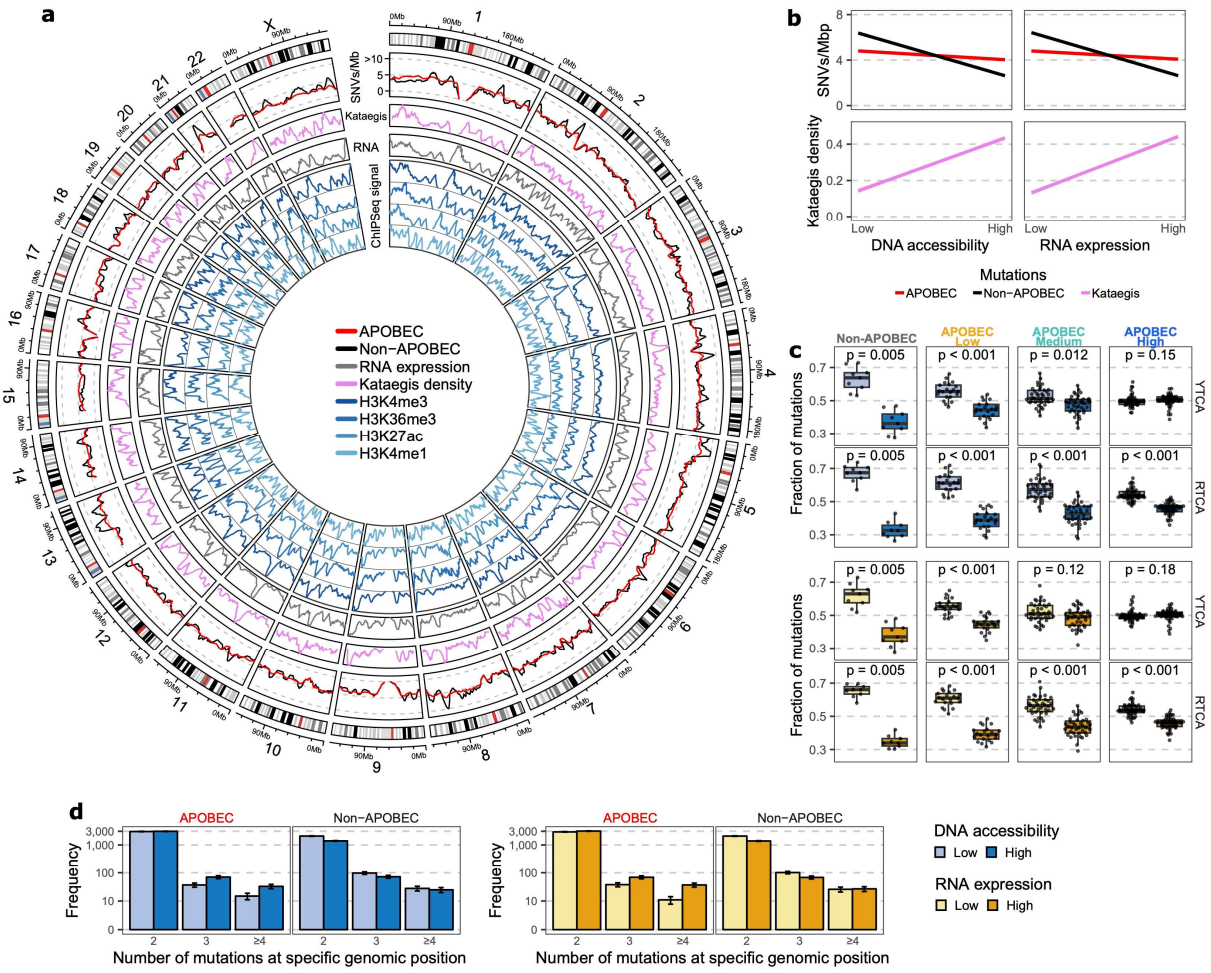


Fig. 2: Distribution of APOBEC-associated mutations across genomic regions of metastatic urothelial carcinoma

(a) Whole-genome sequencing data (n = 115) were analyzed to estimate the mean number of single nucleotide variants (SNVs) in windows of one mega base pair (Mbp) 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; the mutational load of APOBEC and non-APOBEC associated mutations (mutations in TpC or non-TpC context, respectively); The density of kataegis events; Average RNA counts (expression) from tumors with available RNA-sequencing data (n = 90); DNA accessibility estimation from different ChIPseq experiments of multiple histone marks from normal urothelial samples derived from ENCODE [19]. Peaks represent highly accessible DNA. **(b)** Linear regression of the mutational load for APOBEC- and non-APOBEC-associated mutations as well as the density of kataegis events across the genome with DNA accessibility and expression data. **(c)** Relative distribution of APOBEC YTCA and RTCA mutations across DNA-accessible and RNA expression regions. Samples were stratified according to the level of APOBEC mutagenesis. The Wilcoxon

187 signed-rank test was applied and p-values were Benjamini-Hochberg corrected. **(d)** Frequency of
188 hotspot mutations grouped according to APOBEC and non-APOBEC-associated mutations, and DNA
189 accessibility or RNA expression level.
190

191 Because of the high correlation between kataegis and APOBEC enrichment, we analyzed the
192 distribution of kataegis loci across the genome. Contrary to the overall distribution of SNVs, our data
193 suggest that kataegis events are more likely to occur in regions with high DNA accessibility and high
194 transcriptional activity (**Fig. 2b**). Moreover, we also evaluated the genome-wide distribution of all hotspot
195 mutations (two or more mutations in a specific genomic position), representing 0.35% of all mutated
196 genomic positions. We found that the frequency of highly recurrent (n≥4) ApoHM were enriched in high
197 DNA accessibility and high transcriptional active regions (**Fig. 2d**). Thus, while general APOBEC
198 mutagenesis seemed to occur uniformly across the genome, kataegis and ApoHM seemed to occur
199 more frequently in open chromatin and highly transcribed loci.

200

201 **Recurrent hotspot mutations correlate with APOBEC mutagenesis**

202 Next, we investigated the genomic consequence of hotspot mutations and found that the most frequent
203 hotspot mutations in mUC occurred in non-coding regions of the genome (**Fig. 3a**). Hotspot mutations
204 in the *TERT* promoter were present in 62% of tumors. In line with previous reports [20, 21], *TERT*
205 expression did not differ between tumors with hotspot mutations and those being *wildtype* (Additional
206 file 1: **Fig. S5a**). However, differential gene expression analysis showed that tumors with hotspot
207 mutations in the *TERT* promoter had high expression of genes related to the biological oxidation
208 pathway (Additional file 1: **Fig. S5b,** Additional file 2: **Table S1**). Besides *TERT*, other frequent hotspot
209 mutations were identified in the non-coding regions near *ADGRG6* (40%), *PLEKHS1* (33%), *LEPROTL1*
210 (20%), and *TBC1D12* (15%). Similarly, these hotspot mutations did not affect the expression of these
211 genes but were associated with transcriptomic effects in several biological pathways (Additional file 1:
212 **Fig. S5,** Additional file 2: **Table S1**). These hotspot mutations strongly correlated with enrichment for
213 APOBEC-associated mutations, suggesting their origin in APOBEC mutagenesis (**Fig. 3b**).

