Mutational impact of APOBEC3B and APOBEC3A in a human cell line

Running title: Mutational signatures of APOBEC3B and APOBEC3A

Matthew C. Jarvis[1,2], Michael A. Carpenter[1,2,3], Nuri A. Temiz[2,4], Margaret R. Brown[1,2], Kyle A. Richards[1,2,5], Prokopios P. Argyris[1,2,3], William L. Brown[1,2], Douglas Yee[2,6], Reuben S. Harris[1,2,3,5,7,8]

[1] Department of Biochemistry, Molecular Biology and Biophysics, Masonic Cancer Center, University of Minnesota, Minneapolis, MN 55455, USA

[2] Masonic Cancer Center, University of Minnesota, Minneapolis, MN 55455, USA

[3] Howard Hughes Medical Institute, University of Minnesota, Minneapolis, MN 55455, USA

[4] Institute for Health Informatics, University of Minnesota, Minneapolis, MN 55455, USA

[5] Department of Biochemistry and Structural Biology, University of Texas Health San Antonio, San Antonio, TX 78229, USA

[6] Department of Medicine, University of Minnesota Medical School, Minneapolis, Minnesota 55455, USA

[7] Howard Hughes Medical Institute, University of Texas Health San Antonio, San Antonio, TX 78229, USA

[8] Correspondence to Reuben S. Harris, email: rsh@uthscsa.edu

Key Words: APOBEC mutation signature; APOBEC3A; APOBEC3B; cancer mutagenesis; DNA damage response; DNA repair; human mutation reporter system; mutation signatures in cancer

Manuscript Information: 64 character title; 194 word abstract; 5900 word main text (excluding methods, references, and legends); 5 Main Figures and 8 Supplementary Figures

Abstract

A prominent source of mutation in cancer is single-stranded DNA cytosine deamination by cellular APOBEC3 enzymes, which results in C-to-T and C-to-G mutations in TCA and TCT motifs. Although multiple enzymes have been implicated, reports conflict and it is unclear which enzyme(s) are responsible. Here we develop a selectable system to quantify genome mutation and compare the mutagenic activities of three leading candidates - APOBEC3A, APOBEC3B, and APOBEC3H. The human cell line, HAP1, was engineered to express the *thymidine kinase* (*TK*) gene of HSV-1, which confers sensitivity to ganciclovir. Clonal expression of APOBEC3A and APOBEC3B, but not catalytic mutant controls or APOBEC3H, triggered elevated DNA damage responses and increased frequencies of *TK* mutation. Mutant *TK* DNA sequences revealed nearly indistinguishable cytosine mutation patterns. Whole genome sequences from *TK* mutant clones confirmed these results and enabled broader bioinformatic analyses. Most importantly, comparisons of "pure" APOBEC3A- and APOBEC3B-inflicted mutation signatures from this system and the actual APOBEC3 signature from breast cancer indicated that most tumors manifest a composite signature. These studies help resolve a long-standing etiologic debate in the cancer field and indicate that future diagnostic and therapeutic efforts should focus on both APOBEC3A and APOBEC3B.

Introduction

Over the past decade, advances in DNA sequencing technologies and bioinformatics have helped to deconvolute a multitude of mutational processes that contribute to the genesis and evolution of cancer [see recent pan-cancer analysis by (Alexandrov et al., 2020) and reviews by (Koh, Degasperi, Zou, Momen, & Nik-Zainal, 2021; M. Petljak & Maciejowski, 2020; Venkatesan et al., 2018)]. Through these approaches and complementary bench experiments, the APOBEC3 family of single-stranded (ss)DNA cytosine deaminases has emerged as one of the top three sources of single base substitution (SBS) mutation in cancer with particularly large contributions

53 to tumors of the bladder, breast, cervix, lung, and head/neck. APOBEC3 signature mutations in
54 cancer are defined as C-to-T transitions and C-to-G transversions in 5'-TCW motifs (W = A or T;
55 SBS2 and SBS13, respectively) (Alexandrov et al., 2020; Alexandrov, Nik-Zainal, Wedge,
56 Aparicio, et al., 2013; Burns, Lackey, et al., 2013; Leonard et al., 2013; Nik-Zainal et al., 2012;
57 Nik-Zainal et al., 2016; Roberts et al., 2013). This definition is conservative because several
58 APOBEC3 enzymes can also accommodate 5'-NCG and 5'Nmethyl-CG ssDNA substrates
59 (Burns, Lackey, et al., 2013; Carpenter et al., 2012; Hultquist et al., 2011; Ito, Fu, Kao, Yang, &
60 Chen, 2017; Ito et al., 2018; Leonard et al., 2013; Schutsky, Nabel, Davis, DeNizio, & Kohli, 2017;
61 Shi et al., 2017; Suspène, Aynaud, Vartanian, & Wain-Hobson, 2013), which can also lead to C-
62 to-T transition mutations and overlap with the mutation signature attributable to spontaneous
63 water-mediated deamination of cytosine and methyl-cytosine nucleotides. Spontaneous methyl-
64 cytosine deamination is a clock-like mutation process that occurs predominantly in 5'Nmethyl-CG
65 motifs and associates positively with a patient's biological age (Alexandrov et al., 2015) (ageing
66 signature), whereas APOBEC-catalyzed deamination is not evident in normal tissues, not
67 associated with ageing, present in many primary tumors, and often enriched in metastases
68 (Alexandrov et al., 2015; Alexandrov et al., 2020; Angus et al., 2019; Burns, Lackey, et al., 2013;
69 Burns, Temiz, & Harris, 2013; Moore et al., 2021; Nik-Zainal et al., 2016; Roberts et al., 2013;
70 Sieuwerts et al., 2017; Sieuwerts et al., 2014).

71 The human APOBEC3 (A3) family is comprised of seven different enzymes with extensive
72 homology and overlapping activities [reviewed by (Harris & Dudley, 2015; Silvas & Schiffer, 2019;
73 Simon, Bloch, & Landau, 2015)], and it is unclear how much (or little) each contributes to the
74 composite APOBEC3 mutation signature evident in tumor DNA sequences. The bulk of evidence
75 favors two enzymes, APOBEC3A (A3A) and APOBEC3B (A3B), though to wildly different degrees
76 depending on the study, and additional work has also implicated APOBEC3H (A3H). Evidence
77 for A3A includes an intrinsic preference for 5'-TC substrates, high catalytic activity (highest of any
78 human DNA deaminase), cell-wide localization, DNA damage responses, induction by the tumor

79 virus HPV, and positive correlations between mRNA levels and tumor APOBEC mutation loads
80 (Carpenter et al., 2012; H. Chen et al., 2006; Cortez et al., 2019; Faden et al., 2017; Green et al.,
81 2016; Ito et al., 2017; Kouno et al., 2017; Shi et al., 2017; Taylor et al., 2013; Warren et al., 2015;
82 Westrich et al., 2018). Evidence for A3B includes all the same points except this enzyme is
83 several-fold less active, localizes constitutively to the nuclear compartment, is induced by multiple
84 DNA tumor viruses (HPV and polyomaviruses), and associates at both mRNA and protein levels
85 with clinical outcomes (Argyris et al., 2020; Burns, Lackey, et al., 2013; Burns, Temiz, et al., 2013;
86 de Bruin et al., 2014; Ito et al., 2017; Roelofs et al., 2020; Salamango et al., 2018; Shi et al., 2017;
87 Sieuwerts et al., 2017; Sieuwerts et al., 2014; Starrett et al., 2019; Taylor et al., 2013; Verhalen,
88 Starrett, Harris, & Jiang, 2016; Vieira et al., 2014; Walker et al., 2015; Warren et al., 2015;
89 Yamazaki et al., 2019; Yamazaki et al., 2020). Importantly, however, the APOBEC3 mutation
90 signature can still accumulate in tumors completely lacking *A3B* due to a naturally occurring full
91 gene deletion allele (T. W. Chen et al., 2017; Nik-Zainal et al., 2014; Starrett et al., 2016). This
92 observation with *A3B*-null tumors helped implicate a cell-wide, 5'-TC preferring variant of A3H
93 (haplotype I) in causing an APOBEC mutation signature (Feng et al., 2015; Hix, Wong, Flath,
94 Chelico, & Cisneros, 2020; Starrett et al., 2016). Another consideration is that A3A and A3B have
95 been reported to elicit broader tetranucleotide preferences upon overexpression in yeast, 5'-
96 YTCW vs RTCW, respectively (Buisson et al., 2019; Chan et al., 2015; Hoopes et al., 2016).
97 Here, we report a human cellular system for mutation research and use it to perform
98 unbiased comparisons of the mutagenic potential of A3A, A3B, and A3H. The human cell line
99 HAP1 was engineered to express a single copy of the HSV-1 *thymidine kinase* (*TK*) gene, which
100 enables the drug ganciclovir to be used to select rare *TK* mutants and quantify mutation
101 frequencies. Moreover, the *TK* gene can be amplified readily from ganciclovir-resistant (Gan^R)
102 clones by high-fidelity PCR and sequenced to provide initial assessments of mutation spectra
103 (signatures) prior to undertaking additional experiments such as whole genome sequencing
104 (WGS). Using this system, only expression of A3A and A3B caused significant increases in DNA

105 damage signaling (γ-H2AX positivity), DNA breakage (alkaline COMETs), and Gan^R mutation
106 frequencies. Sanger sequences from panels of individual *TK* mutant clones revealed a clear 5'-
107 TC-biased mutation pattern including two hotspots, which had no obvious stem-loop secondary
108 structures. A global perspective was obtained by analyzing the full genomic DNA sequences of
109 *TK* mutant clones. These WGS results demonstrated that both A3A and A3B can generate the
110 APOBEC3 mutational signatures SBS2 and SBS13. Moreover, although these "pure" A3A- and
111 A3B-inflicted mutation signatures each cluster with the APOBEC3 signature extracted from
112 different subsets of breast tumors, neither the A3A- nor the A3B-preferred motifs from the HAP1-
113 TK system are able to fully explain the observed composite APOBEC3 mutation signature in
114 cancer suggesting that both enzymes may be acting in concert in most tumors.
115
116 **Results**
117 **HAP1-TK-M9 – a human cellular system to report DNA damage and mutagenesis**
118 Model organisms such as *E. coli* and yeast are powerful systems for studying mutagens
119 including DNA deaminases [*e.g.*, original studies with A3 enzymes (Chan et al., 2012; Harris,
120 Petersen-Mahrt, & Neuberger, 2002; Mayorov et al., 2005; Petersen-Mahrt, Harris, & Neuberger,
121 2002; Schumacher, Nissley, & Harris, 2005)]. However, these model organisms do not
122 recapitulate all of the DNA repair and regulatory mechanisms found in human cells. We therefore
123 sought to combine strengths of both approaches by introducing a single copy of a selectable
124 reporter, the HSV-1 *thymidine kinase* (*TK*) gene, into the genome of the human cell line HAP1.
125 TK confers exquisite sensitivity to the drug ganciclovir and, like many antimicrobial agents, only
126 *TK*-mutant, ganciclovir-resistant (Gan^R) cells are able to survive selection by this drug. The vast
127 majority of Gan^R mutants can be characterized rapidly by conventional Sanger DNA sequencing
128 because *TK* is a single open reading frame. Moreover, once informative Gan^R mutants are
129 revealed by selection, secondary analyses including WGS can be used to uncover additional and
130 potentially global features of a given mutation process.

131 The overall experimental workflow is shown in **Figure 1A**. To generate a genetically clean
132 "mother" clone of the commercially available HAP1 cell line, Sleeping Beauty (SB)-mediated
133 transposition was used to introduce a single copy of a *TK-Neo* cassette into the genome (Burns,
134 Lackey, et al., 2013). Neo^R clones were selected with G418, expanded into healthy clonal
135 populations (ca. 10^6 cells/ml), and screened for ganciclovir sensitivity (Gan^S). One mother clone,
136 HAP1-TK-M9, was selected for further studies because it is Gan^S (apart from rare resistant
137 mutants) and mostly diploid (apart from some visible chromosome aberrations and S/G2/M cells),
138 it cultures, engineers, and clones well (below), and it shows a favorable *A3* expression profile
139 (**Figure 1B**; **Figure 1 – Figure Supplement 1**). In particular, RT-qPCR showed that its *A3A* and
140 *A3B* mRNA levels are lower than those of the original parent line and that *A3H* mRNA levels are
141 very low and near the detection threshold. Accordingly, virtually no ssDNA cytosine deaminase
142 activity is detected in extracts (vector control lanes in **Figure 1C-E**). In addition, genomic DNA
143 sequencing showed that the only *A3H* allele is haplotype III (ΔAsn15), which is known to produce
144 an unstable protein (Chesarino & Emerman, 2020; Ebrahimi et al., 2018; Refsland et al., 2014).
145

146 **DNA damage phenotypes of HAP1-TK-M9 overexpressing A3A, A3B, and A3H**

147 A panel of MLV-based human A3 expression constructs was assembled and used to
148 produce viral supernatants and transduce the HAP1-TK-M9 mother clone (**Methods**). Uniform
149 A3-expressing pools were selected with puromycin. A3A expression was demonstrated by
150 immunoblotting and immunofluorescent (IF) microscopy using a newly developed rabbit anti-
151 human A3A monoclonal antibody, UMN13, that specifically binds this enzyme through a unique
152 N-terminal epitope (**Figure 1C**, **1F**; **Figure 1 – Figure Supplement 2**). Expression of A3B was
153 similarly shown by immunoblotting and IF microscopy using a rabbit anti-human A3B monoclonal
154 antibody 5210-87-13 (Brown et al., 2019) (**Figure 1D**, **1F**). Expression of A3H haplotypes I and II
155 was shown by immunoblotting and IF microscopy with a rabbit polyclonal antibody (Refsland et
156 al., 2014; Shaban et al., 2018) (**Figure 1E**, **1F**). Endogenous A3A, A3B, and A3H proteins were

157 undetectable as expected based on low mRNA levels (above). In addition, whole cell lysates of
158 A3A-, A3B-, and A3H-expressing HAP1-TK-M9 cells elicited ssDNA cytosine deaminase activity,
159 well above background levels from vector control or catalytic mutant pools constructed in parallel
160 (**Figure 1C-E**).