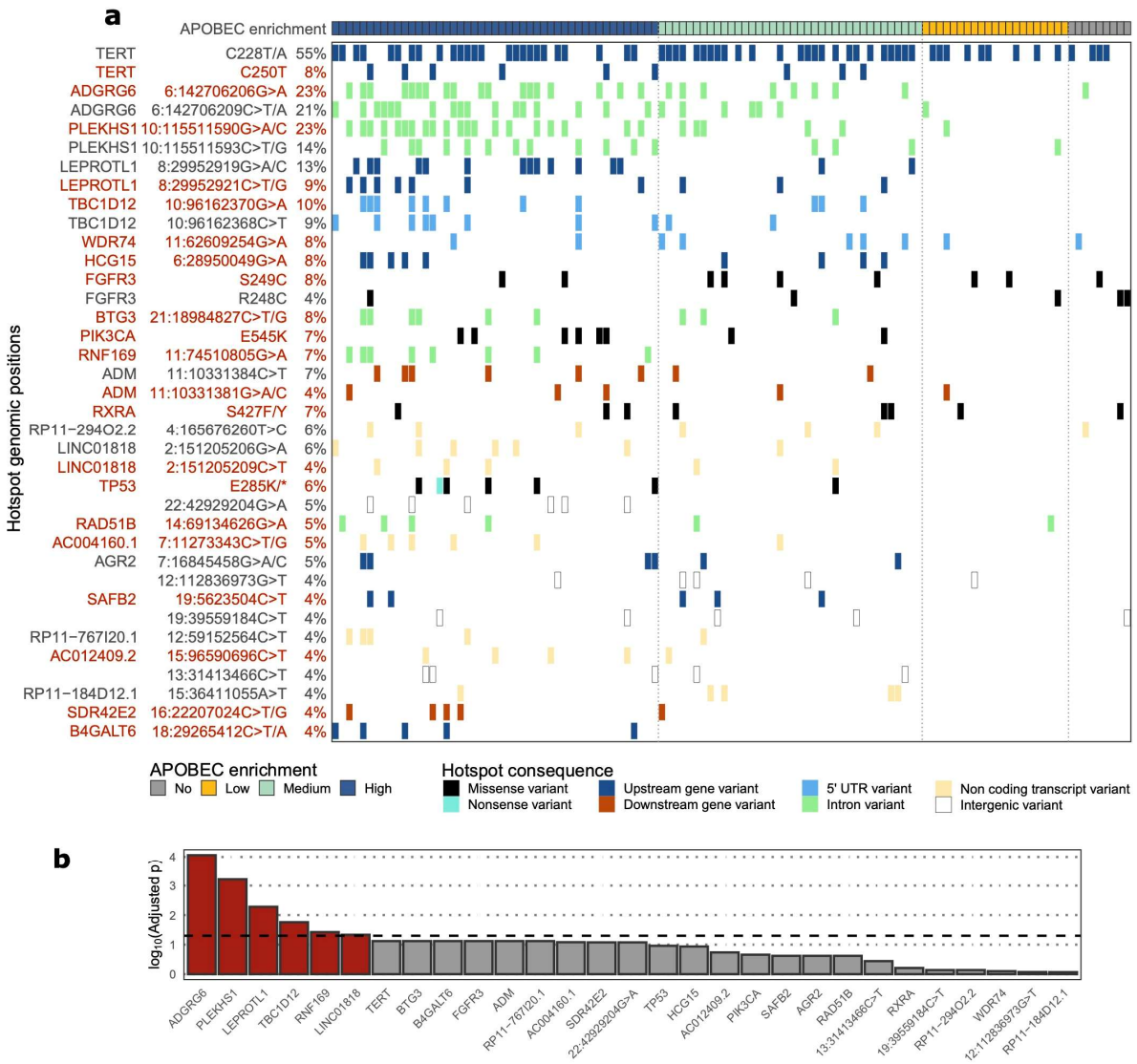


Fig. 3: Recurrent hotspot mutations of metastatic urothelial carcinoma correlate with APOBEC mutagenesis

(a) Overview of recurrent hotspot mutations present in at least five samples, including all substitutions occurring in the same genomic position. Hotspot mutations occurring in the TpC context are highlighted in red. **(b)** The association of hotspot mutations and APOBEC fold enrichment (continuous values) was interrogated with a logistic regression analysis applying the Wald test. P-values were corrected using the Benjamini-Hochberg method and ordered accordingly. Bars above the dashed line ($-\log_{10}(0.05)$) are statistically significant and are indicated in red.

224 All frequent hotspot mutations in the coding region have been previously described and affected known
225 driver genes: *FGFR3* S249C/R248C (8%, 4%), *PIK3CA* E54K (7%), *RXRA* S427F/Y (7%) and *TP53*
226 E285K/* (6%). Comparing the expression of these known driver genes affected by hotspot mutations
227 *vs.* the wildtype, only *FGFR3* hotspot mutations significantly affected the expression of this gene
228 (Additional file 1: **Fig. S5a**).

229

230 **Hairpin-loops are targets of twin hotspot mutations called Didymi**

231 We noticed that the hotpot mutations near *ADGRG6*, *PLEKHS1*, *LEPROTL1* and *TBC1D12* are located
232 within DNA hairpin-loop structures. It is known that DNA hairpin-loops are targets of APOBEC3A (**Fig.**
233 **4a**) [22, 23], therefore, we predicted DNA hairpin-loops for all mutated genomic positions (see Methods).
234 The predicted hairpin-loops near *ADGRG6*, *PLEKHS1*, *LEPROTL1* and *TBC1D12* are each affected by
235 two hotspot mutations, which are referred to as twin mutations [23]. Moreover, we noticed that the twin
236 mutations were not mutually exclusive, which differs from the hotspot mutations in the *TERT* promoter
237 (mutual exclusivity test, $p < 0.001$). Only the twin hotspot mutations in *TBC1D12* co-occurred more
238 frequently than expected among APOBEC-high tumors ($p = 0.02$). Further analysis of co-occurred twin
239 hotspot mutations revealed that very few had identical variant allele frequencies, suggesting that most
240 twin mutations in the same tumor occurred in independent events as they were also found on different
241 alleles (Additional file 1: **Fig. S6**).

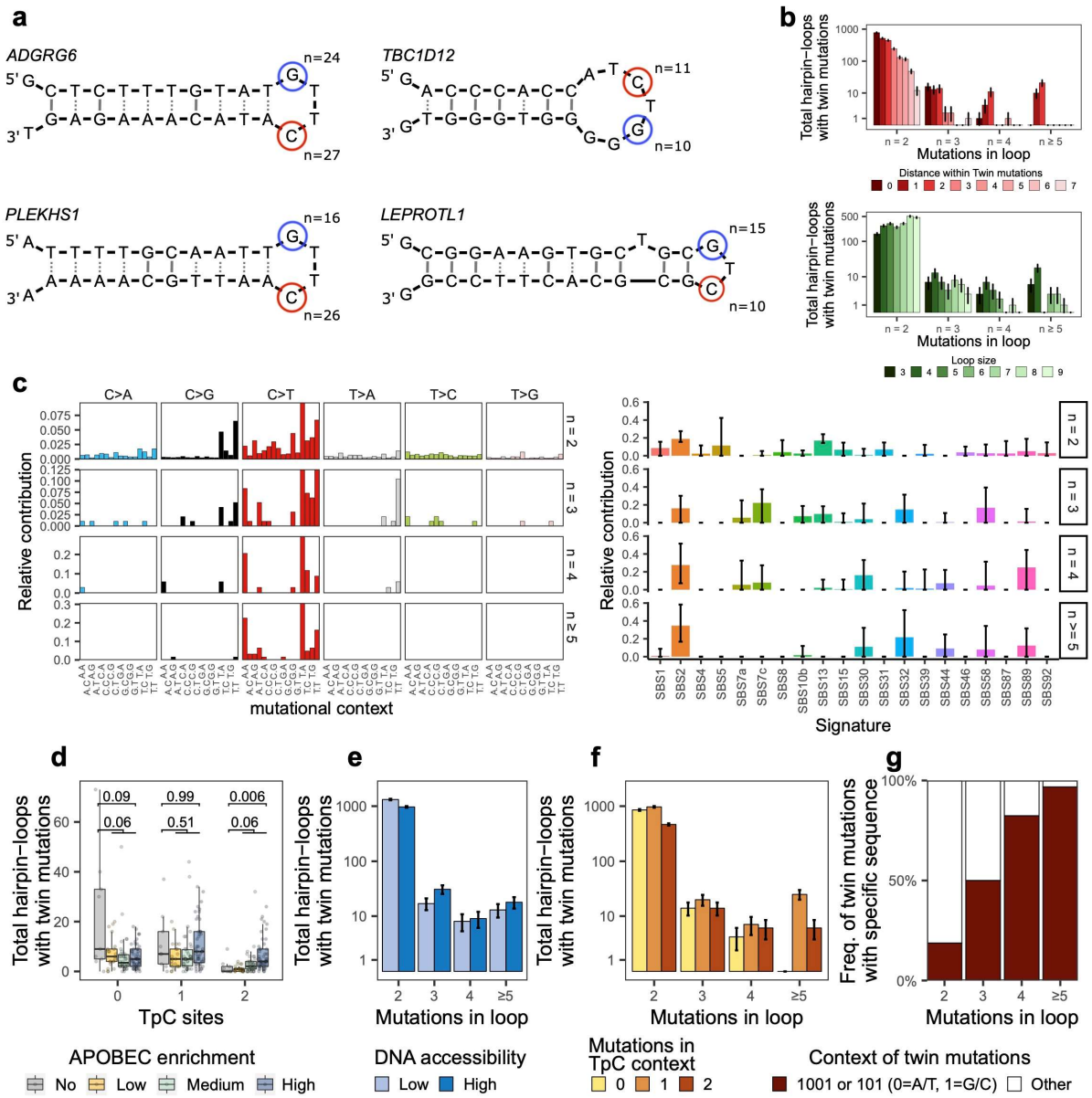


Fig. 4: Genomic characteristics of twin hotspot mutations in metastatic urothelial carcinoma **(a)** Hairpin-loop structures affected by frequent hotspot mutations in *ADGRG6*, *TBC1D12*, *PLEKHS1* and *LEPROTL1*. The positions of hotspot mutations are marked in red for TpC context and blue otherwise. **(b)** Distribution of twin mutations according to the distance between twin mutations and loop size. **(c)** Mutational signatures (COSMIC v3.3) of twin mutations according to their frequency. The stability of the signature call was tested by applying 1,000 bootstrap iterations. Only SBS2 was very stable in highly frequent (n ≥ 5) twin mutations. **(d)** Frequency distribution of hairpin loops affected by twin mutations according to APOBEC mutagenesis, **(e)** DNA accessibility, **(f)** number of mutations in TpC context within a loop and **(g)** DNA sequence between twin mutations. Wilcoxon rank-sum test was

252 applied to compare APOBEC *vs* non-APOBEC tumors and p-values were Benjamini-Hochberg
253 corrected.