161 We next asked whether these A3 enzymes trigger a DNA damage response and cause
162 DNA breakage in HAP1-TK-M9 cells. The populations described above were stained for the DNA
163 damage marker γ-H2AX and subjected to IF microscopy (representative images in **Figure 1F**,
164 quantification in **Figure 1G**). Only A3A and A3B triggered elevated levels of DNA damage as
165 characterized by γ-H2AX staining. Catalytic mutant derivatives of A3A and A3B as well as active
166 A3H-I and A3H-II showed vector control levels of γ-H2AX staining. A3A and A3B also elicited DNA
167 breaks as evidenced by larger tail moments in alkaline COMET assays (representative images in
168 **Figure 1F**, quantification in **Figure 1H**). These results combined to demonstrate that both A3A
169 and A3B are enzymatically active and capable of exerting DNA damage in HAP1-TK-M9 cells,
170 consistent with prior DNA damage studies using different cell lines (Akre et al., 2016; Burns,
171 Lackey, et al., 2013; Landry, Narvaiza, Linfesty, & Weitzman, 2011; Mussil et al., 2013; Nikkila et
172 al., 2017; Yamazaki et al., 2020). Despite showing cell-wide (including nuclear) localization and
173 measurable DNA deaminase activity, A3H-I did not trigger significant increases in DNA damage
174 or breakage. This result differs from an earlier report (Hix et al., 2020), which may have expressed
175 A3H-I to higher levels with an inducible promoter in comparison to the stable levels established
176 here with an MND promoter from a single-copy, integrated MLV-based construct.

177

178 ***TK* mutation spectra of HAP1-TK-M9 with A3A, A3B, and A3H**

179 To directly test which A3 enzymes cause genomic mutation in the HAP1-TK-M9 system,
180 24 independent single-cell derived daughter clones were obtained for A3A, A3B, A3H-I, and A3H-
181 II expressing conditions as well as catalytic mutant controls. A classical fluctuation analysis was
182 performed by growing each single cell clone to $>10^7$ cells, subjecting each population to selection

183 by ganciclovir, and allowing time for single Gan^R mutant cells to grow into countable colonies.
184 Vector control conditions yielded a median Gan^R mutation frequency of 3 mutants per 5 million
185 cells (mean = 3.6, SD = 2.8). In contrast, A3A- and A3B-expressing clones caused median Gan^R
186 mutation frequencies to rise above 10 mutants per 5 million cells (A3A WT: median = 16, mean =
187 14, SEM = 2.2, p-value range = 1.7×10^{-4} to 3.1×10^{-5}; A3B WT: median = 11, mean = 12, SEM
188 = 1.8, p-value range = 2.2×10^{-4} to 1.9×10^{-5} by Welsh's t-test; **Figure 2A**). In comparison,
189 expression of A3H-I or A3H-II or catalytic mutant derivatives of any of these DNA enzymes failed
190 to trigger increased Gan^R mutation frequencies (A3H-I WT: median = 3.0, mean = 3.6, SEM =
191 0.71, p-value range = 0.38 to 0.60; A3H-II WT: median = 2.0, mean = 2.8, SEM = 0.58, p-value
192 range = 0.74 to 0.89 by Welsh's t-test; **Figure 2A**).

193 We next asked what types of genetic alterations led to inactivation of the *TK* gene in Gan^R
194 granddaughter clones derived from the different A3-expressing and control conditions. The *TK*
195 gene was PCR-amplified from genomic DNA of Gan^R granddaughter clones and sequenced via
196 Sanger sequencing. C-to-T and C-to-G mutations in an APOBEC-signature trinucleotide motif
197 (APOBEC), all other single base substitution mutations (other SBS), and all insertion/deletion
198 mutations (INDEL) were placed into groups for comparison (red, black, and blue tics in **Figure**
199 **2B**, respectively; individual sequence schematics in **Figure 2 – Figure Supplement 1**). *TK*
200 sequences derived from A3A-expressing granddaughter clones harbored a greater number of
201 APOBEC mutations [12/20 clones contained at least 1 APOBEC mutation, 22 T(C>T/G)W
202 mutations total, range of 0 to 3 SBS per sequence] relative to catalytic mutant control clones [3/19
203 clones contained 1 APOBEC mutation, 3 T(C>T/G)W mutations total, range of 0 to 1 SBS per
204 sequence]. Similarly, *TK* sequences derived from A3B-expressing granddaughter clones also
205 harbored a greater number of APOBEC mutations [11/20 clones contained at least 1 APOBEC
206 mutation, 19 T(C>T/G)W mutations total, range of 0 to 3 SBS per sequence] relative to catalytic
207 mutant control clones [2/18 clones contained 1 APOBEC mutation, 2 T(C>T/G)W mutations total,
208 range of 0 to 1 SBS per sequence] (**Figure 2B; Figure 2 – Figure Supplement 1**). Additionally,

209 *TK* cytosine bases 22 (Q8X) and 635 (R212K) emerged as mutational hotspots in both the A3A-
210 and A3B-expressing clones (**Figure 2B**; considered further in **Discussion**). No differences were
211 found in the number of other SBS or INDEL mutations between A3A- or A3B-expressing clones
212 and controls.

213 We next examined the sequence context of the 22 A3A- and 19 A3B-induced APOBEC
214 signature mutations that occurred at 5'-TC dinucleotides in *TK* (**Figure 2C-D**). In both instances,
215 A was preferred over T at the +1 nucleobase position relative to the mutated C, although this bias
216 was not significantly different between the two enzymes (68% for A3A and 74% for A3B; p=0.367
217 by Fisher's exact test). Similarly, no obvious biases were evident at the +2 or -2 nucleobase
218 positions with all four nucleotides observed at similar frequencies for both enzymes. Moreover,
219 even when pyrimidines and purines were grouped for comparison, A3A did not show an overt
220 preference for C/T (Y) or A/G (R) at the -2 or +2 nucleobase positions (51% vs 49% and 52% vs
221 48%, respectively). Likewise, A3B also failed to show an overt preference for C/T (Y) or A/G (R)
222 at the -2 or +2 nucleobase positions (45% vs 55% and 53% vs 47%, respectively). These
223 similarities underscore the fact that small mutation numbers are primarily useful for delineating
224 major differences such as the shifts described above from a heterogeneous pattern in catalytic
225 mutant- or vector control-expressing cells towards a predominantly 5'TC-focused SBS mutation
226 pattern in A3A- and A3B-expressing cells.

227

228 **WGS of A3A and A3B clones reveals genome-wide SBS mutation signatures including**
229 **biases unique to each enzyme**

230 To prepare for WGS of Gan^R granddaughter clones, RNA sequencing was done to confirm
231 expression of each exogenously expressed *A3* construct and compare mRNA levels relative to
232 established breast cell lines and primary breast tumors. All expression values were determined
233 relative to those of the conserved housekeeping gene *TBP* in order to be able to compare with
234 RNAseq data from different cell lines and tumors. Interestingly, the average *A3A* and *A3A-E72A*

235 mRNA levels in Gan^R granddaughter clones was over 5-fold higher than the average endogenous
236 *A3A* expression levels of APOBEC3-signature enriched breast cancer cell lines BT474 and MDA-
237 MD-453, breast cancer cell lines of the CCLE, or breast tumors of TCGA (**Figure 3 – Figure**
238 **Supplement 1A**). In other words, A3A levels were high in this system due to difficulty identifying
239 a sufficiently weak promoter for constitutive expression. In comparison, *A3B* and *A3B-E255A*
240 mRNA levels in Gan^R granddaughter clones were comparable to the averages reported for breast
241 cancer cell lines of the CCLE and breast tumors of the TCGA, and approximately 2-fold lower
242 than those of BT474 and MDA-MD-453 (**Figure 3 – Figure Supplement 1A**). *A3H* (haplotype I)
243 mRNA levels were higher but only two clones were analyzed by RNAseq and WGS due to
244 negative results above with the *TK* mutation analysis.

245 The mRNA expression levels of the other four *A3* genes, as well as *AICDA* (*AID*),
246 *APOBEC1*, *APOBEC2*, and *APOBEC4*, were also quantified and compared with those of *A3A*,
247 *A3B*, and *A3H* (**Figure 3 – Figure Supplement 1B**). Endogenous *A3C* was expressed at similarly
248 high levels in all granddaughter clones, providing a robust internal control. Endogenous *A3F* and
249 *A3G* were expressed at lower but still detectable levels, and endogenous *AICDA*, *APOBEC1*,
250 *APOBEC2*, *APOBEC4*, and *A3D* were expressed at very low or undetectable levels. As expected,
251 levels of ectopically expressed *A3A*, *A3B*, and *A3H* mRNA exceeded those in the HAP1-TK-M9
252 parent clone as well as those in vector expressing granddaughter controls. In addition, protein
253 level expression of A3A, A3B, and A3H was confirmed in *TK* mutant granddaughter clones by
254 immunoblotting and activity by ssDNA deamination assays (**Figure 3 – Figure Supplement 2**).
255 Finally, the integration site of the *TK* reporter was determined using WGS reads that spanned
256 both the 5' and 3'-ends of the integrated *TK* construct and the flanking human genome sequence.
257 Data from multiple clones demonstrated a single integration site in chromosome 3, which was
258 ~35 kbp from the nearest annotated coding or regulatory element.

259 To investigate mutational differences genome-wide, Illumina short-read WGS was done
260 for randomly selected Gan^R granddaughter clones (specified in **Figure 2 – Figure Supplement**

261 **1**). Mutations unique to each granddaughter were identified by calling SBS variations versus the
262 genomic DNA sequence of the HAP1-TK-M9 mother clone. This approach eliminated any somatic
263 variation that accumulated in the Gan^S mother clone prior to transduction with each of the A3 or
264 control expression constructs. Thus, all new SBS mutations were present in a significant
265 proportion of reads from the Gan^R granddaughter clones and absent from the reads from the Gan^S
266 mother clone and, as such, must have accumulated in the presence of an active A3 enzyme or a
267 catalytic mutant control.

268 In A3A-expressing Gan^R granddaughter clones, the total number of unique SBSs ranged
269 from 2595 to 5821 (n=6, median=3784, mean=4191, SD=1179). The total number of SBSs in
270 A3B-expressing clones was similar, ranging from 2794 to 2990 (n=5, median=2951, mean=2921,
271 SD=69). In comparison, the total number of SBSs in catalytic mutant control Gan^R granddaughter
272 clones ranged from 2094 to 2825 (n=4, mean=2526, SEM=273). Most importantly, analyses of
273 the trinucleotide contexts of all unique SBS mutations revealed strong C-to-T and C-to-G mutation
274 biases in 5'-TCA and 5'-TCT motifs in A3A-expressing clones and weaker, but still significant,
275 mutation biases in the same motifs in A3B-expressing clones (**Figure 3A**; **Figure 3 – Figure**
276 **Supplement 3**). In other words, only A3A- and A3B-expressing granddaughter clones elicited
277 APOBEC3 signature mutations. This important result was confirmed using an independent metric,
278 APOBEC3 signature enrichment scores (Chan, 2015 #5423;Jarvis, 2018 #6333}, which indicated
279 that 6/6 A3A-expressing clones and 4/5 A3B-expressing clones had significant enrichments for
280 APOBEC3 signature mutations, whereas clones expressing catalytically inactive A3A or A3B, as
281 well as clones expressing A3H-I or vector control had none (**Figure 3 – Figure Supplement 4A**).
282 An orthologous bioinformatics approach, non-negative matrix factorization [NMF (Alexandrov,
283 Nik-Zainal, Wedge, Campbell, & Stratton, 2013)], yielded similar results with "signature A"
284 resembling SBS2 and SBS13 in A3A- and A3B-expressing clones (*i.e.*, APOBEC3 signature) and
285 "signature B" occurring in all clones regardless of A3 functionality (**Figure 3 – Figure Supplement**
286 **4B**). Taken together, these experiments constitute the first cause-and-effect demonstration in a

287 human cell line that both A3A and A3B are capable of eliciting a clear APOBEC3 mutation
288 signature.

289 We next analyzed the broader pentanucleotide contexts of the APOBEC3 signature
290 mutations that accumulated in A3A- and A3B-expressing clones (n=6520 and n=1590 C-to-T or
291 C-to-G mutations in 5'-TCW motifs, respectively; **Figure 3B-C**). First, a bias for +1 A over +1 T
292 emerged in A3A-expressing clones (43.3% A vs 39.6% T), whereas an opposite bias was evident
293 in A3B-expressing clones (32.5% A vs 43.2% T). Second, no significant bias was noted at the +2
294 position. Third, an exceptionally strong bias for a pyrimidine nucleobase (C or T) occurred at the
295 -2 position in A3A-expressing clones (69.1% YTCW vs 30.9% RTCW). These A3A results
296 resemble the strong -2 pyrimidine bias reported for human A3A mutations in murine hepatocellular
297 carcinomas (70% YTCW vs 30% RTCW) (Law et al., 2020) and for human A3A expression in
298 yeast (Chan et al., 2015; Hoopes et al., 2016). Fourth, a weaker bias was apparent for pyrimidines
299 at the -2 position in A3B-expressing clones (51.6% YTCW vs 48.4% RTCW). This latter result
300 opposes prior results from yeast, where human A3B appears to cause a bias toward RTCW motifs
301 [*i.e.*, RTCW > YTCW (Chan et al., 2015; Hoopes et al., 2016)]. Possible explanations for these
302 contrasting results are mentioned in **Discussion**).

303

304 **Mesoscale features of A3A and A3B mutagenesis in the HAP1-TK-M9 system**

305 X-ray structures have revealed a U-shaped bend in ssDNA substrates bound by A3A and
306 A3B (Kouno et al., 2017; Shi et al., 2017), and other studies have indicated that similarly bent
307 ssDNA loop regions of hairpins (*i.e.*, DNA cruciform structures) may be preferred substrates for
308 deamination by A3A but not A3B (Buisson et al., 2019; Jalili et al., 2020; Langenbucher et al.,
309 2021). To ask whether this might be a global feature, we analyzed our A3A and A3B *TK* PCR
310 sequences and granddaughter clone WGS data for evidence of mutagenesis in the single-
311 stranded loop regions of DNA hairpin structures. First, neither of the two A3A/B mutation hotspots
312 in the *TK* gene reported above are part of obvious stem-loop structures. Second, none of the top-

313 100 cruciform structures reported previously to harbor recurring APOBEC3 signature mutations
314 in tumors (Buisson et al., 2019) were hotspots in our HAP1 WGS data sets. Third, in global
315 comparisons of APOBEC3 signature mutations in predicted loop regions of stem-loop structures
316 versus APOBEC3 signature mutations in non-cruciform structures, neither the A3A- nor the A3B-
317 expressing conditions showed an increased proportion of TCW mutations in stem-loop regions
318 relative to control conditions (Fisher's exact test, q = 1 for all comparisons; **Figure 4 – Figure**
319 **Supplement 1**).

320 To assess relative rates of A3A and A3B-catalyzed deamination of hairpin versus non-
321 hairpin substrates, we used ssDNA substrates representing two previously reported A3A
322 mutational hotspots in *SDHB* and *NUP93* (Buisson et al., 2019; Sharma et al., 2015). A3A and
323 A3B were affinity-purified from human cells and incubated with these C-containing hairpin
324 substrates over time. First, both A3A and A3B showed a strong preference for deaminating the
325 *SDHB* hairpin substrate in comparison to a linear control with the same nucleobase content
326 scrambled (~4- and ~8-fold preference, respectively; **Figure 4A**). In addition, A3A shows higher
327 rates of deamination than A3B on both the hairpin and the linear substrate in agreement with prior
328 studies (Buisson et al., 2019). Thus, the relative deamination rates for *SDHB* substrates are:
329 A3A/hairpin > A3B/hairpin > A3A/linear > A3B linear (120, 41, 31, and 5.1 nM/min, respectively).

330 However, a different picture emerged from analyses of deamination of *NUP93*-based
331 substrates **Figure 4B**). Rates of A3A-catalyzed deamination were similarly high for the *NUP93*
332 hairpin and linear control with the same nucleobase content scrambled. A3B was able to
333 deaminate the hairpin substrate but at low efficiencies and linear kinetics. In contrast, A3B showed
334 higher rates of deamination of the linear substrate. Thus, the relative deamination rates for *NUP93*
335 substrates are: A3A/hairpin = A3A/linear > A3B linear >> A3B/hairpin (93, 83, 7.7, and 1.3 nM/min,
336 respectively). These results show that reaction rates can vary massively between substrates and
337 suggest that caution should be used when attempting to extend individual structural and
338 biochemical preferences to whole genomes. Nevertheless, these data conservatively show that

339 *both* A3A and A3B can deaminate hairpin and linear substrates and, as proposed (Buisson et al.,
340 2019), it is possible to find sites such as the *NUP93* hairpin that are strongly (but not exclusively)
341 preferred by a single A3 enzyme.

342 Another mesoscale feature of APOBEC3 mutagenesis in human cancer is clusters of
343 strand-coordinated cytosine SBS mutations in TCA and TCT motifs most likely caused by
344 processive deamination of exposed tracts of ssDNA (aka. *kataegis*) (Alexandrov, Nik-Zainal,
345 Wedge, Aparicio, et al., 2013; Alexandrov, Nik-Zainal, Wedge, Campbell, et al., 2013; Nik-Zainal
346 et al., 2012; Roberts et al., 2012). No APOBEC3 signature *kataegic* events were observed in
347 control conditions. However, several APOBEC3 signature *kataegic* events were evident in the
348 genomic DNA of both A3A and A3B expressing granddaughter clones (**Figure 4C**). For instance,
349 one A3A-attributable *kataegic* event was comprised of 7 T(C>T/G)A/T events within a 10 kbp
350 window, and an A3B-attributable *kataegic* event included 8 T(C>T/G)A/T events within a 10 kbp
351 window. Interestingly, the frequency of *kataegic* events did not differ significantly between A3A-
352 and A3B-expressing granddaughter clones (A3A WT: median = 30, mean = 35 events, SEM =
353 8.3; A3B WT: median = 12, mean = 15, SEM = 6.9; p = 0.28 by Welch's t-test). These results are
354 the first to indicate in a human cell line that *both* A3A and A3B can cause kataegis.
355

356 **APOBEC signature etiology in primary breast tumors**

357 Sequencing data from model systems such as HAP1-TK-M9 are powerful because the
358 resulting mutation signatures can be considered relatively pure and, therefore, can help to
359 establish cause-and-effect relationships for comparison to more complex tumor WGS data sets
360 to identify similarities and, potentially, to infer which enzyme(s) generated the observed mutation
361 signature in a given tumor. We therefore performed a Euclidean distance analysis to compare the
362 pentanucleotide cytosine mutation signatures derived from sequencing A3A, A3B, and A3H
363 expressing HAP1 clones and those from 794 primary breast tumors with WGS available through
364 the ICGC data portal resource. This analysis revealed three distinct tumor groups with respect to

365 APOBEC3 signature mutations: 1) a group that showed similarity to A3A-expressing HAP1-TK-
366 M9 clones, 2) a group that showed similarity to A3B-expressing HAP1-TK-M9 clones, and 3) a
367 group that showed little to no significant APOBEC3 mutation signature (**Figure 5**). As expected
368 from a lack of substantial APOBEC signature mutations above in the *TK* reporter or in
369 representative WGS, A3H-I expressing clones and catalytic-inactive clones clustered with the
370 non-APOBEC signature tumors.

371 Importantly, both the A3A-like and A3B-like groups were comprised of tumors that showed
372 high levels of APOBEC3 signature mutations and correspondingly high enrichment scores.
373 However, even in the A3A-like group, none of the breast tumors showed ≥70% APOBEC3
374 signature mutations in YTCW motifs [the overall bias observed here in the HAP1-TK-M9 system
375 and previously in murine tumors (Law et al., 2020)]. This key result strongly suggests that the
376 observed APOBEC3 mutation signatures in most breast tumors may be a composite resulting
377 from both A3A and A3B activity. In comparison, the homologous recombination repair deficiency
378 signature (SBS3) appeared underrepresented in both the A3A- and A3B-like tumor groups, and
379 the ageing signature (SBS1) occurred in all 3 groups regardless of A3 mutation status.

380

381 **Discussion**

382 Here we report the development and implementation of a novel system to investigate
383 mutational processes in human cells (**Figure 1**). Like many bacterial and yeast model systems,
384 the HAP1-TK-M9 system enables a uniform cytotoxic selection with ganciclovir such that only *TK*
385 mutant cells survive. An analysis of clonally derived, A3A and A3B expressing *TK* mutants by
386 Sanger sequencing of high-fidelity PCR amplicons demonstrates a strong shift in the mutational
387 pattern from a variety of different base substitution mutations in control conditions to a strongly
388 5'TCW-biased pattern (**Figure 2**). Interestingly, the *TK* mutation spectra inflicted by A3A and A3B
389 are virtually indistinguishable including two shared hotspots (Q8X, R212K) and no obvious
390 mesoscale features. This may be due to the limited number of mutable cytosines in *TK* that confer

391 resistance to ganciclovir (local base composition) and/or to selective pressure. However, a striking
392 analogy can be made with the mutational spectrum of the *PIK3CA* gene in breast, head/neck, and
393 others cancers, which has two prominent APOBEC3 signature hotspots (E542K, E545K) and no
394 obvious mesoscale structures (Buisson et al., 2019; Henderson, Chakravarthy, Su, Boshoff, &
395 Fenton, 2014). These observations combine to suggest that selective pressure has the potential
396 to overshadow the intrinsic preferences of the different APOBEC3 enzymes and complicate
397 assignment of direct cause-and-effect relationships. Drawing connections between A3A and/or
398 A3B and a given mutation, even a prominent hotspot, is additionally challenging due to the fact
399 that both enzymes can deaminate DNA cytosines in both linear substrates as well as single-
400 stranded loop regions of stem-loop structures [*e.g.*, **Figure 4** and prior biochemical studies
401 (Buisson et al., 2019; Jalili et al., 2020; Langenbucher et al., 2021)]. Thus, the *TK*-based system
402 described here is capable of yielding informative, rapid, and inexpensive mutation data sets with
403 positive results motivating genome-wide analyses where the vast majority of mutations will be
404 unselected and the larger mutation data sets will enable even stronger conclusions. Earlier work
405 also leveraged HAP1 cells to characterize a variety of different mutation sources but these studies
406 did not incorporate a selectable reporter nor did they address the enzyme(s) responsible for
407 APOBEC3 signature mutations (Zou et al., 2021; Zou et al., 2018).

408 The studies here are the first to demonstrate unambiguously that both A3A and A3B can
409 inflict an APOBEC3 mutation signature in human genomic DNA with, in both instances, ssDNA
410 deamination events immortalizing predominantly as C-to-T and C-to-G mutations in TCA and TCT
411 trinucleotide motifs (**Figure 3**). Over identical timeframes, A3A causes 4.1-fold more APOBEC3
412 signature mutations in comparison to A3B (6520 vs 1590 mutations in 6 and 5 subclone WGSs,
413 respectively). This difference can be explained by higher A3A expression levels in the HAP1-TK-
414 M9 system (beyond levels in tumor) and the higher intrinsic activity of this enzyme biochemically
415 in comparison to A3B. Regardless, both A3A and A3B yielded thousands of TC-focused signature
416 mutations and the most striking difference between these enzymes is the stronger preference of

417 A3A for a pyrimidine at the -2 position (70.5% YTCW for A3A vs 54.3% YTCW for A3B). This -2
418 pyrimidine bias mirrors recent WGS results from human A3A-induced murine hepatocellular
419 carcinomas [70% YTCW (Law et al., 2020)], and it also resembles the bias reported originally for
420 human A3A expression in yeast (Chan et al., 2015; Hoopes et al., 2016). It also mirrors a strong
421 -2 pyrimidine bias reported for A3A in the chicken B cell line DT40 [(DeWeerd et al., 2022) that
422 was published during preparation of this manuscript]. However, importantly, A3B also shows a
423 YTCW bias here in human cells, which is opposite to the RTCW bias reported in yeast (Chan et
424 al., 2015; Hoopes et al., 2016). This conflicting result may be due to differences in genetic and/or
425 epigenetic factors including but not limited to base content, chromatin state, DNA repair
426 processes, and/or regulatory factors specific to human A3B expression in human cells.
427 Regardless of the precise explanation(s), the fact that both A3A and A3B can elicit YTCW
428 mutational biases in human cells helps to inform interpretations of the APOBEC3 enzyme
429 responsible for the overall APOBEC3 mutation program in cancer. For instance, the observed
430 APOBEC3 mutation signature in tumors is likely to be a composite of A3A and A3B because very
431 few (if any) human cancers elicit a 70% YTCW mutation biases indicative of A3A exclusivity or a
432 52% YTCW bias indicative of A3B exclusivity. In support of this inference, our Euclidean distance
433 analysis APOBEC3-signature motifs from HAP1-TK-M9 WGSs here and nearly 800 breast tumor
434 WGSs identified both A3A- and A3B-like subgroups, with neither enzyme's preferred motif fully
435 explaining the composite signature across breast cancer (**Figure 5**).

436 In addition to selective pressures and mesoscale features, additional factors are likely to
437 influence the APOBEC3 signature component of an overall tumor mutational landscape including
438 whether A3A and/or A3B is expressed, expression level, duration of expression, intrinsic activity,
439 and accessibility of chromosomal DNA (replication stress, R-loop levels, chromatin state, *etc.*).
440 With regards to studies here with the HAP1-TK-M9 system, both A3A and A3B are expressed
441 constitutively from the same construct for identical durations prior to ganciclovir selection, A3A is
442 intrinsically more active than A3B, A3A is cell-wide and A3B predominantly nuclear, and yet these

443 enzymes and cellular factors combine to yield remarkably similar *TK* mutation frequencies and
444 only a 4-fold difference in overall genome-wide SBS mutation level. With respect to cancer, the
445 *A3A* gene is expressed at lower levels than *A3B* in almost all cell lines and tumors, A3A is cell-
446 wide or predominantly cytoplasmic where A3B is constitutively nuclear, and A3A has higher
447 enzymatic activity that can vary from 2- to 100-fold above that of A3B depending on substrate
448 [*e.g.*, **Figure 4** and prior biochemical studies (Buisson et al., 2019; Ito et al., 2017)]. It is therefore
449 remarkable here that the overall genome-wide level of APOBEC3 signature mutation from A3A is
450 only ~4-fold higher than that attributable to A3B. Endogenous *A3A* and *A3B* also have both distinct
451 and overlapping transcription programs, and both genes can be induced by a variety of conditions
452 including viral infection and inflammation (Argyris et al., 2020; Cho et al., 2018; Leonard et al.,
453 2015; Maruyama et al., 2016; Oh et al., 2021; Periyasamy et al., 2015; Periyasamy et al., 2017;
454 Roelofs et al., 2020; Starrett et al., 2019; Verhalen et al., 2016; Vieira et al., 2014; Warren,
455 Westrich, Doorslaer, & Pyeon, 2017; Warren et al., 2015; Westrich et al., 2018). *In vivo*, *A3A* and
456 *A3B* gene expression is also likely to be affected by the local tumor microenvironment, which can
457 vary both between and within cancer types, as well as by a patient's global state of health. Taken
458 together with unknown and likely lengthy multi-year durations of tumor development prior to
459 clinical manifestation, deducing the precise fractions of mutations attributable to A3A and/or A3B
460 may be fruitless.