254

255 Next, we investigated the properties and origin of these twin mutations. A comprehensive analysis of all
256 DNA hairpin-loop structures in the human genome affected by two mutations in their loops revealed
257 2,387 twin mutations (4,774 altered genomic positions), representing 0.16% of all mutated genomic
258 positions. The distance between twin mutations varied, but when the frequency of mutations increased,
259 the distance decreased to mainly one or two bases and the loop sizes to mainly three to four bases (**Fig.**
260 **4b**). Additional examination of the 96 tri-nucleotide contexts of all twin mutations revealed that both
261 APOBEC COSMIC signatures, SBS2 and SBS13, were dominant (**Fig. 4c**). However, at higher
262 mutational frequencies (n ≥ 5), only signature SBS2 remained. We also observed a secondary signature
263 of C>T mutations in the ApC context that does not resemble any known COSMIC signature (Additional
264 file 1: **Fig. S7,** Additional file 2: **Table S2**). The absolute contribution of this signature was similar across
265 all APOBEC tumors and its prevalence in the mUC cohort correlated with spontaneous deamination
266 (SBS1) and defective DNA mismatch repair signatures (Additional file 1: **Fig. S8**; SBS6, SBS15, SBS20
267 and SBS26).

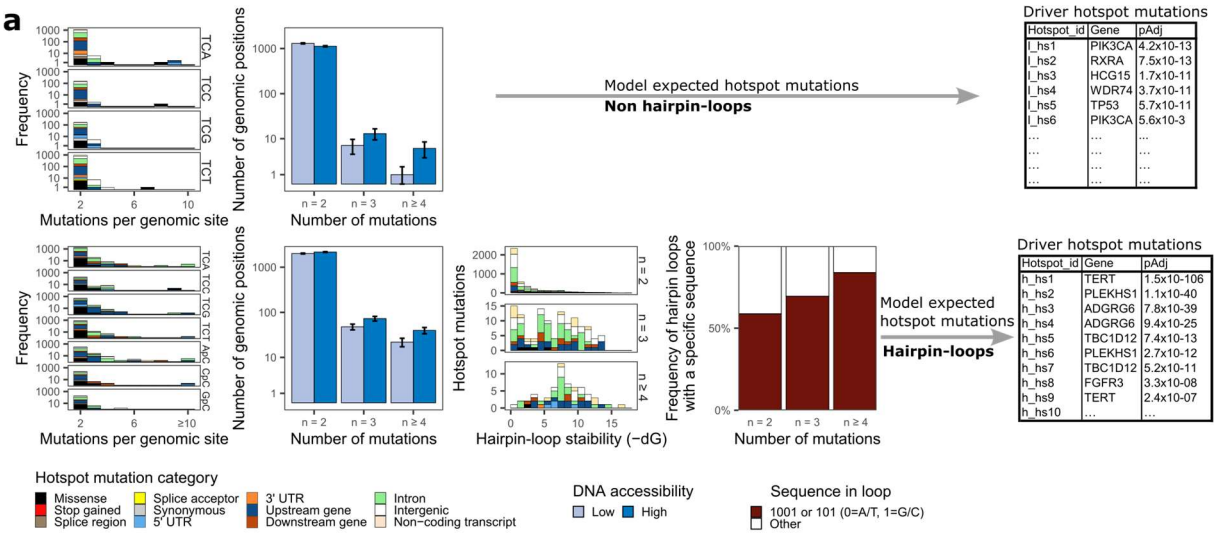
268 Furthermore, APOBEC-driven tumors were enriched for twin mutations occurring only in the TpC context
269 (**Fig. 4d**). Contrary to the general pattern of ApoHM (enriched in DNA accessible regions), twin
270 mutations with a high number of alterations were similarly distributed between high and low DNA
271 accessible regions (**Fig. 4e**).

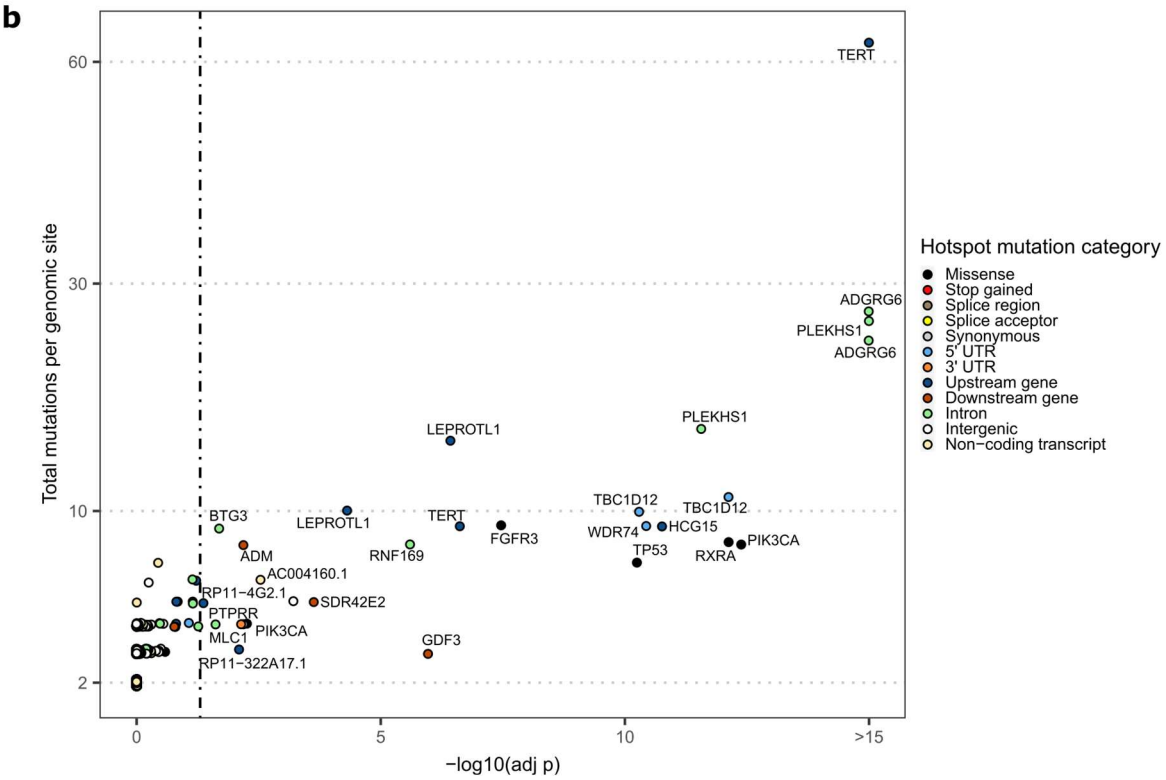
272 Additionally, we found that at higher mutational frequency, at least one of the twin mutations occurs in
273 the TpC context (**Fig. 4f**) and the sequence between the two is 1001 or 101 (0 = A/T, 1 = G/C; underlined
274 are the positions of the twin mutations) (**Fig. 4g**). Because of the unique characteristics of frequently
275 affected twin mutations, we named them didymi (twins in Greek). In summary, didymi are two C>T
276 hotspot mutations found in DNA hairpin-loops separated by one or two A/T base-pairs in which at least
277 one of the twin mutations is located in TpC context and the other in NpC (N = any base-pair; most N
278 bases are A or T). Applying this definition, we identified 231 didymi in the mUC cohort (Additional file 2:
279 **Table S3**).

280

Driver hotspot mutations in urothelial carcinoma

After identifying several hotspot mutations that could be attributed to APOBEC activity, we assessed whether these hotspot mutations had a selective advantage (drivers) or not (passengers). Recent attempts relying on the stability of hairpin-loops have been proposed to differentiate passengers from driver ApoHM [8, 22]. We confirmed that a more stable loop (Gibbs free energy ΔG; see Methods) leads to a higher number of alterations (**Fig. 5a**). Taking this into account, we developed a statistical model to identify driver hotspot mutations considering not only the stability of hairpin-loops but also the tri-nucleotide context, DNA accessibility and the potential for didymi via sequence in the loop (**Fig. 5a**).





290 **Fig. 5: Driver hotspot mutations associated with APOBEC mutagenesis in urothelial carcinoma**
291 **(a)** Frequency distribution of variables (trinucleotide context, DNA accessibility, DNA hairpin-loop
292 stability and sequence in the loop) considered to identify hotspot mutations that were more frequently
293 mutated than expected outside (top) and within (bottom) hairpin-loop structures. **(b)** Driver hotpot
294 mutations were estimated separately for mutations outside and within hairpin loops. Per group, p-values
295 were corrected using the Benjamini-Hochberg method.

296 Putative ApoHM were divided into those located outside or inside the loop of DNA hairpin-loop
297 structures. In case of ApoHM outside hairpin-loops, only those in the TpC context were considered. For
298 ApoHM within loops, all alterations in TpC, ApC, CpC and GpC context were included in the analysis to
299 account for didymi. In case of TpC, the distribution of hotspot mutations in the trinucleotide context was
300 considered. A background distribution was modeled as a Poisson process and the significant enrichment
301 of mutations in a particular genomic site was estimated.

302 We identified 0.40% (n = 27) of ApoHM as drivers (adjusted $p < 0.05$; **Fig 5b,** Additional file 2: **Table**
303 **S4**). Known driver genes affected by hotspot mutations included coding alterations in *TP53, PIK3CA,*
304 *FGFR3, RXRA* and the *TERT* promoter. All other putative driver ApoHM affected non-coding regions
305 including didymi in *ADGRG6*, *PLEKHS1*, *TBC1D12* and *LEPROTL1* proposed as drivers by other
306 studies [9, 24]. Other potential driver ApoHM include *RNF169* (involved in DNA damage repair) [25],
307 *BTG3* (angiogenesis) [26], *ADM* (adrenomedullin; vasodilator) [27], *GDF3* (regulation of TGF-beta) [28]
308 and WDR74 (ribosome biogenesis) [29].

309 To validate our method and confirm the driver ApoHM assessment, we used an independent cohort of
310 non-metastatic urothelial carcinoma of the bladder (n = 23) of the pan-cancer analysis of whole-genomes
311 (PCAWG) study [30] (Additional file 1: **Fig. S9**). This analysis confirmed the previously identified ApoHM
312 as potential cancer drivers of urothelial carcinoma. Moreover, in this cohort, 96% of tumors were
313 APOBEC-driven, APOBEC enrichment correlated with kataegis and the largest group (35%) had high
314 enrichment for APOBEC-associated mutations.

315

316 **Driver hotspot mutations in metastatic breast cancer**

317 To test our statistical framework in other cancer types and to evaluate if our findings were UC-specific,
318 we analyzed a cohort of 442 mBC (**Fig. 6**) [11]. Similar to UC, breast cancer is commonly affected by

319 APOBEC mutagenic activity [1]. We identified APOBEC-enriched tumors in 76% of patients, and only
320 19% of mBC tumors were APOBEC-high. In most patients (39%), tumors were classified as APOBEC-
321 low and were enriched for HR deficiency (Fisher's exact $p < 0.001$).

322 The most frequent coding hotspot mutations affected *PIK3CA*, *ESR1* and *AKT1*. Twin mutations in
323 hairpin-loops displayed a similar mutational signature as those in mUC, including the APOBEC signature
324 SBS2 in conjunction with the uncharacterized C>T mutations in ApC context that define didymi
325 (Additional file 1: **Fig. S10a-b**). Didymi in *PLEKHS1* and *ADGRG6* were the most frequent non-coding
326 hotspot mutations, while *LEPROTL1* and *TBC1D12*, two other didymi frequently found in mUC, only
327 affected ≤1% of mBC. A total of 694 didymi were identified in mBC (Additional file 2: **Table S5**), but only
328 19 (2.7%) were shared with mUC (Additional file 1: **Fig. S10c**).

329 Our analysis revealed 51 driver ApoHM in APOBEC-enriched mBC (**Fig. 6b,** Additional file 2: **Table S6**),
330 representing only 0.07% of all ApoHM. Drivers included missense hotspot mutations in *PIK3CA*, *AKT1*
331 and *TP53*, and hotspot mutations outside of the protein-coding region of *MAPKAPK2*, *STAG1* and
332 including didymi in *PLEKHS1* and *ADGRG6*. In contrast to mUC, and despite being one of the most
333 frequently affected genes by hotspot mutations in mUC, *LEPROTL1* was not a driver of mBC. This
334 analysis suggests that driver mutations derived from APOBEC activity are cancer-type-specific.

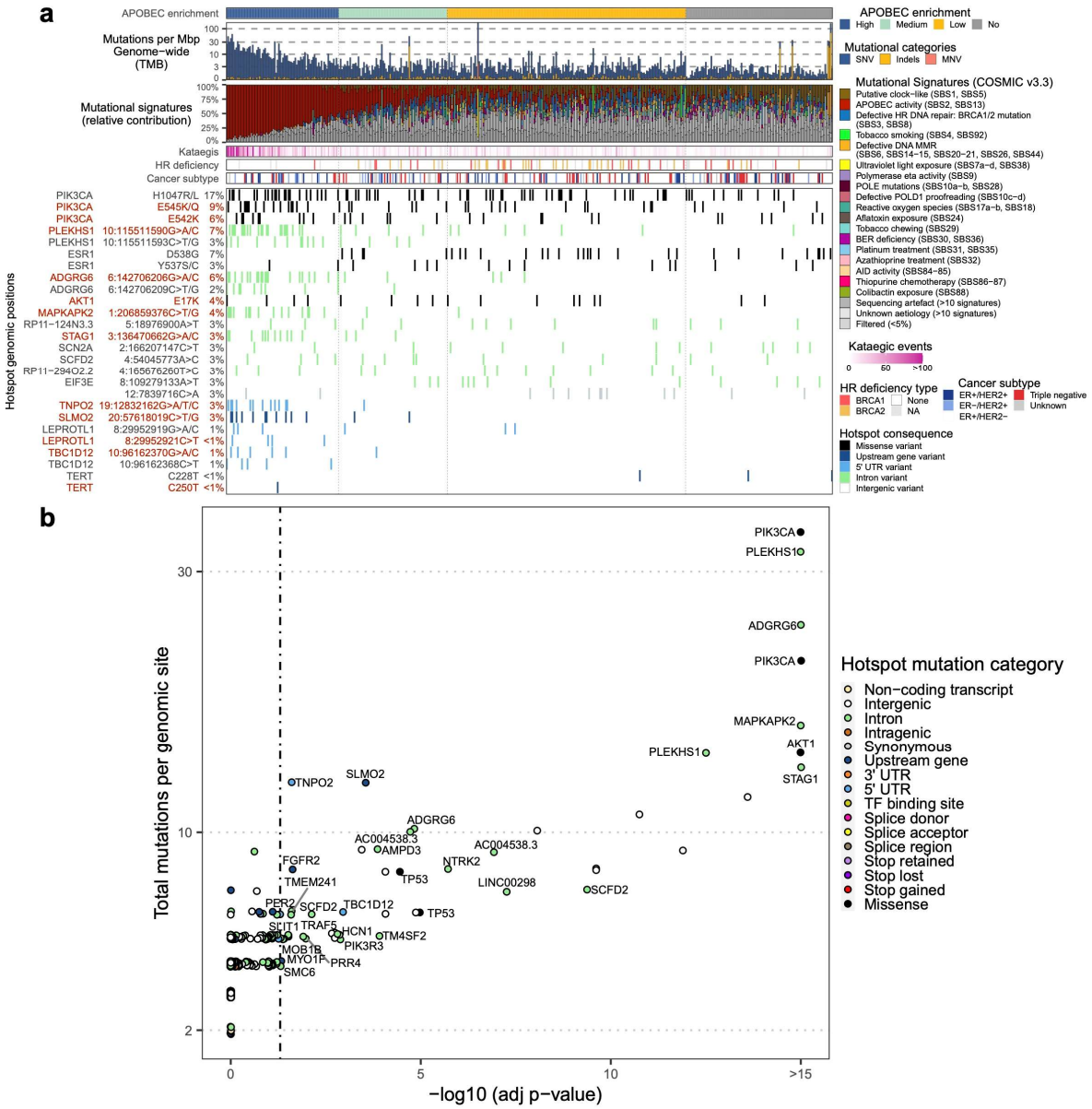


Fig. 6: Genomic landscape and driver hotspot mutations associated with APOBEC mutagenesis in metastatic breast cancer

(a) Whole-genome sequencing data from 442 metastatic breast cancer were analyzed and patients were classified according to the enrichment of APOBEC-associated mutations. The genomic features are displayed from top to bottom as follows: APOBEC mutagenesis; genome-wide tumor mutational burden (TMB); COSMIC mutational signatures; frequency of kataegis; homologous recombination (HR) deficiency and its probable origin; cancer subtype and; the most frequent hotspot mutations (known hotspot mutations in *LEPROTL1*, *TBC1D12* and *TERT* were also included). **(b)** Putative driver hotspot mutations in APOBEC-enriched breast cancer. P-values were adjusted using the Benjamini-Hochberg method.

346 ## Discussion

347 In this study, we describe the genomic landscape of APOBEC-driven tumors, characterize ApoHM and
348 identify potential cancer drivers in mUC. The in-depth analysis of 115 whole-genomes of mUC identified
349 chromatin accessibility, hairpin-loop stability and specific sequences within the hairpin-loop as variables
350 associated with ApoHM. These variables, in combination with the mutational context, were used to
351 identify ApoHM that were more frequently mutated than expected by chance.