461 Independent whole genome sequencing experiments have provided additional information
462 on the APOBEC3 mutation process. Initial studies induced overexpression of A3B in HEK-293-
463 derived cell lines, documented the resulting DNA damage responses, and performed WGS to
464 assess genome-wide associations (Akre et al., 2016; Nikkila et al., 2017). However, an
465 unambiguous APOBEC3 mutation signature was difficult to extract from these whole genome
466 sequences due to large numbers of mutations attributable to compromised mismatch repair (Akre
467 et al., 2016; Nikkila et al., 2017). A more recent study compared *de novo* mutations occurring in
468 APOBEC3 signature positive cell lines during multiple generations of clonal outgrowth (M. Petljak

469 et al., 2019). An intriguing finding from this work is that APOBEC3 signature mutations may be
470 able to occur in an episodic manner, accumulating in some generations and not others, consistent
471 with evidence discussed above that A3A and A3B expression can be induced by multiple signal
472 transduction pathways. Episodic mutagenesis, however, is unexpected in cell-based systems in
473 which stochastic mutagenesis should predominate given defined media and well-controlled
474 growth conditions. These studies were followed-up more recently by WGS comparisons of
475 subclones of the same cancer cell lines CRISPR-engineered to lack A3A or A3B (M. Petljak et
476 al., 2021). Interestingly, disruption of A3A but not A3B caused a near-complete loss of SBS2 and
477 SBS13 and a strong diminution of YTCA- and YTCN-focused APOBEC3 signature mutations. A
478 simplistic interpretation of these results is that A3A is the primary enzyme responsible for
479 APOBEC3 signature mutagenesis in these cancer cell lines. However, as discussed above,
480 conditions in culture are necessarily controlled and, at best, are likely to only recapitulate a subset
481 of the more complex local and global conditions of A3 expression, mutagenesis, and tumor
482 development *in vivo*. Thus, roles for A3B may be underestimated under such experimental
483 conditions.

484 A role for A3H, haplotypes I or II, in cancer is disfavored by our results here showing that
485 these enzymes are incapable of eliciting DNA damage responses or increasing the *TK* mutation
486 frequency. Two A3H-I expressing *TK* mutant clones were subjected to WGS and no APOBEC3
487 mutation signature was evident. In addition, no evidence for A3H emerged from sequencing
488 clonally-derived breast and B cell lymphoma cancer cell lines (M. Petljak et al., 2021). However,
489 all of these cell-based studies have limitations as discussed above and have yet to fully eliminate
490 A3H as a source of APOBEC3 signature mutations in cancer. For instance, A3H-I may take more
491 time to inflict detectable levels of mutation, it may be subject to different transcriptional and post-
492 transcriptional regulatory processes, and/or it may only be mutagenic in a subset of cancer types
493 subject to different stresses and different selective pressures.

494 Ultimately, the studies here show that both A3A and A3B are capable of triggering overt

495 DNA damage responses and inflicting a robust APOBEC3 mutation signature in the HAP1-TK-
496 M9 system and, taken together with other work summarized above, support a model in which both
497 of these enzymes contribute to the composite APOBEC3 mutation signature reported in many
498 different tumor types. This conclusion is supported by clinical studies implicating A3A and/or A3B
499 in a variety of different tumor phenotypes including drug resistance/susceptibility, metastasis, and
500 immune responsiveness (Argyris et al., 2020; Burns, Lackey, et al., 2013; Cannataro et al., 2019;
501 T. W. Chen et al., 2017; Z. Chen et al., 2019; Cortez et al., 2019; Driscoll et al., 2020; Faden et
502 al., 2017; Glaser et al., 2018; Law et al., 2016; Middlebrooks et al., 2016; Sieuwerts et al., 2017;
503 Sieuwerts et al., 2014; Tsuboi et al., 2016). Thus, efforts to diagnose and treat APOBEC3
504 signature-positive tumors should endeavor to take into account both enzymes, not simply one or
505 the other. Such longer-term goals are not trivial given the high degree of identity between A3A
506 and the A3B catalytic domain (>90%), the related difficulty of developing specific and versatile
507 antibodies for detecting each enzyme, and the fact that each can be regulated differentially by a
508 wide variety of common factors including virus infection and inflammation. Thus, we are hopeful
509 that the HAP1-TK-M9 system and A3A-specific rabbit monoclonal antibody described here will
510 help to expedite the achievement of these goals.
511
512 **Materials and Methods**
513 **Key resources table**

Reagent type (species) or resource	**Designation**	**Source or reference**	**Identifiers**	**Additional information**
Cell line (*homo sapiens,* male*)*	HAP1	Horizon	Cat#: C859	

Cell line (*homo sapiens,* male*)*	HAP1-TK-M9	Original clone, this study		Near diploid clone; expresses HSV-1 *TK* and is Gan^S
Cell line (*homo sapiens,* female*)*	293T	ATCC	Cat#: CRL-3216 RRID: CVCL_0063	
Cell line (*homo sapiens,* male*)*	THP1	ATCC	Cat#: TIB-202 RRID: CVCL_0006	
Antibody	Anti-A3A (rabbit Monoclonal)	This study	UMN13	WB (20 ng/mL)
Antibody	Anti-A3B (rabbit monoclonal)	(Brown et al., 2019)	Cat#:12397 RRID:AB_2721202	WB (1:1,000) IF (1:300)
Antibody	Anti-A3H (rabbit polyclonal)	Novus ARP10	Cat#: NBP1-91682 RRID: AB_2057523	WB (1:1,000) IF (1:300)
Antibody	Anti-γ-H2AX (S139) (rabbit polyclonal)	Millipore Sigma JW301	Cat#: H5912 RRID: AB_310406	IF (1:300)
Antibody	Anti-Tubulin (mouse monoclonal)	Sigma-Aldrich	Cat#: T5168 RRID: AB_477579	WB (1:20,000)

Recombinant DNA reagent	pQCXIP MLV vector	NovoPro	Cat#: V010396#	Parental transduction vector
Recombinant DNA reagent	pQCXIP-MND-A3Ai-IRES-Puro	This study	A3A with intron (pRH9978)	Request by contacting RSH
Recombinant DNA reagent	pQCXIP-MND-A3Ai-E72A-IRES-Puro	This study	A3A-E72A with intron (pRH9979)	Request by contacting RSH
Recombinant DNA reagent	pQCXIP-MND-A3Bi-IRES-Puro	This study	A3B with intron (pRH9980)	Request by contacting RSH
Recombinant DNA reagent	pQCXIP-MND-A3Bi-E255A-IRES-Puro	This study	A3B E255A with intron (pRH9981)	Request by contacting RSH
Recombinant DNA reagent	pQCXIP-MND-eGFP-IRES-Puro	This study	eGFP vector control (pRH9977)	Request by contacting RSH
Recombinant DNA reagent	pQCXIP-MND-A3H-Ii-IRES-Puro	This study	A3H hapI with intron (pRH9984)	Request by contacting RSH
Recombinant DNA reagent	pQCXIP-MND-A3H-IIi-IRES-Puro	This study	A3H hapII with intron (pRH9985)	Request by contacting RSH
Recombinant DNA reagent	pQCXIP-MND-A3H-IIi-E56A-IRES-Puro	This study	A3H hapII E56A with intron (pRH9986)	Request by contacting RSH
Software, algorithm	Burrows-Wheeler Aligner (BWA)		RRID: SCR_010910	

Software, algorithm	Fiji	Fiji	RRID: SCR_002285	
Software, algorithm	GraphPad Prism 6	GraphPad	RRID: SCR_002798	
Software, algorithm	ImageQuant	GE Healthcare	RRID: SCR_014246	
Software, algorithm	Image Studio	LI-COR Biosciences	RRID: SCR_015795	
Software, algorithm	MaxQuant version 1.5.2.8	MaxQuant	RRID: SCR_014485	
Software, algorithm	R for Statistical Computing		RRID: SCR_001905	
Software, algorithm	SpeedSeq		RRID: SCR_000469	
Software, algorithm	VarScan2		RRID: SCR_006849	
Other	BD LSRFortessa Flow Cytometer	BD Biosciences	RRID: SCR_019600	PI stain quantification
Other	LI-COR Odyssey FC	LI-COR	Cat#: 2800	WB imaging

Other	Typhoon FLA 7000	GE Healthcare	Cat#: 29-0044-13	Oligo cleavage assay imaging
Other	MiSeq Sequencing System	Illumina	RRID: SCR_016379	WGS library balancing validation
Other	NovaSeq 6000 Sequencing System	Illumina	RRID: SCR_016387	WGS

514

515

516 **Cell lines and culture conditions**

517 All cells lines were cultured at 37°C under 5% CO_2. HAP1 cells and derivatives were grown
518 in IMDM (Invitrogen) supplemented with 10% fetal bovine serum (Sigma), penicillin (100 U/mL),
519 and streptomycin (100 µg/mL). 293T cells were cultured in DMEM (Invitrogen) with 10% fetal
520 bovine serum (Sigma), penicillin (100 U/mL), and streptomycin (100 µg/mL). THP1 cells were
521 cultured in RPMI (ThermoFisher) supplemented with 10% fetal bovine serum (Sigma), penicillin
522 (100 U/mL), and streptomycin (100 µg/mL). TransIT-LT1 (Mirus) was used for all transfections.
523 All parent and clonal lineage cell lines tested negative for mycoplasma using a PCR-based assay
524 (Uphoff & Drexler, 2011). Puromycin (ThermoFisher) and neomycin (ThermoFisher) were used
525 at 1 µg/mL and 50 µg/mL, respectively. Ganciclovir (ThermoFisher) was used at 5 µM to select
526 *TK* mutant clones.

527

528 **Creating the HAP1-TK-M9 system**

529 The HAP1-TK-M9 system was generated by co-transfecting HAP1 parent cells with a
530 plasmid expressing the Sleeping Beauty transposase and a separate plasmid with *TK-Neo* coding
531 sequences flanked by SB recognition sites (Burns, Lackey, et al., 2013; Stenglein, Burns, Li,
532 Lengyel, & Harris, 2010). Semi-confluent cells in 6 well plates were transfected, treated 24 hrs

533 later with G418 (1 μg/mL), and subcloned by limiting dilution in 96 well plates to create single cell
534 derivatives. Single cell clones were then expanded and characterized as described in the main
535 text. The *A3H* genotype was determined by sequencing exon-specific PCR amplicons (Refsland
536 et al., 2010; Wang et al., 2019) and further confirmed by WGS (below).

537 Standard molecular cloning procedures were used to create derivatives of MLV pQCXIP
538 for expressing each A3 protein. First, pQCXIP was cut with MluI and PacI to excise the strong
539 CMV promoter and replace it with a weaker MND promoter (a synthetic promoter containing
540 regions of both the MLV LTR and the myeloproliferative sarcoma virus enhancer). Second, this
541 new construct was cut with SfiI and BsiWI to insert intron-disrupted A3 coding sequences (Law et
542 al., 2020). This was done for *A3A*, *A3B*, *A3H* haplotype-I, *A3H* haplotype-II, and appropriate
543 catalytic mutant derivatives (E-to-A). An eGFP expressing construct was generated in parallel to
544 use as a control in various experiments. All new constructs were confirmed by Sanger
545 sequencing. Each MLV-based construct was co-transfected into 293T cells with appropriate
546 packaging vectors and 48 hrs later the resulting viral supernatants were filtered (0.2 μm) and used
547 to transduce semi-confluent HAP1-TK-M9 cells. After an additional 48 hrs incubation, transduced
548 cells were selected with puromycin (1 μg/mL) and subcloned by limiting dilution to obtain A3
549 expressing daughter clones. These A3 expressing and control daughter clones were expanded
550 and characterized as described in the main text. Mutation frequencies were determined by plating
551 5 x 10^6 cells in 96 well flat bottom plates, treating with 5 μM ganciclovir (ThermoFisher), and after
552 14 days incubation counting the number of *TK* mutant colonies that survived selection. Single
553 Gan^R granddaughter clones were counted using a light microscope and expanded and
554 characterized as described in the main text including immunoblotting, DNA deaminase activity
555 assays, *TK* sequencing, and WGS.

556

557 **Immunoblots (IB)**

558 Cells were treated with trypsin EDTA and collected, washed in 1X PBS, and re-suspended
559 in 100 μL of reducing sample buffer per one million cells [0.5 M Tris-HCl pH 6.8, 1% 2-
560 mercaptoethanol, 10% sodium dodecyl sulfate (SDS), 50% glycerol]. Proteins were denatured by
561 boiling samples for 20 min and resolved by SDS-polyacrylamide gel electrophoresis. Proteins
562 were transferred to a PVDF-FL membrane (Millipore Sigma) and blocked in 5% milk in 1X PBS.
563 Primary and secondary antibodies were incubated in blocking buffer, with the addition of 0.2%
564 SDS for fluorescent antibodies. The primary antibodies used were anti-A3A/B [5210-87-13, IB:
565 1:1,000, IF: 1:300 (Brown et al., 2019)], anti-A3H [Novus, IB: 1:1,000, IF: 1:300 (Refsland et al.,
566 2014)], anti-Tubulin (Sigma Aldrich, 1:20,000), and anti-γ-H2AX (S139) (Millipore Sigma JBW301,
567 1:300). The secondary antibodies used were anti-rabbit HRP-linked (CST 7074, 1:2,000), IRDye
568 800CW goat anti-mouse (LI-COR Biosciences, 1:10,000), Alexa Fluor 680 goat anti-rabbit
569 (Molecular Probes, Eugene, OR, USA, 1:10,000), and Alexa Fluor 680 goat anti-mouse
570 (Molecular Probes, Eugene, OR, USA, 1:10,000). Membranes were imaged using a LI-COR
571 Odyssey instrument or LI-COR Odyssey-Fc instrument for HRP visualization (LI-COR
572 Biosciences).
573
574 **DNA deaminase activity assays**
575 Whole cell extract (WCE) assays: ssDNA deamination activities were measured using
576 WCE prepared using 100 μL of HED lysis buffer per 1 million cells (25 mM HEPES, 5 mM EDTA,
577 10% Glycerol, 1 mM DTT, and 1 protease inhibitor tablet). The samples were sonicated 3 times
578 for 5 sec to ensure complete lysis. A3-containing lysates were incubated at 37°C for 2 hrs with
579 purified human UNG2 and a single-stranded DNA substrate containing a single TCA trinucleotide
580 motif and 3' FAM label (5'-ATTATTATTATTCAAATGGATTTATTTATTTATTTATTTATTT-FAM-
581 3') following established protocols (Burns, Lackey, et al., 2013; Lackey et al., 2012). After this
582 initial incubation, the reaction was treated with 1 mM NaOH for 10 min at 98°C. The reaction was