352 The substrate of APOBEC enzymes is single-stranded DNA (ssDNA) [1], which has led to the following
353 conflicting hypotheses: 1) APOBEC enzymes are mainly active during replication [31] and 2) APOBEC
354 is mainly active in open chromatin and transcriptionally active genomic regions [32]. The equal
355 distribution of all APOBEC-associated mutations across genomic regions supports the hypothesis that
356 these mutations had been generated during replication when APOBEC enzymes have equal access to
357 ssDNA across the genome [31]. However, kataegis which has previously been linked to APOBEC
358 activity [33], and ApoHM were enriched in high DNA accessible and highly transcribed regions. This
359 observation reconciliates both views claiming that APOBEC is active during DNA replication (non-
360 clustered and non-hotspot mutations) and transcriptionally active regions (clustered and hotspot
361 mutations). Additionally, we observed that APOBEC3A-preferred YTCA mutations are dominant in mUC
362 and are evenly distributed across genomic regions. This result is in line with experimental observations
363 in human cancer cell lines [33], suggesting that APOBEC3A is the main driver of APOBEC mutagenesis.
364 For YTCA mutations as well as for APOBEC3B-preferred RTCA mutations, the gap between the number
365 of mutations across genomic regions is smaller at higher APOBEC mutagenesis, which strongly
366 suggests that APOBEC enzymes do not have a preference for specific genomic regions. However, the
367 enrichment of highly frequent ApoHM (n≥4) in open chromatin, of which many are drivers according to
368 our analysis, may imply a functional effect of these putative driver mutations occurring near gene
369 regulatory elements [24].

370 The extensive examination of ApoHM in hairpin-loops revealed twin mutations we termed didymi that
371 are characterized by a unique mutational pattern. Didymi comprises the APOBEC SBS2 signature and
372 an unknown signature delineated by C>T mutations in the ApC context. It is remarkable to see only
373 one of the APOBEC signatures in didymi loci, when they usually appear together in tumor samples
374 with APOBEC mutagenesis [34, 35]. The fact that only C>T mutations that characterize SBS2 are
375 present in didymi suggests that these mutations may arise predominantly by replication across the

376 uracil bases [36, 37] and that the mechanisms to generate C>G mutations that characterize SBS13
377 are not operational in this context. Furthermore, the unknown mutational signature of didymi correlates
378 with spontaneous deamination (SBS1) and defective DNA mismatch repair signatures (SBS6, SBS15,
379 SBS20 and SBS26), suggesting a potentially different mechanism in ApC from the putative TpC
380 APOBEC mutations. Although there is a strong correlation with APOBEC mutagenesis [23], it is
381 unclear if both mutations in didymi loci are direct targets of APOBEC3A or whether the non-TpC
382 mutations are just the result of spontaneous deamination followed by DNA mismatch repair.

383 Compared to breast cancer, twice as many bladder cancer tumors were APOBEC-high (19% vs. 41%).
384 Most driver ApoHM were cancer-specific, possibly reflecting different selective pressures that each
385 cancer type endured. In both tumor types, APOBEC-low patients had enrichment for HR deficiency,
386 while APOBEC-high tumors had high tumor mutational burden which may indicate different treatment
387 options for these two groups of patients with different levels of APOBEC mutagenesis [38-41].

388 The in-depth analysis performed in this study to characterize the mutational landscape of APOBEC
389 mutagenesis, revealed the correlation of ApoHM with the stability of DNA hairpin-loops, DNA
390 accessibility and the potential for didymi associated with these recurrent mutations. These features are
391 key to modeling the background distribution of hotspot mutations and identifying potential drivers. Most
392 potential driver ApoHM were in the non-protein coding regions, including didymi. The similar frequency
393 of these drivers in the metastatic and primary settings of UC indicates a general phenomenon in UC
394 and the drivers could cause early events of tumorigenesis of UC. APOBEC-associated mutations have
395 also been identified in normal tissue [42, 43], however, it is unclear if the driver ApoHM we report here
396 are also present in healthy tissue and to what extent they contribute to cancer development from
397 normal cells. Nevertheless, experimental validation will be needed to confirm the cancer driver status
398 of these hotspot mutations.

399 Conclusion

400 In this study, we characterized the genomic landscape of APOBEC-driven mUC and identified novel
401 mechanisms of genomic alteration patterns associated with APOBEC mutagenesis. The mutational
402 signatures associated with DNA hairpin-loops targeted by APOBEC in two distinct hotspot positions are
403 unique, demonstrating the exclusive mutational signature of APOBEC-derived hotspot mutations. These
404 findings were confirmed in non-metastatic UC and in metastatic breast cancer. Additional studies are

needed to clarify the role of APOBEC in these recurrent twin mutations. Also, the enrichment of ApoHM and kataegis in high DNA accessible regions, suggests a different mechanism compared to the general APOBEC mutagenesis (non-hotspot mutations) that seems to occur independently of genomic regions, which may be linked to different mechanisms of APOBEC3A and APOBEC3B [14]. As APOBEC is a major source of hotspot mutations, it is crucial to identify those in coding and non-coding regions of whole-genomes that may play an important role in cancer development. The statistical framework we developed could aid to identify potential driver hotspot mutations derived from APOBEC activity which may offer novel targeted therapy options for APOBEC-driven cancer patients.

415 # Methods

416 **Patient cohort**

417 The mUC cohort of this study has been previously described (NCT01855477 and NCT02925234) [10].
418 In short, patients with advanced or mUC were prospectively enrolled in these multicenter clinical trials
419 and were scheduled for 1st or 2nd line palliative systemic treatment. Following protocols of the Hartwig
420 Medical Foundation (HMF) [44], WGS, with a depth close to 100X [10], was successfully performed on
421 DNA from freshly obtained biopsies from metastatic sites, and matched RNA-sequencing (RNA-seq)
422 was available for 90 patients (97 samples). Sequential biopsies of a metastatic lesion taken at the time
423 of clinical or radiological disease progression from eight patients were additionally sequenced. Similarly,
424 the mBCa cohort has been previously described [11], and the DNA and RNA extraction and sequencing
425 were performed according to the HMF protocols [44].

426

427 **Whole-genome sequencing and analysis**

428 Alignment and pre-processing of WGS data, detection of mutational signatures, genomic subtypes, HR
429 deficiency, MSI, structural variants, chromothripsis events, kataegis, APOBEC mutagenesis, and
430 pathway activity have been previously described [4, 10, 44-46]. APOBEC enriched tumors (adj. $p < 0.05$,
431 otherwise non-APOBEC tumors) were classified as high when the fold enrichment (E) for C>T and C>G
432 mutations in T<u>C</u>W (W = A or T) context was $E \geq 3$, medium when $2 \leq E < 3$ and low when $E < 2$. Similarly,
433 the fold enrichment for C>T and C>G mutations in the tetra YT<u>C</u>A (Y = T or C) and RT<u>C</u>A (R = G or A)
434 context was calculated.

435

436 **Clonal fraction and cancer cell fraction**

437 The clonal fraction of mutations was estimated as previously described [47]. Correcting for tumor purity
438 and copy number, the variant copy number n_{SNV} of each SNV was calculated as follows

439 $$n_{SNV} = \frac{f_m}{p}[pC_t + (1-p)C_h], \quad (1)$$

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

442 autosomes and 1 for allosomes). In this study, mutations were considered clonal when the variant copy
443 number was >0.75.

444 To identify the proportion of cancer cells carrying a specific mutation, the cancer cell fraction (CCF) was
445 estimated as previously described [48, 49]. Given the number of reference and mutant reads and
446 assuming binomial distribution, we estimated the expected number of allelic copies (n_{chr}) carrying the
447 observed SNV resulting from f_m values when the mutation is present in 1, 2, 3,…, N_{chr} allelic copies. The
448 resulting estimated n_{chr} with the maximum likelihood is used to calculate the CCF as n_{SNV}/n_{chr}.

449

450 **Mutational load across genomic regions**

451 The genome was divided into regions (bins) of one mega base-pair (Mbp). The number of SNVs was
452 counted in each bin, and the mean number of SNVs was estimated from the entire cohort. These values
453 represented the average SNVs/Mbp reflecting the mutational load in each genomic region. The average
454 SNVs/Mbp was smoothed by applying a moving average with a window of k = 3. For visualization
455 reasons, in Fig. 2 a k = 9 was used.

456

457 **DNA accessibility estimation (ChIPseq)**

458 ChIPseq data for healthy urinary bladder and breast cancer (H3K4me1, H3K4me3, H3K36me3 and
459 H3K27ac) were downloaded from the ENCODE portal (https://www.encodeproject.org) to our local
460 server. The bed.gz files were imported with narrowPeak format for analysis. Only peaks with q < 0.05
461 were kept for analysis. The signal of each experiment was divided into regions of one Mbp, and a moving
462 average with k = 3 bins was applied. The signals were normalized using the mean and standard
463 deviation. This procedure was applied to each chromosome. The sum of all four ChIPseq experiments
464 was considered an approximation of DNA accessibility. High DNA accessible regions (open chromatin)
465 had values above the median considering the whole genome. All other regions were considered to be
466 of low DNA accessibility (condensed chromatin).

467

468 **Detection of hairpin loops**

469 All SNVs were assessed to determine whether they occur in the loop of hairpin-loop structures and their
470 thermodynamic stability. A total of 50 bases upstream and downstream of the mutation site were

471 considered. The minimum length of the stem was 2 base-pairs and the minimum and maximum loop
472 size was 3 and 10 bases, respectively (not considering the closing base-pair). Since multiple
473 configurations are possible, only the structure with the highest stability was considered. One mismatch
474 was allowed which could be either a non-matching base-pair or a single nucleotide bulge loop.

475

476 **Stability of hairpin loops**

477 We implemented the nearest neighbor stability algorithm (NNSA) to estimate the thermodynamic
478 stability of hairpin-loops . This algorithm calculates the Gibbs free energy (ΔG) of the DNA hairpin-loop
479 structure based on known biophysics properties of the base pairs and their interactions [50, 51].

480 The NNSA was applied to the target DNA sequence allowing mismatches in the stem. Each base-pair
481 contributes to the stability of the stem considering the immediate neighboring base-pairs. Adding the
482 local $\Delta G's$ calculated per base and accounting for the size and sequence of the loop results in a final ΔG
483 for the whole hairpin-loop structure. All parameters are available in the literature and the UNAFold web
484 server was used to infer parameters for mismatches [50-52]. For loops larger than 4, no data was
485 available for specific sequences and only the loop size was considered.

486

487 **Driver hotspot mutations**

488 To identify driver hotspot mutations, all genomic positions with 2 or more mutations were considered for
489 analysis. Hotspot mutations were divided into two groups either located within loops of hairpin-loops or
490 outside of these DNA structures. For hotspot mutations outside of loops, only those in the TpC context
491 were considered, as these are likely initiated by APOBEC enzymes. For hotspot mutations falling within
492 loops, all hotspot mutations in NpC (N = any base) context were considered, as APOBEC3A may also
493 mutate these bases that are not in the TpC context [23].

494 In cancer, only a few somatic mutations are drivers, while the vast majority are passenger mutations
495 [53]. Under this consideration, we modeled the distribution of the remaining hotspot mutations as a
496 Poisson process per focal hotspot mutation. For more accurate modeling, we considered the tri-
497 nucleotide context TCW (TCA, TCC, TCG, TCT). In the case of hotspot mutations that do not occur in
498 hairpin loops, DNA accessibility was used as a predictor variable that can influence the distribution of
499 hotspot mutations. We modelled this in R as model_TCW_noloop=glm(n_mut_genpos+DNA_access).

500 Where, n_muts_genpos is a vector with the number of mutations per genomic position linked to the DNA
501 accessibility region (DNA_access). Accessibility regions were divided into 10 regions based on
502 percentiles. Since DNA-accessibility varies per tumor-tissue of origin [54, 55], this was estimated for
503 mUC and mBC using ChIPseq experiments from normal tissue as described above.

504 In the case of hotspot mutations within hairpin-loops, the model was extended to include mutations in
505 non-TCW context (grouped as ApC, CpC or GpC context), the hairpin loop stability (hairpin_stab) and
506 the DNA sequence in the loop (loopSeq):
507 model_TCW_loop=n_mut_genpos+DNA_access+hairpin_stab+loopSeq. The loop sequence was a
508 binary variable to indicate whether the mutation occurred in the following sequence: 1001 or 101 (0 =
509 A/T, 1 = G/C; underlined is the position of the hotspot mutation).

510 Using these models, we estimated the expected number of mutations of the specified hotspot mutation
511 for which the model was built. Then, the exact Poisson test was applied to estimate the significance of
512 observing the same or more mutations than expected in a specific genomic position. P-values were
513 Benjamini-Hochberg adjusted.

514

515 **RNA-sequencing**

516 Alignment, pre-processing of RNA-seq data and transcript normalization have been previously
517 described [10, 11]. The transcriptomic subtype of each mUC sample was identified when the mean
518 (normalized) expression of all genes associated with a specific subtype was the highest across all
519 subtypes.

520

521 **mRNA editing**

522 Jalili, et al. 2020 identified hotspot mutations in the mRNA of *DDOST* that is targeted by APOBEC3A
523 [18]. The genomic position of this hotspot mutation reveals a hairpin-loop structure that is an ideal
524 substrate for APOBEC3A. Due to the short life-time of mRNA molecules, the presence of this hotspot
525 mutation reflects ongoing APOBEC mutagenesis. The proportion of C>U mutations in chr1:20981977
526 was estimated to identify the RNA-editing activity of APOEBC3A.

527

Transcriptome expression data mapped to genomic regions.

MultiBamSummary from deepTools v1.30.0 [56] was used to read BAM files and estimate the number of reads in genomic regions with a size of one Mbp. The average raw read count per Mbp was calculated, and a moving average with k = 3 bins was applied. The scale of the read counts was normalized per chromosome using the mean and standard deviation. High transcriptional regions were defined as such when the expression value of one region was above the median of the whole genome.

Statistical analysis

Analyses were performed using the statistical analysis platform R v4.1.0 [57]. Fisher's exact, Wilcoxon-rank sum and Wilcoxon signed-rank tests were used for comparison between groups. The correlation values of continuous values with categorical values were estimated with logistic regression analysis applying the Wald test. The exact Poisson test was applied to identify potential driver hotspot mutations. The Poisson-binomial method was applied for mutually exclusive mutation events using Rediscover [58] and the Fisher's exact test was applied for the significance of co-occurred mutations. In all cases, p-values were adjusted using the Benjamini-Hochberg method.

553 # Declarations

554 **Ethics approval and consent to participate**

555 The study protocols of patients analyzed in this publication have been approved by the corresponding
556 authorities at the time of access granted by HMF and the NCBI dbGAP.
557

558 **Consent for publication**
559 N/A

560

561 **Availability of data and materials**
562 WGS, RNA-seq and clinical data from mUC and mBC are available through the Hartwig Medical
563 Foundation at https://www.hartwigmedicalfoundation.nl, under request numbers DR-041 and DR-026,
564 respectively. For mUC, samples that were previously analyzed (DR-031) by Nakauma et. al, 2022 [10]
565 were retrieved from DR-041. WGS data from primary UC was requested to the NCBI dbGAP and granted
566 access through request #33427.

567 ChIPseq data experiments are freely available through The ENCODE Project Consortium [59] and the
568 Roadmap Epigenomics Consortium [60] on the ENCODE portal (https://www.encodeproject.org) [19].

569 The scripts, including the algorithm to find hairpin-loops and estimate the thermodynamic stability have
570 been deposited in a public repository available at https://github.com/ANakauma/ApobecHM_drivers.
571 Other scripts used for processing sequencing data can be found at https://github.com/J0bbie/R2CPCT
572 and https://github.com/hartwigmedical/hmftools.

573

574 **Competing interests**
575 Martijn P. J. Lolkema has received research support from JnJ, Sanofi, Astellas and MSD, and
576 consultancy fees from Incyte, Amgen, JnJ, Bayer, Servier, Roche, INCa, Pfizer, Sanofi, Astellas,
577 AstraZeneca, Merck Sharp & Dohme, Novartis, Julius Clinical and the Hartwig Medical Foundation (all
578 paid to the Erasmus MC Cancer Institute).

579 Joost L. Boormans has received research support from Merck AG / Pfizer, Janssen and Merck Sharp &
580 Dohme, and consultancy fees from Merck Sharp & Dohme, Bristol-Myers Squibb, Astellas, AstraZeneca,
581 Ipsen and Janssen (all paid to the Erasmus MC Cancer Institute)

John W. M. Martens has received research support from Pfizer, Sanofi, GSK, Therawis Cergentis, and Philips (all paid to the Erasmus MC Cancer Institute) and one consultancy fee from Novartis.

Astrid A.M. van der Veldt has received consultancy fees from BMS, MSD, Pfizer, Novartis, Eisai, Sanofi, Pierre Fabre, Ipsen and Roche (all paid to the Erasmus MC Cancer Institute).

J. Alberto Nakauma-González, Maud Rijnders, Minouk Noordsij and Harmen J. G. van de Werken declare no competing interests.

Funding

The Stichting Dutch Uro-Oncology Study (DUOS) group and the Daniel den Hoed Foundation supported this research.

Authors' contributions

Conceptualization: JAN and HJGvdW; Methodology: JAN and HJGvdW; Software: JAN, HJGvdW and MN; Validation: JAN and HJGvdW; Formal Analysis: JAN and HJGvdW; Investigation: All authors; Resources: JLB; Data Curation: JAN, MR and HJGvdW; Writing – Original Draft: JAN and HJGvdW; Writing – Review & Editing: All authors; Visualization: JAN; Supervision: JLB, HJGvdW and MPJL; Project Administration: JAN, HJGvdW, MPJL and JLB; Funding Acquisition: JLB , HJGvdW, and MPJL

Acknowledgements

Hartwig Medical Foundation and the Center of Personalized Cancer Treatment are acknowledged for making the clinical, genomic and transcriptomic data available to the study. We would also like to thank the Pan-Cancer Analysis of Whole Genomes study for sharing the genomic data from 23 primary UC patients. We are particularly grateful to all participating patients and their families.