583 run out on a 15% TBE acrylamide urea gel to separate substrate oligo from cleaved product oligo
584 and imaged on a Typhoon FLA 7000 with ImageQuant TL 8.2.0 software (GE Healthcare).
585 Recombinant enzyme assays: A3A- and A3B-mycHis were prepared from transfected
586 293T as reported (Carpenter et al., 2012; McCann et al., 2021; Shaban et al., 2022; Shi et al.,
587 2017; Shi, Carpenter, Kurahashi, Harris, & Aihara, 2015). 25 nM of each protein was incubated
588 for the indicated times with 800 nM substrate in HEPES buffer (25 nM HEPES pH 7.4, 50 mM
589 NaCl, 5 mM imidazole) at 37°C. Reactions were stopped by freezing in liquid nitrogen and then
590 were heated to 95°C to denature the enzymes. Reactions were then treated with 0.5 U/reaction
591 Uracil DNA Glycosylase (NEB, USA) for 10 min at 37°C. Sodium hydroxide was added to each
592 reaction to 100 mM and reactions were heated to 95°C for 5 minutes to cleave the DNA. Products
593 were separated by 20% TBE-Urea PAGE, imaged on a Typhoon FLA-7000 (GE Healthcare,
594 USA), and quantified using ImageQuant TL 8.2.0 software (GE Healthcare). Deamination of the
595 target cytosine was calculated by dividing the total reaction product (and any subsequent off-
596 target cytosine deamination events) by the total amount of starting substrate.
597
598 **PI staining and flow cytometry**
599 Propidium iodide staining was used to assess the ploidy of HAP1 clones relative to THP1
600 as a confirmed diploid control. Cells were trypsinized and suspended in 100 µL of 1X PBS per 1
601 million cells. Then, 500 µL ice-cold ethanol was added to cell suspensions and incubated at -20°C
602 for 1 hour to fix the cells. After fixation, cells were pelleted and washed in 1X PBS. Cells were
603 finally suspended in 500 µL of FxCycle PI stain (Invitrogen) and incubated for 30 minutes at RT
604 in the dark to stain the cells. Cells were spun down and resuspended in 300 µL of the PI stain
605 solution and placed in a 96 well round bottom plate for flow cytometry analysis using a BD
606 LSRFortessa flow cytometer (with high-throughput 96 well adapter system). A minimum of ten
607 thousand events were acquired for each cell line.
608

609 **RT-qPCR**

610 Total RNA was extracted using the High Pure RNA isolation kit (Roche). cDNA was
611 synthesized using SuperScript First-Strand RT (ThermoFisher). Quantification of mRNA was
612 done using validated primer sets for all human *A3* genes relative to the housekeeping gene *TBP*
613 (Burns, Lackey, et al., 2013; Leonard et al., 2015; Refsland et al., 2010; Stenglein et al., 2010).
614 All RT-qPCR reaction were performed using SsoFast SYBR Green mastermix (Bio-Rad) in 384-
615 well plates on a LightCycler 480 (Roche) following the manufacturer's protocol. Statistical
616 analyses were done using GraphPad Prism 6 and R.

617

618 **Immunofluorescent microscopy (IF)**

619 IF was done as described (Lackey, Law, Brown, & Harris, 2013; McNamara, McCann,
620 Gudipaty, & D'Orso, 2013). Cells were grown in 6 well tissue culture plates at low density prior to
621 fixation. Cells were fixed with 4% paraformaldehyde in PBS for 15 minutes at RT. Cells were then
622 permeabilized using PBS containing 0.2% Triton X-100 (Sigma Aldrich), then rinsed with PBS.
623 Cells were blocked using IF blocking solution and 0.1% Triton X-100 for 1 hr at RT, then incubated
624 with primary antibody overnight at 4°C. The primary antibodies used were anti-A3A/B (5210-87-
625 13, 1:300) (Brown et al., 2019), anti-A3H (Novus, 1:300) (Refsland et al., 2014), and anti-γ-H2AX
626 (JBW301, Millipore Sigma, 1:300). Cells were then washed with PBS and incubated with a fluoro-
627 conjugated secondary antibody for 2 hrs at RT in the dark. The secondary antibodies used were
628 IRDye 800CW goat anti-mouse (LI-COR Biosciences, 1:10,000) and Alexa Fluor 680 goat anti-
629 rabbit (Molecular Probes, Eugene, OR, USA, 1:10,000). Primary and secondary antibodies were
630 diluted in blocking buffer. Hoescht 33342 (Mirus) was used at a final concentration of 1 μg/mL to
631 stain nuclei.

632

633 **COMET assays**

634 HAP1-TK-M9 clones expressing individual A3 enzymes were harvested at 70%
635 confluency and resuspended in ice cold 1X phosphor-buffered saline (PBS, Ca^{2+} and Mg^{2+} free,
636 $1.0x10^5$ cells/mL). The CometAssay kit and manufacturer's protocol were used for all alkaline
637 comet assays (Trevigen). Cells were spread at low density on a glass slide and covered in low-
638 melt agarose to keep the cells in position for lysis. After cell lysis, slides were washed 2 times with
639 1X UDG buffer (New England Biolabs), and then incubated with purified human UNG2 in 1X UDG
640 buffer (4 ng/μL) for 1 hour at 37°C. After incubation, DNA unwinding and electrophoresis were
641 done according to the manufacturer's alkaline protocol. Comets were analyzed using the
642 OpenComet plugin for Image J (Gyori, Venkatachalam, Thiagarajan, Hsu, & Clement, 2014) and
643 statistical analyses were done in GraphPad Prism 6 and R.
644
645 ***TK* sequencing**
646 Cells were harvested and genomic DNA was isolated using the Puregene DNA isolation
647 protocol. The *TK* cassette was amplified from genomic DNA using 5'-
648 ATCTTGGTGGCGTGAAACTC-3' and 5'-CTTCCGGTATTGTCTCCTTCC-3'. PCR products
649 were cleaned-up using the GeneJet Gel Extraction Kit (Thermo Scientific) and Sanger sequenced
650 with 4 different primers to cover the full open reading frame (5'-ATCTTGGTGGCGTGAAACTC-
651 3', 5'-GGTCATGCTGCCCATAAGGTA-3', 5'-CCGTTCTGGCTCCTCATATC-3', and 5'-
652 CTTCCGGTATTGTCTCCTTCC-3'.
653
654 **Whole genome sequencing (WGS)**
655 Genomic DNA was prepared from cell pellets (1 million cells) using Allprep DNA/RNA mini
656 kit (Qiagen). Whole genome libraries were sequenced 150x2 bp on a NovaSeq 6000 (Illumina) to
657 a target read depth of 30X coverage for all granddaughter clones as well as the parental HAP1-
658 TK-M9 mother clone. Resulting sequences were aligned to the human genome (hg38) using
659 SpeedSeq (Chiang et al., 2015), which relies on the Burrows-Wheeler Aligner, BWA (version

660 0.7.17). PCR duplicates were removed using Picard (version 2.18.16). Reads were locally
661 realigned around InDels using GATK3 (version 3.6.0) tools RealignerTargetCreator to create
662 intervals, followed by IndelRealigner on the aligned bam files. Single base substitutions and small
663 InDels were called in each clone relative to the bam file generated from the HAP1-TK-M9 mother
664 clone using Mutect2 from GATK3 (version 3.6.0). SBSs that passed the internal GATK3 filter with
665 minimum 4 reads supporting each variant, minimum 20 total reads at each variant site and a
666 variant allele frequency over 0.05 were used for downstream analysis. SBSs were analyzed in R
667 (version 4.0.5) using the MutationalPatterns (Blokzijl, Janssen, van Boxtel, & Cuppen, 2018) and
668 deconstructSigs R packages [version 1.8.0 (Rosenthal, McGranahan, Herrero, Taylor, &
669 Swanton, 2016)]. All visualizations were generated using the ggplot2 package (version 3.3.5).
670 COSMIC single base substitution mutation signatures (v3 – May 2019
671 https://cancer.sanger.ac.uk/cosmic/signatures/SBS/) were obtained from
672 https://www.synapse.org/#!Synapse:syn11738319. *De-novo* non-negative matrix factorization of
673 mutational signatures was performed with the "extract_signature" command from the
674 MutationalPatterns package, with a rank of 2 and 100 iterations. TCW mutation enrichment scores
675 were calculated as described (Chan et al., 2015; Jarvis, Ebrahimi, Temiz, & Harris, 2018).
676 Sequence logos of -2 to +2 sequence surrounding C-to-T mutations were created using the
677 ggseqlogo (version 0.1) package.
678
679 **APOBEC expression and mutation signature analyses in TCGA and ICGC data sets**
680 TCGA primary breast tumors represented by both RNA-seq and whole exome sequencing
681 were downloaded from the Firehose GDAC resource through the Broad Institute pipeline
682 (http://gdac.broadinstitute.org/) for multiple tumor tissue types. APOBEC mutation signatures
683 were determined as described (Alexandrov, Nik-Zainal, Wedge, Aparicio, et al., 2013; Jarvis et
684 al., 2018) using the deconstructSigs R package (Rosenthal et al., 2016). APOBEC mutation
685 enrichment scores were calculated using the hg19 reference genome and published methods

686 (Chan et al., 2015). Enrichment score significance was assessed using a Fisher exact test with
687 Benjamini-Hochberg false discovery rate (FDR) correction. All dataset analyses and
688 visualizations were conducted using R and the ggplot2 package (https://www.R-project.org/).
689

690 ***TK* integration site determination**

691 To determine the integration site of the single copy *TK-Neo* construct, a TK reference
692 sequence was provided as an additional chromosome during alignment of the WGS reads to the
693 reference genome (hg38). Reads that mapped to this region were then categorized as discordant
694 and realigned using GRIDSS (v2.2) (Cameron et al., 2017) to determine site of integration.
695

696 **Euclidean distance analysis**

697 All primary breast tumor whole genome sequencing variant information from International
698 Cancer Genome Consortium (ICGC) was downloaded from the ICGC data portal
699 (https://dcc.icgc.org). SBSs that were used in these analyses included only C-to-T variants in a
700 TC dinucleotide, excluding all mutations as CG motifs. SBSs meeting these inclusion criteria from
701 all clones expressing A3A, A3B, A3A-E72A, A3B-E255A, and A3H-I were pooled per condition
702 for this analysis. A matrix comprised of the number of mutations within a tetranucleotide across
703 all samples within a cancer type was generated, and counts were normalized to frequency within
704 each cancer type. This matrix was converted to a hierarchical distance matrix using the hclust
705 function in R, which was then plotted as dendrograms. Mutation signatures were calculated using
706 deconstructSigs as above.
707

708 **Data and code availability**

709 All alignment files (FASTQ and BAM format) are available through the Sequence Read
710 Archive under the BioProject accession number PRJXXXXXXXX.
711

712 **Author contributions**

713 R.S. Harris and M.C. Jarvis conceptualized and designed the overall project. M.C. Jarvis,
714 M.A. Carpenter, M.R. Brown, K. Richards, P.P. Argyris, and W.L. Brown performed experiments.
715 M.C. Jarvis, M.A. Carpenter, N.A. Temiz, M.R. Brown, P.P. Argyris, and W.L. Brown did formal
716 data analysis. R.S. Harris, M.C. Jarvis, and D. Yee contributed to funding acquisition. R.S. Harris
717 and M.C. Jarvis drafted the manuscript, and all authors contributed to manuscript proofing and
718 revision.

719

720 **Additional contributions**

721 We thank Arad Moghadasi and Sofia Moraes for assistance with flow cytometry and
722 immunofluorescence microscopy, respectively. We also thank the UMN Cytogenetics Core for
723 expert assistance with metaphase chromosome analyses and Scott McIvor for sharing a plasmid
724 construct with the MND promoter. The results presented here are in part based upon data
725 generated by the TCGA Research Network: http://www.cancer.gov/tcga. Additionally, data from
726 the International Cancer Genome Consortium (ICGC) were used in these analyses:
727 https://dcc.icgc.org/.

728

729 **Additional information**

730 **Funding**

731 These studies were supported in part by NCI P01-CA234228 (to RSH). Salary support for
732 MCJ was provided in part by T32-CA009138 and subsequently F31-CA243306. RSH is the Ewing
733 Halsell President's Council Distinguished Chair at University of Texas San Antonio and an
734 Investigator of the Howard Hughes Medical Institute.

735

736 **Compliance with ethical standards**

737 Conflict of interest: The authors have no conflicts to declare.

738

739 **References**

740 Akre, M. K., Starrett, G. J., Quist, J. S., Temiz, N. A., Carpenter, M. A., Tutt, A. N., . . . Harris, R.
741 S. (2016). Mutation processes in 293-based clones overexpressing the DNA cytosine
742 deaminase apobec3b. *PLoS One, 11*(5), e0155391. doi:10.1371/journal.pone.0155391

743 Alexandrov, L. B., Jones, P. H., Wedge, D. C., Sale, J. E., Campbell, P. J., Nik-Zainal, S., &
744 Stratton, M. R. (2015). Clock-like mutational processes in human somatic cells. *Nat Genet,*
745 *47*(12), 1402-1407. doi:10.1038/ng.3441

746 Alexandrov, L. B., Kim, J., Haradhvala, N. J., Huang, M. N., Tian Ng, A. W., Wu, Y., . . .
747 Consortium, P. (2020). The repertoire of mutational signatures in human cancer. *Nature,*
748 *578*(7793), 94-101. doi:10.1038/s41586-020-1943-3

749 Alexandrov, L. B., Nik-Zainal, S., Wedge, D. C., Aparicio, S. A., Behjati, S., Biankin, A. V., . . .
750 Stratton, M. R. (2013). Signatures of mutational processes in human cancer. *Nature,*
751 *500*(7463), 415-421. doi:10.1038/nature12477 [pii]

752 Alexandrov, L. B., Nik-Zainal, S., Wedge, D. C., Campbell, P. J., & Stratton, M. R. (2013).
753 Deciphering signatures of mutational processes operative in human cancer. *Cell Rep,*
754 *3*(1), 246-259. doi:10.1016/j.celrep.2012.12.008

755 Angus, L., Smid, M., Wilting, S. M., van Riet, J., Van Hoeck, A., Nguyen, L., . . . Martens, J. W.
756 M. (2019). The genomic landscape of metastatic breast cancer highlights changes in
757 mutation and signature frequencies. *Nat Genet, 51*(10), 1450-1458. doi:10.1038/s41588-
758 019-0507-7