607 ## References

608 1. Roberts SA, Lawrence MS, Klimczak LJ, Grimm SA, Fargo D, Stojanov P, Kiezun A, Kryukov GV, Carter
609 SL, Saksena G, et al: **An APOBEC cytidine deaminase mutagenesis pattern is widespread in human**
610 **cancers.** *Nature Genetics* 2013, **45:**970-976.
611 2. Tate JG, Bamford S, Jubb HC, Sondka Z, Beare DM, Bindal N, Boutselakis H, Cole CG, Creatore C,
612 Dawson E, et al: **COSMIC: The Catalogue Of Somatic Mutations In Cancer.** *Nucleic Acids Research*
613 2019, **47:**D941-D947.
614 3. Boichard A, Pham TV, Yeerna H, Goodman A, Tamayo P, Lippman S, Frampton GM, Tsigelny IF, Kurzrock
615 R: **APOBEC-related mutagenesis and neo-peptide hydrophobicity: implications for response to**
616 **immunotherapy.** *OncoImmunology* 2019, **8:**1550341.
617 4. Rijnders M, Nakauma-González JA, Robbrecht DGJ, Gil-Jimenez A, Aarts MJB, Boormans JL, Hamberg
618 P, van der Heijden MS, Szabados BE, van Leenders GJLH, et al: **T Cell-to-Stroma Enrichment (TSE)**
619 **score: a gene expression metric that predicts response to immune checkpoint inhibitors in patients**
620 **with urothelial cancer.** *bioRxiv* 2022**:**2022.2005.2030.493997.
621 5. Law EK, Levin-Klein R, Jarvis MC, Kim H, Argyris PP, Carpenter MA, Starrett GJ, Temiz NA, Larson LK,
622 Durfee C, et al: **APOBEC3A catalyzes mutation and drives carcinogenesis in vivo.** *Journal of*
623 *Experimental Medicine* 2020, **217**.
624 6. Sharma S, Baysal BE: **Stem-loop structure preference for site-specific RNA editing by APOBEC3A**
625 **and APOBEC3G.** *PeerJ* 2017, **5:**e4136.
626 7. Kumar S, Warrell J, Li S, McGillivray PD, Meyerson W, Salichos L, Harmanci A, Martinez-Fundichely A,
627 Chan CWY, Nielsen MM, et al: **Passenger Mutations in More Than 2,500 Cancer Genomes: Overall**
628 **Molecular Functional Impact and Consequences.** *Cell* 2020, **180:**915-927.e916.
629 8. Shi MJ, Meng XY, Fontugne J, Chen CL, Radvanyi F, Bernard-Pierrot I: **Identification of new driver and**
630 **passenger mutations within APOBEC-induced hotspot mutations in bladder cancer.** *Genome*
631 *Medicine* 2020, **12:**85-85.
632 9. Wong JKL, Aichmüller C, Schulze M, Hlevnjak M, Elgaafary S, Lichter P, Zapatka M: **Association of**
633 **mutation signature effectuating processes with mutation hotspots in driver genes and non-coding**
634 **regions.** *Nature Communications* 2022, **13:**178.
635 10. Nakauma-González JA, Rijnders M, van Riet J, van der Heijden MS, Voortman J, Cuppen E, Mehra N,
636 van Wilpe S, Oosting SF, Rijstenberg LL, et al: **Comprehensive Molecular Characterization Reveals**
637 **Genomic and Transcriptomic Subtypes of Metastatic Urothelial Carcinoma.** *European Urology* 2022,
638 **81:**331-336.
639 11. Angus L, Smid M, Wilting SM, van Riet J, Van Hoeck A, Nguyen L, Nik-Zainal S, Steenbruggen TG, Tjan-
640 Heijnen VCG, Labots M, et al: **The genomic landscape of metastatic breast cancer highlights**
641 **changes in mutation and signature frequencies.** *Nature Genetics* 2019, **51:**1450-1458.
642 12. Chan K, Roberts SA, Klimczak LJ, Sterling JF, Saini N, Malc EP, Kim J, Kwiatkowski DJ, Fargo DC,
643 Mieczkowski PA, et al: **An APOBEC3A hypermutation signature is distinguishable from the signature**
644 **of background mutagenesis by APOBEC3B in human cancers.** *Nature Genetics* 2015, **47:**1067-1072.
645 13. Roelofs PA, Timmermans MAM, Stefanovska B, den Boestert MA, van den Borne AWM, Balcioglu HE,
646 Trapman AM, Harris RS, Martens JWM, Span PN: **Aberrant APOBEC3B Expression in Breast Cancer**
647 **Is Linked to Proliferation and Cell Cycle Phase.** *Cells* 2023, **12:**1185.
648 14. Hirabayashi S, Shirakawa K, Horisawa Y, Matsumoto T, Matsui H, Yamazaki H, Sarca AD, Kazuma Y,
649 Nomura R, Konishi Y, et al: **APOBEC3B is preferentially expressed at the G2/M phase of cell cycle.**
650 *Biochemical and Biophysical Research Communications* 2021, **546:**178-184.
651 15. Cortez LM, Brown AL, Dennis MA, Collins CD, Brown AJ, Mitchell D, Mertz TM, Roberts SA: **APOBEC3A**
652 **is a prominent cytidine deaminase in breast cancer.** *PLoS Genetics* 2019, **15:**e1008545-e1008545.
653 16. Glaser AP, Fantini D, Wang Y, Yu Y, Rimar KJ, Podojil JR, Miller SD, Meeks JJ: **APOBEC-mediated**
654 **mutagenesis in urothelial carcinoma is associated with improved survival, mutations in DNA**
655 **damage response genes, and immune response.** *Oncotarget* 2018, **9:**4537-4548.
656 17. Robertson AG, Kim J, Al-Ahmadie H, Bellmunt J, Guo G, Cherniack AD, Hinoue T, Laird PW, Hoadley KA,
657 Akbani R, et al: **Comprehensive Molecular Characterization of Muscle-Invasive Bladder Cancer.** *Cell*
658 2017, **171:**540-556.e525.
659 18. Jalili P, Bowen D, Langenbucher A, Park S, Aguirre K, Corcoran RB, Fleischman AG, Lawrence MS, Zou
660 L, Buisson R: **Quantification of ongoing APOBEC3A activity in tumor cells by monitoring RNA**
661 **editing at hotspots.** *Nature Communications* 2020, **11:**2971.
662 19. Davis CA, Hitz BC, Sloan CA, Chan ET, Davidson JM, Gabdank I, Hilton JA, Jain K, Baymuradov UK,
663 Narayanan AK, et al: **The Encyclopedia of DNA elements (ENCODE): data portal update.** *Nucleic Acids*
664 *Research* 2018, **46:**D794-D801.
665 20. Allory Y, Beukers W, Sagrera A, Flández M, Marqués M, Márquez M, Van Der Keur KA, Dyrskjot L, Lurkin
666 I, Vermeij M, et al: **Telomerase reverse transcriptase promoter mutations in bladder cancer: High**
667 **frequency across stages, detection in urine, and lack of association with outcome.** *European*
668 *Urology* 2014, **65:**360-366.
669 21. Siraj AK, Bu R, Iqbal K, Parvathareddy SK, Siraj N, Siraj S, Diaz MRF, Rala DR, Benito AD, Sabido MA,
670 et al: **Telomerase reverse transcriptase promoter mutations in cancers derived from multiple organ**
671 **sites among middle eastern population.** *Genomics* 2020, **112:**1746-1753.