759 Argyris, P. P., Wilkinson, P. E., Jarvis, M. C., Magliocca, K. R., Patel, M. R., Vogel, R. I., . . .
760 Harris, R. S. (2020). Endogenous APOBEC3B overexpression characterizes HPV-positive
761 and HPV-negative oral epithelial dysplasias and head and neck cancers. *Mod Pathol*.
762 doi:10.1038/s41379-020-0617-x

763 Blokzijl, F., Janssen, R., van Boxtel, R., & Cuppen, E. (2018). Mutationalpatterns: Comprehensive
764 genome-wide analysis of mutational processes. *Genome Med, 10*(1), 33.
765 doi:10.1186/s13073-018-0539-0
766 Brown, W. L., Law, E. K., Argyris, P. P., Carpenter, M. A., Levin-Klein, R., Ranum, A. N., . . .
767 Harris, R. S. (2019). A rabbit monoclonal antibody against the antiviral and cancer
768 genomic DNA mutating enzyme APOBEC3B. *Antibodies (Basel), 8*(3).
769 doi:10.3390/antib8030047
770 Buisson, R., Langenbucher, A., Bowen, D., Kwan, E. E., Benes, C. H., Zou, L., & Lawrence, M.
771 S. (2019). Passenger hotspot mutations in cancer driven by APOBEC3A and mesoscale
772 genomic features. *Science, 364*(6447). doi:10.1126/science.aaw2872
773 Burns, M. B., Lackey, L., Carpenter, M. A., Rathore, A., Land, A. M., Leonard, B., . . . Harris, R.
774 S. (2013). APOBEC3B is an enzymatic source of mutation in breast cancer. *Nature,*
775 *494*(7437), 366-370. doi:10.1038/nature11881 [pii]
776 Burns, M. B., Temiz, N. A., & Harris, R. S. (2013). Evidence for APOBEC3B mutagenesis in
777 multiple human cancers. *Nat Genet, 45*(9), 977-983. doi:10.1038/ng.2701 [pii]
778 Cameron, D. L., Schroder, J., Penington, J. S., Do, H., Molania, R., Dobrovic, A., . . . Papenfuss,
779 A. T. (2017). Gridss: Sensitive and specific genomic rearrangement detection using
780 positional de bruijn graph assembly. *Genome Res, 27*(12), 2050-2060.
781 doi:10.1101/gr.222109.117
782 Cannataro, V. L., Gaffney, S. G., Sasaki, T., Issaeva, N., Grewal, N. K. S., Grandis, J. R., . . .
783 Townsend, J. P. (2019). APOBEC-induced mutations and their cancer effect size in head
784 and neck squamous cell carcinoma. *Oncogene, 38*(18), 3475-3487. doi:10.1038/s41388-
785 018-0657-6
786 Carpenter, M. A., Li, M., Rathore, A., Lackey, L., Law, E. K., Land, A. M., . . . Harris, R. S. (2012).
787 Methylcytosine and normal cytosine deamination by the foreign DNA restriction enzyme
788 APOBEC3A. *J Biol Chem, 287*(41), 34801-34808. doi:10.1074/jbc.M112.385161 [pii]

789 Chan, K., Roberts, S. A., Klimczak, L. J., Sterling, J. F., Saini, N., Malc, E. P., . . . Gordenin, D. A.
790 (2015). An APOBEC3A hypermutation signature is distinguishable from the signature of
791 background mutagenesis by APOBEC3B in human cancers. *Nat Genet, 47*(9), 1067-1072.
792 doi:10.1038/ng.3378
793 Chan, K., Sterling, J. F., Roberts, S. A., Bhagwat, A. S., Resnick, M. A., & Gordenin, D. A. (2012).
794 Base damage within single-strand DNA underlies in vivo hypermutability induced by a
795 ubiquitous environmental agent. *PLoS Genet, 8*(12), e1003149.
796 doi:10.1371/journal.pgen.1003149
797 Chen, H., Lilley, C. E., Yu, Q., Lee, D. V., Chou, J., Narvaiza, I., . . . Weitzman, M. D. (2006).
798 APOBEC3A is a potent inhibitor of adeno-associated virus and retrotransposons. *Curr*
799 *Biol, 16*(5), 480-485.
800 Chen, T. W., Lee, C. C., Liu, H., Wu, C. S., Pickering, C. R., Huang, P. J., . . . Chang, Y. S. (2017).
801 APOBEC3A is an oral cancer prognostic biomarker in Taiwanese carriers of an APOBEC
802 deletion polymorphism. *Nat Commun, 8*(1), 465. doi:10.1038/s41467-017-00493-9
803 Chen, Z., Wen, W., Bao, J., Kuhs, K. L., Cai, Q., Long, J., . . . Guo, X. (2019). Integrative genomic
804 analyses of APOBEC-mutational signature, expression and germline deletion of apobec3
805 genes, and immunogenicity in multiple cancer types. *BMC Med Genomics, 12*(1), 131.
806 doi:10.1186/s12920-019-0579-3
807 Chesarino, N. M., & Emerman, M. (2020). Polymorphisms in human APOBEC3H differentially
808 regulate ubiquitination and antiviral activity. *Viruses, 12*(4). doi:10.3390/v12040378
809 Chiang, C., Layer, R. M., Faust, G. G., Lindberg, M. R., Rose, D. B., Garrison, E. P., . . . Hall, I.
810 M. (2015). Speedseq: Ultra-fast personal genome analysis and interpretation. *Nat*
811 *Methods, 12*(10), 966-968. doi:10.1038/nmeth.3505
812 Cho, R. J., Alexandrov, L. B., den Breems, N. Y., Atanasova, V. S., Farshchian, M., Purdom, E.,
813 . . . South, A. P. (2018). APOBEC mutation drives early-onset squamous cell carcinomas

814 in recessive dystrophic epidermolysis bullosa. *Sci Transl Med, 10*(455).
815 doi:10.1126/scitranslmed.aas9668
816 Cortez, L. M., Brown, A. L., Dennis, M. A., Collins, C. D., Brown, A. J., Mitchell, D., . . . Roberts,
817 S. A. (2019). APOBEC3A is a prominent cytidine deaminase in breast cancer. *PLoS*
818 *Genet, 15*(12), e1008545. doi:10.1371/journal.pgen.1008545
819 de Bruin, E. C., McGranahan, N., Mitter, R., Salm, M., Wedge, D. C., Yates, L., . . . Swanton, C.
820 (2014). Spatial and temporal diversity in genomic instability processes defines lung cancer
821 evolution. *Science, 346*(6206), 251-256. doi:10.1126/science.1253462
822 DeWeerd, R. A., Nemeth, E., Poti, A., Petryk, N., Chen, C. L., Hyrien, O., . . . Green, A. M. (2022).
823 Prospectively defined patterns of APOBEC3A mutagenesis are prevalent in human
824 cancers. *Cell Rep, 38*(12), 110555. doi:10.1016/j.celrep.2022.110555
825 Driscoll, C. B., Schuelke, M. R., Kottke, T., Thompson, J. M., Wongthida, P., Tonne, J. M., . . .
826 Vile, R. G. (2020). APOBEC3B-mediated corruption of the tumor cell immunopeptidome
827 induces heteroclitic neoepitopes for cancer immunotherapy. *Nat Commun, 11*(1), 790.
828 doi:10.1038/s41467-020-14568-7
829 Ebrahimi, D., Richards, C. M., Carpenter, M. A., Wang, J., Ikeda, T., Becker, J. T., . . . Harris, R.
830 S. (2018). Genetic and mechanistic basis for APOBEC3H alternative splicing, retrovirus
831 restriction, and counteraction by HIV-1 protease. *Nat Commun, 9*(1), 4137.
832 doi:10.1038/s41467-018-06594-3
833 Faden, D. L., Thomas, S., Cantalupo, P. G., Agrawal, N., Myers, J., & DeRisi, J. (2017). Multi-
834 modality analysis supports APOBEC as a major source of mutations in head and neck
835 squamous cell carcinoma. *Oral Oncol, 74*, 8-14. doi:10.1016/j.oraloncology.2017.09.002
836 Feng, Y., Love, R. P., Ara, A., Baig, T. T., Adolph, M. B., & Chelico, L. (2015). Natural
837 polymorphisms and oligomerization of human APOBEC3H contribute to single-stranded
838 DNA scanning ability. *J Biol Chem, 290*(45), 27188-27203. doi:10.1074/jbc.M115.666065

839 Glaser, A. P., Fantini, D., Wang, Y., Yu, Y., Rimar, K. J., Podojil, J. R., . . . Meeks, J. J. (2018).
840 APOBEC-mediated mutagenesis in urothelial carcinoma is associated with improved
841 survival, mutations in DNA damage response genes, and immune response. *Oncotarget,*
842 *9*(4), 4537-4548. doi:10.18632/oncotarget.23344
843 Green, A. M., Landry, S., Budagyan, K., Avgousti, D. C., Shalhout, S., Bhagwat, A. S., &
844 Weitzman, M. D. (2016). APOBEC3A damages the cellular genome during DNA
845 replication. *Cell Cycle, 15*(7), 998-1008. doi:10.1080/15384101.2016.1152426
846 Gyori, B. M., Venkatachalam, G., Thiagarajan, P. S., Hsu, D., & Clement, M. V. (2014).
847 OPENCOMET: An automated tool for comet assay image analysis. *Redox Biol, 2*, 457-
848 465. doi:10.1016/j.redox.2013.12.020
849 Harris, R. S., & Dudley, J. P. (2015). APOBECs and virus restriction. *Virology, 479-480C*, 131-
850 145. doi:10.1016/j.virol.2015.03.012
851 Harris, R. S., Petersen-Mahrt, S. K., & Neuberger, M. S. (2002). RNA editing enzyme apobec1
852 and some of its homologs can act as DNA mutators. *Mol Cell, 10*(5), 1247-1253.
853 Henderson, S., Chakravarthy, A., Su, X., Boshoff, C., & Fenton, T. R. (2014). Apobec-mediated
854 cytosine deamination links PIK3CA helical domain mutations to human papillomavirus-
855 driven tumor development. *Cell Rep, 7*(6), 1833-1841. doi:10.1016/j.celrep.2014.05.012
856 Hix, M. A., Wong, L., Flath, B., Chelico, L., & Cisneros, G. A. (2020). Single-nucleotide
857 polymorphism of the DNA cytosine deaminase APOBEC3H haplotype I leads to enzyme
858 destabilization and correlates with lung cancer. *NAR Cancer, 2*(3), zcaa023.
859 doi:10.1093/narcan/zcaa023
860 Hoopes, J. I., Cortez, L. M., Mertz, T. M., Malc, E. P., Mieczkowski, P. A., & Roberts, S. A. (2016).
861 APOBEC3A and APOBEC3B preferentially deaminate the lagging strand template during
862 DNA replication. *Cell Rep, 14*(6), 1273-1282. doi:10.1016/j.celrep.2016.01.021
863 Hultquist, J. F., Lengyel, J. A., Refsland, E. W., LaRue, R. S., Lackey, L., Brown, W. L., & Harris,
864 R. S. (2011). Human and rhesus APOBEC3D, APOBEC3F, APOBEC3G, and

865 APOBEC3H demonstrate a conserved capacity to restrict Vif-deficient HIV-1. *J Virol, 85*(21), 11220-11234.
867 Ito, F., Fu, Y., Kao, S. A., Yang, H., & Chen, X. S. (2017). Family-wide comparative analysis of
868 cytidine and methylcytidine deamination by eleven human APOBEC proteins. *J Mol Biol, 429*(12), 1787-1799. doi:10.1016/j.jmb.2017.04.021
870 Ito, F., Yang, H., Xiao, X., Li, S. X., Wolfe, A., Zirkle, B., . . . Chen, X. S. (2018). Understanding
871 the structure, multimerization, subcellular localization and MC selectivity of a genomic
872 mutator and anti-HIV factor APOBEC3H. *Sci Rep, 8*(1), 3763. doi:10.1038/s41598-018-
873 21955-0
874 Jalili, P., Bowen, D., Langenbucher, A., Park, S., Aguirre, K., Corcoran, R. B., . . . Buisson, R.
875 (2020). Quantification of ongoing APOBEC3A activity in tumor cells by monitoring rna
876 editing at hotspots. *Nat Commun, 11*(1), 2971. doi:10.1038/s41467-020-16802-8
877 Jarvis, M. C., Ebrahimi, D., Temiz, N. A., & Harris, R. S. (2018). Mutation signatures including
878 APOBEC in cancer cell lines. *JNCI Cancer Spectr, 2*(1). doi:10.1093/jncics/pky002
879 Koh, G., Degasperi, A., Zou, X., Momen, S., & Nik-Zainal, S. (2021). Mutational signatures:
880 Emerging concepts, caveats and clinical applications. *Nat Rev Cancer, 21*(10), 619-637.
881 doi:10.1038/s41568-021-00377-7
882 Kouno, T., Silvas, T. V., Hilbert, B. J., Shandilya, S. M. D., Bohn, M. F., Kelch, B. A., . . . Schiffer,
883 C. A. (2017). Crystal structure of APOBEC3A bound to single-stranded DNA reveals
884 structural basis for cytidine deamination and specificity. *Nat Commun, 8*, 15024.
885 doi:10.1038/ncomms15024
886 Lackey, L., Demorest, Z. L., Land, A. M., Hultquist, J. F., Brown, W. L., & Harris, R. S. (2012).
887 APOBEC3B and AID have similar nuclear import mechanisms. *J Mol Biol, 419*(5), 301-
888 314. doi:10.1016/j.jmb.2012.03.011

889 Lackey, L., Law, E. K., Brown, W. L., & Harris, R. S. (2013). Subcellular localization of the apobec3
890 proteins during mitosis and implications for genomic DNA deamination. *Cell Cycle, 12*(5),
891 762-772. doi:10.4161/cc.23713 [pii]
892 Landry, S., Narvaiza, I., Linfesty, D. C., & Weitzman, M. D. (2011). APOBEC3A can activate the
893 DNA damage response and cause cell-cycle arrest. *EMBO Rep, 12*(5), 444-450.
894 Langenbucher, A., Bowen, D., Sakhtemani, R., Bournique, E., Wise, J. F., Zou, L., . . . Lawrence,
895 M. S. (2021). An extended APOBEC3A mutation signature in cancer. *Nat Commun, 12*(1),
896 1602. doi:10.1038/s41467-021-21891-0
897 Law, E. K., Levin-Klein, R., Jarvis, M. C., Kim, H., Argyris, P. P., Carpenter, M. A., . . . Harris, R.
898 S. (2020). APOBEC3A catalyzes mutation and drives carcinogenesis *in vivo*. *J Exp Med,*
899 *217*(12). doi:10.1084/jem.20200261
900 Law, E. K., Sieuwerts, A. M., LaPara, K., Leonard, B., Starrett, G. J., Molan, A. M., . . . Harris, R.
901 S. (2016). The DNA cytosine deaminase APOBEC3B promotes tamoxifen resistance in
902 ER-positive breast cancer. *Sci Adv, 2*(10), e1601737. doi:10.1126/sciadv.1601737
903 Leonard, B., Hart, S. N., Burns, M. B., Carpenter, M. A., Temiz, N. A., Rathore, A., . . . Harris, R.
904 S. (2013). APOBEC3B upregulation and genomic mutation patterns in serous ovarian
905 carcinoma. *Cancer Res, 73*(24), 7222-7231. doi:10.1158/0008-5472.CAN-13-1753
906 Leonard, B., McCann, J. L., Starrett, G. J., Kosyakovsky, L., Luengas, E. M., Molan, A. M., . . .
907 Harris, R. S. (2015). The PKC/NF-kappaB signaling pathway induces APOBEC3B
908 expression in multiple human cancers. *Cancer Res, 75*(21), 4538-4547.
909 doi:10.1158/0008-5472.CAN-15-2171-T
910 Maruyama, W., Shirakawa, K., Matsui, H., Matsumoto, T., Yamazaki, H., Sarca, A. D., . . . Takaori-
911 Kondo, A. (2016). Classical NF-kappaB pathway is responsible for APOBEC3B
912 expression in cancer cells. *Biochem Biophys Res Commun, 478*(3), 1466-1471.
913 doi:10.1016/j.bbrc.2016.08.148

914 Mayorov, V. I., Rogozin, I. B., Adkison, L. R., Frahm, C., Kunkel, T. A., & Pavlov, Y. I. (2005).
915 Expression of human aid in yeast induces mutations in context similar to the context of
916 somatic hypermutation at G-C pairs in immunoglobulin genes. *BMC Immunol, 6*(1), 10.
917 McCann, J. L., Cristini, A., Law, E. K., Lee, S. Y., Tellier, M., Carpenter, M. A., . . . Harris, R. S.
918 (2021). R-loop homeostasis and cancer mutagenesis promoted by the DNA cytosine
919 deaminase APOBEC3B. *BioRxiv*.
920 McNamara, R. P., McCann, J. L., Gudipaty, S. A., & D'Orso, I. (2013). Transcription factors
921 mediate the enzymatic disassembly of promoter-bound 7SK SNRNP to locally recruit p-
922 tefb for transcription elongation. *Cell Rep, 5*(5), 1256-1268.
923 doi:10.1016/j.celrep.2013.11.003
924 Middlebrooks, C. D., Banday, A. R., Matsuda, K., Udquim, K. I., Onabajo, O. O., Paquin, A., . . .
925 Prokunina-Olsson, L. (2016). Association of germline variants in the APOBEC3 region
926 with cancer risk and enrichment with apobec-signature mutations in tumors. *Nat Genet,*
927 *48*(11), 1330-1338. doi:10.1038/ng.3670
928 Moore, L., Cagan, A., Coorens, T. H. H., Neville, M. D. C., Sanghvi, R., Sanders, M. A., . . .
929 Rahbari, R. (2021). The mutational landscape of human somatic and germline cells.
930 *Nature, 597*(7876), 381-386. doi:10.1038/s41586-021-03822-7
931 Mussil, B., Suspene, R., Aynaud, M. M., Gauvrit, A., Vartanian, J. P., & Wain-Hobson, S. (2013).
932 Human APOBEC3A isoforms translocate to the nucleus and induce DNA double strand
933 breaks leading to cell stress and death. *PLoS One, 8*(8), e73641.
934 doi:10.1371/journal.pone.0073641
935 Nik-Zainal, S., Alexandrov, L. B., Wedge, D. C., Van Loo, P., Greenman, C. D., Raine, K., . . .
936 Stratton, M. R. (2012). Mutational processes molding the genomes of 21 breast cancers.
937 *Cell, 149*(5), 979-993. doi:S0092-8674(12)00528-4 [pii]

938 Nik-Zainal, S., Davies, H., Staaf, J., Ramakrishna, M., Glodzik, D., Zou, X., . . . Stratton, M. R.
939 (2016). Landscape of somatic mutations in 560 breast cancer whole-genome sequences.
940 *Nature, 534*(7605), 47-54. doi:10.1038/nature17676
941 Nik-Zainal, S., Wedge, D. C., Alexandrov, L. B., Petljak, M., Butler, A. P., Bolli, N., . . . Stratton,
942 M. R. (2014). Association of a germline copy number polymorphism of APOBEC3A and
943 APOBEC3B with burden of putative APOBEC-dependent mutations in breast cancer. *Nat*
944 *Genet, 46*(5), 487-491. doi:10.1038/ng.2955
945 Nikkila, J., Kumar, R., Campbell, J., Brandsma, I., Pemberton, H. N., Wallberg, F., . . . Lord, C. J.
946 (2017). Elevated APOBEC3B expression drives a kataegic-like mutation signature and
947 replication stress-related therapeutic vulnerabilities in p53-defective cells. *Br J Cancer,*
948 *117*(1), 113-123. doi:10.1038/bjc.2017.133
949 Oh, S., Bournique, E., Bowen, D., Jalili, P., Sanchez, A., Ward, I., . . . Buisson, R. (2021).
950 Genotoxic stress and viral infection induce transient expression of APOBEC3A and pro-
951 inflammatory genes through two distinct pathways. *Nat Commun, 12*(1), 4917.
952 doi:10.1038/s41467-021-25203-4
953 Periyasamy, M., Patel, H., Lai, C. F., Nguyen, V. T., Nevedomskaya, E., Harrod, A., . . . Ali, S.
954 (2015). APOBEC3B-mediated cytidine deamination is required for estrogen receptor
955 action in breast cancer. *Cell Rep, 13*(1), 108-121. doi:10.1016/j.celrep.2015.08.066
956 Periyasamy, M., Singh, A. K., Gemma, C., Kranjec, C., Farzan, R., Leach, D. A., . . . Ali, S. (2017).
957 P53 controls expression of the DNA deaminase APOBEC3B to limit its potential mutagenic
958 activity in cancer cells. *Nucleic Acids Res, 45*(19), 11056-11069. doi:10.1093/nar/gkx721
959 Petersen-Mahrt, S. K., Harris, R. S., & Neuberger, M. S. (2002). AID mutates *E. coli* suggesting
960 a DNA deamination mechanism for antibody diversification. *Nature, 418*(6893), 99-103.
961 Petljak, M., Alexandrov, L. B., Brammeld, J. S., Price, S., Wedge, D. C., Grossmann, S., . . .
962 Stratton, M. R. (2019). Characterizing mutational signatures in human cancer cell lines

963 reveals episodic APOBEC mutagenesis. *Cell, 176*(6), 1282-1294 e1220.
964 doi:10.1016/j.cell.2019.02.012
965 Petljak, M., Chu, K., Dananberg, A., Bergstrom, E. N., von Morgen, P., Alexandrov, L. B., . . .
966 Maciejowski, J. (2021). The APOBEC3A deaminase drives episodic mutagenesis in
967 cancer cells. *BioRxiv,* doi: https://doi.org/10.1101/2021.02.14.431145.
968 Petljak, M., & Maciejowski, J. (2020). Molecular origins of APOBEC-associated mutations in
969 cancer. *DNA Repair (Amst), 94*, 102905. doi:10.1016/j.dnarep.2020.102905
970 Refsland, E. W., Hultquist, J. F., Luengas, E. M., Ikeda, T., Shaban, N. M., Law, E. K., . . . Harris,
971 R. S. (2014). Natural polymorphisms in human APOBEC3H and HIV-1 Vif combine in
972 primary T lymphocytes to affect viral G-to-A mutation levels and infectivity. *PLoS Genet,*
973 *10*(11), e1004761. doi:10.1371/journal.pgen.1004761
974 Refsland, E. W., Stenglein, M. D., Shindo, K., Albin, J. S., Brown, W. L., & Harris, R. S. (2010).
975 Quantitative profiling of the full *APOBEC3* mRNA repertoire in lymphocytes and tissues:
976 Implications for hiv-1 restriction. *Nucleic Acids Res, 38*, 4274-4284.
977 Roberts, S. A., Lawrence, M. S., Klimczak, L. J., Grimm, S. A., Fargo, D., Stojanov, P., . . .
978 Gordenin, D. A. (2013). An APOBEC cytidine deaminase mutagenesis pattern is
979 widespread in human cancers. *Nat Genet, 45*(9), 970-976. doi:10.1038/ng.2702
980 Roberts, S. A., Sterling, J., Thompson, C., Harris, S., Mav, D., Shah, R., . . . Gordenin, D. A.
981 (2012). Clustered mutations in yeast and in human cancers can arise from damaged long
982 single-strand DNA regions. *Mol Cell, 46*(4), 424-435. doi:10.1016/j.molcel.2012.03.030
983 Roelofs, P. A., Goh, C. Y., Chua, B. H., Jarvis, M. C., Stewart, T. A., McCann, J. L., . . . Harris, R.
984 S. (2020). Characterization of the mechanism by which the RB/E2F pathway controls
985 expression of the cancer genomic DNA deaminase APOBEC3B. *Elife, 9*.
986 doi:10.7554/eLife.61287
987 Rosenthal, R., McGranahan, N., Herrero, J., Taylor, B. S., & Swanton, C. (2016). Deconstructsigs:
988 Delineating mutational processes in single tumors distinguishes DNA repair deficiencies

989 and patterns of carcinoma evolution. *Genome Biol, 17*, 31. doi:10.1186/s13059-016-0893-
990 4
991 Salamango, D. J., McCann, J. L., Demir, O., Brown, W. L., Amaro, R. E., & Harris, R. S. (2018).
992 APOBEC3B nuclear localization requires two distinct N-terminal domain surfaces. *J Mol*
993 *Biol, 430*(17), 2695-2708. doi:10.1016/j.jmb.2018.04.044
994 Schumacher, A. J., Nissley, D. V., & Harris, R. S. (2005). APOBEC3G hypermutates genomic
995 DNA and inhibits TY1 retrotransposition in yeast. *Proc Natl Acad Sci U S A, 102*(28), 9854-
996 9859.
997 Schutsky, E. K., Nabel, C. S., Davis, A. K. F., DeNizio, J. E., & Kohli, R. M. (2017). APOBEC3A
998 efficiently deaminates methylated, but not tet-oxidized, cytosine bases in DNA. *Nucleic*
999 *Acids Res, 45*(13), 7655-7665. doi:10.1093/nar/gkx345
1000 Shaban, N. M., Shi, K., Lauer, K. V., Carpenter, M. A., Richards, C. M., Salamango, D., . . . Harris,
1001 R. S. (2018). The antiviral and cancer genomic DNA deaminase APOBEC3H is regulated
1002 by an RNA-mediated dimerization mechanism. *Mol Cell, 69*(1), 75-86 e79.
1003 doi:10.1016/j.molcel.2017.12.010
1004 Shaban, N. M., Yan, R., Shi, K., Moraes, S. N., Cheng, A. Z., Carpenter, M. A., . . . Harris, R. S.
1005 (2022). Cryo-em structure of the EBV ribonucleotide reductase BORF2 and mechanism
1006 of APOBEC3B inhibition. *Science Advances, In press*.
1007 Sharma, S., Patnaik, S. K., Taggart, R. T., Kannisto, E. D., Enriquez, S. M., Gollnick, P., & Baysal,
1008 B. E. (2015). APOBEC3A cytidine deaminase induces RNA editing in monocytes and
1009 macrophages. *Nat Commun, 6*, 6881. doi:10.1038/ncomms7881
1010 Shi, K., Carpenter, M. A., Banerjee, S., Shaban, N. M., Kurahashi, K., Salamango, D. J., . . .
1011 Aihara, H. (2017). Structural basis for targeted DNA cytosine deamination and
1012 mutagenesis by APOBEC3A and APOBEC3B. *Nat Struct Mol Biol, 24*(2), 131-139.
1013 doi:10.1038/nsmb.3344

1014 Shi, K., Carpenter, M. A., Kurahashi, K., Harris, R. S., & Aihara, H. (2015). Crystal structure of
1015 the DNA deaminase APOBEC3B catalytic domain. *J Biol Chem, 290*(47), 28120-28130.
1016 doi:10.1074/jbc.M115.679951
1017 Sieuwerts, A. M., Schrijver, W. A., Dalm, S. U., de Weerd, V., Moelans, C. B., Ter Hoeve, N., . . .
1018 van Deurzen, C. H. (2017). Progressive APOBEC3B mRNA expression in distant breast
1019 cancer metastases. *PLoS One, 12*(1), e0171343. doi:10.1371/journal.pone.0171343
1020 Sieuwerts, A. M., Willis, S., Burns, M. B., Look, M. P., Meijer-Van Gelder, M. E., Schlicker, A., . .
1021 . Martens, J. W. (2014). Elevated APOBEC3B correlates with poor outcomes for estrogen-
1022 receptor-positive breast cancers. *Horm Cancer, 5*(6), 405-413. doi:10.1007/s12672-014-
1023 0196-8
1024 Silvas, T. V., & Schiffer, C. A. (2019). APOBEC3s: DNA-editing human cytidine deaminases.
1025 *Protein Sci, 28*(9), 1552-1566. doi:10.1002/pro.3670
1026 Simon, V., Bloch, N., & Landau, N. R. (2015). Intrinsic host restrictions to HIV-1 and mechanisms
1027 of viral escape. *Nat Immunol, 16*(6), 546-553. doi:10.1038/ni.3156
1028 Starrett, G. J., Luengas, E. M., McCann, J. L., Ebrahimi, D., Temiz, N. A., Love, R. P., . . . Harris,
1029 R. S. (2016). The DNA cytosine deaminase APOBEC3H haplotype I likely contributes to
1030 breast and lung cancer mutagenesis. *Nat Commun, 7*, 12918. doi:10.1038/ncomms12918
1031 Starrett, G. J., Serebrenik, A. A., Roelofs, P. A., McCann, J. L., Verhalen, B., Jarvis, M. C., . . .
1032 Harris, R. S. (2019). Polyomavirus T antigen induces APOBEC3B expression using an
1033 LXCXE-dependent and TP53-independent mechanism. *MBio, 10*(1).
1034 doi:10.1128/mBio.02690-18
1035 Stenglein, M. D., Burns, M. B., Li, M., Lengyel, J., & Harris, R. S. (2010). APOBEC3 proteins
1036 mediate the clearance of foreign DNA from human cells. *Nat Struct Mol Biol, 17*(2), 222-
1037 229.