22. Buisson R, Langenbucher A, Bowen D, Kwan EE, Benes CH, Zou L, Lawrence MS: **Passenger hotspot mutations in cancer driven by APOBEC3A and mesoscale genomic features.** *Science (New York, NY)* 2019, **364:**eaaw2872-eaaw2872.
23. Langenbucher A, Bowen D, Sakhtemani R, Bournique E, Wise JF, Zou L, Bhagwat AS, Buisson R, Lawrence MS: **An extended APOBEC3A mutation signature in cancer.** *Nature Communications* 2021, **12:**1602.
24. Wu S, Ou T, Xing N, Lu J, Wan S, Wang C, Zhang X, Yang F, Huang Y, Cai Z: **Whole-genome sequencing identifies ADGRG6 enhancer mutations and FRS2 duplications as angiogenesis-related drivers in bladder cancer.** *Nature Communications* 2019, **10:**1-12.
25. An L, Dong C, Li J, Chen J, Yuan J, Huang J, Chan KM, Yu C-h, Huen MSY: **RNF169 limits 53BP1 deposition at DSBs to stimulate single-strand annealing repair.** *Proceedings of the National Academy of Sciences* 2018, **115:**E8286-E8295.
26. Cheng Y-C, Chiang H-Y, Cheng S-J, Chang H-W, Li Y-J, Shieh S-Y: **Loss of the tumor suppressor BTG3 drives a pro-angiogenic tumor microenvironment through HIF-1 activation.** *Cell Death & Disease* 2020, **11:**1046.
27. Fischer J-P, Els-Heindl S, Beck-Sickinger AG: **Adrenomedullin – Current perspective on a peptide hormone with significant therapeutic potential.** *Peptides* 2020, **131:**170347.
28. Levine AJ, Brivanlou AH: **GDF3 at the crossroads of TGF-beta signaling.** *Cell Cycle* 2006, **5:**1069-1073.
29. Lo Y-H, Romes EM, Pillon MC, Sobhany M, Stanley RE: **Structural Analysis Reveals Features of Ribosome Assembly Factor Nsa1/WDR74 Important for Localization and Interaction with Rix7/NVL2.** *Structure* 2017, **25:**762-772.e764.
30. Campbell PJ, Getz G, Korbel JO, Stuart JM, Jennings JL, Stein LD, Perry MD, Nahal-Bose HK, Ouellette BFF, Li CH, et al: **Pan-cancer analysis of whole genomes.** *Nature* 2020, **578:**82-93.
31. Hoopes James II, Cortez Luis MM, Mertz Tony MM, Malc Ewa PP, Mieczkowski Piotr AA, Roberts Steven AA: **APOBEC3A and APOBEC3B Preferentially Deaminate the Lagging Strand Template during DNA Replication.** *Cell Reports* 2016, **14:**1273-1282.
32. Kazanov MD, Roberts SA, Polak P, Stamatoyannopoulos J, Klimczak LJ, Gordenin DA, Sunyaev SR: **APOBEC-Induced Cancer Mutations Are Uniquely Enriched in Early-Replicating, Gene-Dense, and Active Chromatin Regions.** *Cell Reports* 2015, **13:**1103-1109.
33. Petljak M, Dananberg A, Chu K, Bergstrom EN, Striepen J, von Morgen P, Chen Y, Shah H, Sale JE, Alexandrov LB, et al: **Mechanisms of APOBEC3 mutagenesis in human cancer cells.** *Nature* 2022, **607:**799-807.
34. Petljak M, Alexandrov LB, Brammeld JS, Price S, Wedge DC, Grossmann S, Dawson KJ, Ju YS, Iorio F, Tubio JMC, et al: **Characterizing Mutational Signatures in Human Cancer Cell Lines Reveals Episodic APOBEC Mutagenesis.** *Cell* 2019, **176:**1282-1294.e1220.
35. Alexandrov LB, Kim J, Haradhvala NJ, Huang MN, Tian Ng AW, Wu Y, Boot A, Covington KR, Gordenin DA, Bergstrom EN, et al: **The repertoire of mutational signatures in human cancer.** *Nature* 2020, **578:**94-101.
36. Helleday T, Eshtad S, Nik-Zainal S: **Mechanisms underlying mutational signatures in human cancers.** *Nature Reviews Genetics* 2014, **15:**585-598.
37. Petljak M, Maciejowski J: **Molecular origins of APOBEC-associated mutations in cancer.** *DNA Repair* 2020, **94:**102905.
38. Patel M, Nowsheen S, Maraboyina S, Xia F: **The role of poly(ADP-ribose) polymerase inhibitors in the treatment of cancer and methods to overcome resistance: a review.** *Cell & Bioscience* 2020, **10:**35.
39. Jekimovs C, Bolderson E, Suraweera A, Adams M, O'Byrne KJ, Richard DJ: **Chemotherapeutic Compounds Targeting the DNA Double-Strand Break Repair Pathways: The Good, the Bad, and the Promising.** *Frontiers in Oncology* 2014, **4**.
40. Gillyard T, Davis J: **Chapter Two - DNA double-strand break repair in cancer: A path to achieving precision medicine.** In *International Review of Cell and Molecular Biology. Volume* 364. Edited by Weyemi U, Galluzzi L: Academic Press; 2021: 111-137
41. Marabelle A, Fakih M, Lopez J, Shah M, Shapira-Frommer R, Nakagawa K, Chung HC, Kindler HL, Lopez-Martin JA, Miller WH, Jr., et al: **Association of tumour mutational burden with outcomes in patients with advanced solid tumours treated with pembrolizumab: prospective biomarker analysis of the multicohort, open-label, phase 2 KEYNOTE-158 study.** *Lancet Oncol* 2020, **21:**1353-1365.
42. Lawson ARJ, Abascal F, Coorens THH, Hooks Y, O'Neill L, Latimer C, Raine K, Sanders MA, Warren AY, Mahbubani KTA, et al: **Extensive heterogeneity in somatic mutation and selection in the human bladder.** *Science (New York, NY)* 2020, **370:**75-82.
43. Wang Y, Robinson PS, Coorens THH, Moore L, Lee-Six H, Noorani A, Sanders MA, Jung H, Katainen R, Heuschkel R, et al: **APOBEC mutagenesis is a common process in normal human small intestine.** *Nature Genetics* 2023, **55:**246-254.
44. Priestley P, Baber J, Lolkema MP, Steeghs N, de Bruijn E, Shale C, Duyvesteyn K, Haidari S, van Hoeck A, Onstenk W, et al: **Pan-cancer whole-genome analyses of metastatic solid tumours.** *Nature* 2019, **575:**210-216.
45. van Dessel LF, van Riet J, Smits M, Zhu Y, Hamberg P, van der Heijden MS, Bergman AM, van Oort IM, de Wit R, Voest EE, et al: **The genomic landscape of metastatic castration-resistant prostate cancers**

reveals multiple distinct genotypes with potential clinical impact. *Nature Communications* 2019, **10:**1-13.

46. Hazelaar DM, van Riet J, Hoogstrate Y, van de Werken HJG: **Katdetectr: An R/Bioconductor package utilizing unsupervised changepoint analysis for robust kataegis detection.** *bioRxiv* 2023**:**2022.2007.2011.499364.
47. Stephens PJ, Tarpey PS, Davies H, Van Loo P, Greenman C, Wedge DC, Nik-Zainal S, Martin S, Varela I, Bignell GR, et al: **The landscape of cancer genes and mutational processes in breast cancer.** *Nature* 2012, **486:**400-404.
48. Bolli N, Avet-Loiseau H, Wedge DC, Van Loo P, Alexandrov LB, Martincorena I, Dawson KJ, Iorio F, Nik-Zainal S, Bignell GR, et al: **Heterogeneity of genomic evolution and mutational profiles in multiple myeloma.** *Nature Communications* 2014, **5:**2997-2997.
49. Gundem G, Van Loo P, Kremeyer B, Alexandrov LB, Tubio JMC, Papaemmanuil E, Brewer DS, Kallio HML, Högnäs G, Annala M, et al: **The evolutionary history of lethal metastatic prostate cancer.** *Nature* 2015, **520:**353-357.
50. SantaLucia J: **A unified view of polymer, dumbbell, and oligonucleotide DNA nearest-neighbor thermodynamics.** *Proceedings of the National Academy of Sciences* 1998, **95:**1460-1465.
51. John SantaLucia J, Hicks D: **The Thermodynamics of DNA Structural Motifs.** *Annual Review of Biophysics and Biomolecular Structure* 2004, **33:**415-440.
52. Zuker M: **Mfold web server for nucleic acid folding and hybridization prediction.** *Nucleic Acids Research* 2003, **31:**3406-3415.
53. Martincorena I, Raine KM, Gerstung M, Dawson KJ, Haase K, Van Loo P, Davies H, Stratton MR, Campbell PJ: **Universal Patterns of Selection in Cancer and Somatic Tissues.** *Cell* 2017, **171:**1029-1041.e1021.
54. Polak P, Karlić R, Koren A, Thurman R, Sandstrom R, Lawrence MS, Reynolds A, Rynes E, Vlahoviček K, Stamatoyannopoulos JA, Sunyaev SR: **Cell-of-origin chromatin organization shapes the mutational landscape of cancer.** *Nature* 2015, **518:**360-364.
55. Kübler K, Karlić R, Haradhvala NJ, Ha K, Kim J, Kuzman M, Jiao W, Gakkhar S, Mouw KW, Braunstein LZ, et al: **Tumor mutational landscape is a record of the pre-malignant state.** *bioRxiv* 2019**:**517565.
56. Ramírez F, Ryan DP, Grüning B, Bhardwaj V, Kilpert F, Richter AS, Heyne S, Dündar F, Manke T: **deepTools2: a next generation web server for deep-sequencing data analysis.** *Nucleic Acids Research* 2016, **44:**W160-W165.
57. Team RC: **R Core Team (2017). R: A language and environment for statistical computing.** *R Foundation for Statistical Computing, Vienna, Austria URL http://wwwR-projectorg/* 2017**:**R Foundation for Statistical Computing-R Foundation for Statistical Computing.
58. Ferrer-Bonsoms JA, Jareno L, Rubio A: **Rediscover: an R package to identify mutually exclusive mutations.** *Bioinformatics* 2021, **38:**844-845.
59. Dunham I, Kundaje A, Aldred SF, Collins PJ, Davis CA, Doyle F, Epstein CB, Frietze S, Harrow J, Kaul R, et al: **An integrated encyclopedia of DNA elements in the human genome.** *Nature* 2012, **489:**57-74.
60. Roadmap Epigenomics C, Kundaje A, Meuleman W, Ernst J, Bilenky M, Yen A, Heravi-Moussavi A, Kheradpour P, Zhang Z, Wang J, et al: **Integrative analysis of 111 reference human epigenomes.** *Nature* 2015, **518:**317-329.

Supplementary Information

Additional file 1: Supplementary figures S1-10

Additional file 2: Supplementary tables S1-6