1038 Suspène, R., Aynaud, M. M., Vartanian, J. P., & Wain-Hobson, S. (2013). Efficient deamination
1039 of 5-methylcytidine and 5-substituted cytidine residues in DNA by human APOBEC3A
1040 cytidine deaminase. *PLoS One, 8*(6), e63461. doi:10.1371/journal.pone.0063461 [pii]
1041 Taylor, B. J., Nik-Zainal, S., Wu, Y. L., Stebbings, L. A., Raine, K., Campbell, P. J., . . . Neuberger,
1042 M. S. (2013). DNA deaminases induce break-associated mutation showers with
1043 implication of APOBEC3B and 3A in breast cancer kataegis. *Elife, 2*, e00534.
1044 doi:10.7554/eLife.00534 [pii]
1045 Tsuboi, M., Yamane, A., Horiguchi, J., Yokobori, T., Kawabata-Iwakawa, R., Yoshiyama, S., . . .
1046 Nishiyama, M. (2016). APOBEC3B high expression status is associated with aggressive
1047 phenotype in japanese breast cancers. *Breast Cancer, 23*(5), 780-788.
1048 doi:10.1007/s12282-015-0641-8
1049 Uphoff, C. C., & Drexler, H. G. (2011). Detecting mycoplasma contamination in cell cultures by
1050 polymerase chain reaction. *Methods Mol Biol, 731*, 93-103. doi:10.1007/978-1-61779-
1051 080-5_8
1052 Venkatesan, S., Rosenthal, R., Kanu, N., McGranahan, N., Bartek, J., Quezada, S. A., . . .
1053 Swanton, C. (2018). Perspective: APOBEC mutagenesis in drug resistance and immune
1054 escape in HIV and cancer evolution. *Ann Oncol, 29*(3), 563-572.
1055 doi:10.1093/annonc/mdy003
1056 Verhalen, B., Starrett, G. J., Harris, R. S., & Jiang, M. (2016). Functional upregulation of the DNA
1057 cytosine deaminase APOBEC3B by polyomaviruses. *J Virol, 90*(14), 6379-6386.
1058 doi:10.1128/JVI.00771-16
1059 Vieira, V. C., Leonard, B., White, E. A., Starrett, G. J., Temiz, N. A., Lorenz, L. D., . . . Harris, R.
1060 S. (2014). Human papillomavirus E6 triggers upregulation of the antiviral and cancer
1061 genomic DNA deaminase APOBEC3B. *MBio, 5*(6). doi:10.1128/mBio.02234-14

1062 Walker, B. A., Wardell, C. P., Murison, A., Boyle, E. M., Begum, D. B., Dahir, N. M., . . . Morgan,
1063 G. J. (2015). APOBEC family mutational signatures are associated with poor prognosis
1064 translocations in multiple myeloma. *Nat Commun, 6*, 6997. doi:10.1038/ncomms7997
1065 Wang, J., Becker, J. T., Shi, K., Lauer, K. V., Salamango, D. J., Aihara, H., . . . Harris, R. S.
1066 (2019). The role of RNA in HIV-1 Vif-mediated degradation of APOBEC3H. *J Mol Biol,*
1067 *431*(24), 5019-5031. doi:10.1016/j.jmb.2019.09.014
1068 Warren, C. J., Westrich, J. A., Doorslaer, K. V., & Pyeon, D. (2017). Roles of APOBEC3A and
1069 APOBEC3B in human papillomavirus infection and disease progression. *Viruses, 9*(8).
1070 doi:10.3390/v9080233
1071 Warren, C. J., Xu, T., Guo, K., Griffin, L. M., Westrich, J. A., Lee, D., . . . Pyeon, D. (2015).
1072 APOBEC3A functions as a restriction factor of human papillomavirus. *J Virol, 89*(1), 688-
1073 702. doi:10.1128/JVI.02383-14
1074 Westrich, J. A., Warren, C. J., Klausner, M. J., Guo, K., Liu, C. W., Santiago, M. L., & Pyeon, D.
1075 (2018). Human papillomavirus 16 E7 stabilizes APOBEC3A protein by inhibiting CULLIN
1076 2-dependent protein degradation. *J Virol, 92*(7). doi:10.1128/JVI.01318-17
1077 Yamazaki, H., Shirakawa, K., Matsumoto, T., Hirabayashi, S., Murakawa, Y., Kobayashi, M., . . .
1078 Takaori-Kondo, A. (2019). Endogenous APOBEC3B overexpression constitutively
1079 generates DNA substitutions and deletions in myeloma cells. *Sci Rep, 9*(1), 7122.
1080 doi:10.1038/s41598-019-43575-y
1081 Yamazaki, H., Shirakawa, K., Matsumoto, T., Kazuma, Y., Matsui, H., Horisawa, Y., . . . Takaori-
1082 Kondo, A. (2020). APOBEC3B reporter myeloma cell lines identify DNA damage response
1083 pathways leading to APOBEC3B expression. *PLoS One, 15*(1), e0223463.
1084 doi:10.1371/journal.pone.0223463
1085 Zou, X., Koh, G. C. C., Nanda, A. S., Degasperi, A., Urgo, K., Roumeliotis, T. I., . . . Nik-Zainal,
1086 S. (2021). A systematic CRISPR screen defines mutational mechanisms underpinning

1087 signatures caused by replication errors and endogenous DNA damage. *Nat Cancer, 2*(6),
1088 643-657. doi:10.1038/s43018-021-00200-0
1089 Zou, X., Owusu, M., Harris, R., Jackson, S. P., Loizou, J. I., & Nik-Zainal, S. (2018). Validating
1090 the concept of mutational signatures with isogenic cell models. *Nat Commun, 9*(1), 1744.
1091 doi:10.1038/s41467-018-04052-8
1092

1093 **Main and Supplementary Figure Legends**

1094

1095 **Figure 1. A3 activity in the HAP1-TK-M9 mutation reporter system.**

1096 (**A**) Schematic of the construction of the HAP1-TK-M9 system and overall experimental workflow

1097 (see **Figure 1 – Figure Supplement 1** for representative HAP1-TK-M9 karyotype images).

1098 (**B**) *A3* mRNA levels in parental HAP1 cells in comparison to the HAP1-TK-M9 daughter clone by

1099 RT-qPCR (*** = $p < 0.001$, NS = non-significant, Welsh's t-test).

1100 (**C-E**) Immunoblots of A3A, A3B, and A3H following transduction of HAP1-TK-M9 cells. Tubulin

1101 (Tub) is a loading control (see **Figure 1 – Figure Supplement 2** for anti-A3A mAb validation).

1102 The lower images show ssDNA deaminase activity of extracts from the same cells (S, substrate;

1103 P, product).

1104 (**F**) Immunofluorescent microscopy images of HAP1-TK-M9 cells expressing the indicated A3

1105 enzymes (see methods for primary antibodies). A3 staining is green, γ-H2AX staining is red, and

1106 nuclei are blue from Hoechst. A3A is cell wide, A3B is predominantly nuclear, and A3H-I and -II

1107 are cell wide with nucleolar concentrations (scare bar = 25 μm). The lower row shows

1108 representative COMET images visualized with SYBR Gold (1:10,000) in the GFP channel.

1109 (**G-H**) Quantification of γ-H2AX staining and COMET moments in experiments shown in panel F

1110 (each data point represents an independent cell; p-values determined with Welch's t-test).

1111

1112 **Figure 1 – Supplementary Figure 1. HAP1-TK-M9 karyotype analysis.**

1113 Representative G-band and spectral karyotype (SKY) images of HAP1-TK-M9 M-phase

1114 chromosomes showing a near diploid DNA content and previously reported aberrations including

1115 the reciprocal chromosome 9:22 translocation (Philadelphia chromosome) characteristic of CML

1116 tumor cells. The Y-chromosome is missing, as reported previously in the KBM7 parent line of

1117 HAP1.

1118

1119 **Figure 1 – Supplementary Figure 2. Validation of a custom rabbit-anti human APOBEC3A**
1120 **mAb UMN-13.**
1121 (**A**) Schematic of human A3A, A3B, and A3G indicating the unique N-terminal epitope used here
1122 to generate an A3A-specific mAb UMN-13. The schematic also shows the C-terminal epitope
1123 used previously to generate the versatile 5210-87-13 mAb that recognizes these three enzymes.
1124 (**B**) Comparative immunoblots of whole cell extracts from 293T cells expressing each of the 7
1125 human A3 family members and probed with either our custom rabbit anti-human A3A mAb UMN-
1126 13 (left) or a commercial anti-HA mAb as an expression control (right).
1127 (**C**) Comparative immunoblots of whole cell extracts from the monocytic cell line THP-1 and a
1128 clonal derivative lacking A3A-through-A3G (ΔA-G) plus/minus treatment with LPS/IFN-α to induce
1129 expression of multiple A3s including A3A and A3G. The UMN-13 mAb blot on the left shows a
1130 single band representing full-length A3A (starting at Met1), and the 5210-87-13 mAb blot on the
1131 right shows A3G (strong top band), A3B (weak band just below A3G), and both A3A translation
1132 products (strong faster-migrating bands for full-length A3A starting at Met1 and the shorter isoform
1133 starting at Met13).
1134
1135 **Figure 2. Characterization of *TK* mutations in ganciclovir-resistant clones.**
1136 (**A**) A dot plot of Gan^R colonies generated under the indicated A3 expression conditions. Each
1137 data point represents the number of Gan^R mutants in a single clonal culture (mean +/- SD shown
1138 with significance assessed using Welch's t-test).
1139 (**B**) Schematics representing all *TK* mutations observed under the indicated A3 expression
1140 conditions (APOBEC3 signature T(C>T/G)W mutations in red, other SBSs in black, and InDels in
1141 blue; see **Figure 2 – Figure Supplement 1** for schematics of individual *TK* mutants for these and
1142 vector control conditions). Q8X and R212K mutation hotspots are labeled.
1143 (**C-D**) Pentanucleotide logos depicting -2 and +2 sequence preferences flanking all T(C>T/G)W
1144 mutations observed under the indicated conditions.

1145

1146 **Figure 2 – Supplementary Figure 1. Mutations in TK genes derived from ganciclovir-**
1147 **resistant clones.**

1148 Schematics of TK mutation frequencies in ganciclovir-resistant clones per individual clone.
1149 T[C>G/T]W mutations are shown in red, and other SBSs are shown in black. InDels are shown in
1150 blue. Composite sequence mutation schematics are shown below each list of sequences per
1151 condition.

1152

1153 **Figure 3. Single base substitution mutation signatures in ganciclovir-resistant clones by**
1154 **whole genome sequencing.**

1155 (**A**) Trinucleotide profiles of pooled SBSs across all clones sequenced for each listed experimental
1156 condition (A3A WT, n=6; A3A E72A, n=2; A3B WT, n=5; A3B E255A, n=2. APOBEC3 signature
1157 T(C>T/G)W mutations are highlighted by red dashed-line boxes. See **Figure 3 – Figure**
1158 **Supplement 1 and 2** for mRNA and protein level expression confirmation, respectively, and
1159 **Figure 3 – Figure Supplement 3** for SBS profiles from each WGS).

1160 (**B-C**) Pentanucleotide logos depicting -2 and +2 sequence preferences flanking all T(C>T/G)W
1161 mutations attributable to A3A and A3B, respectively.

1162

1163 **Figure 3 – Supplementary Figure 1. *A3* mRNA expression in the HAP1-TK-M9 system.**

1164 (**A**) *A3A*, *A3B*, and *A3H* mRNA expression levels relative to those of the housekeeping gene *TBP*
1165 for the indicated HAP1-TK-M9 conditions (RNA-seq FKPM from $n \geq 2$ Gan^R clones for each
1166 condition; mean +/- SD shown). RNA-seq data from A3 signature-high breast cancer cell lines
1167 (BT-474 and MDA-MB-453), CCLE breast-derived cell lines (n=52), and TGCA primary breast
1168 cancers (n=1093) are presented alongside for comparison (mean +/- SD).

1169 (**B**) A heatmap depicting mean expression levels of all 7 human *APOBEC3* family members, in
1170 addition to *APOBEC1*, *APOBEC2*, *APOBEC3*, and *AICDA*, relative to those of the housekeeping

1171 gene *TBP* (RNA-seq values are FKPM; n≥2 for each condition to provide matching data sets for
1172 the Gan^R clones subjected to WGS). Endogenous *A3C* provides a robust internal control.
1173
1174 **Figure 3 – Supplementary Figure 2. A3 protein expression in Gan^R clones.**
1175 (**A**) Immunoblots of A3A, A3B, and A3H in the indicated Gan^R clones. Tubulin (TUB) is a loading
1176 control.
1177 (**B**) Deaminase activity of WCE on ssDNA from the same clones (S, substrate; P, product).
1178
1179 **Figure 3 – Supplementary Figure 3. Single base substitution mutation signatures in**
1180 **individual ganciclovir-resistant clones by whole genome sequencing.**
1181 Trinucleotide profiles of the single base substitution mutations found in WGSs from the indicated
1182 Gan^R clones. Composite profiles for each condition are shown in **Figure 3**.
1183
1184 **Figure 3 – Supplementary Figure 4. APOBEC3 mutation signature extracted from WGSs**
1185 **using complementary methods.**
1186 (**A**) Mutation signature profiles extracted from Gan^R clone WGS using an NMF-based approach.
1187 For each clone, all SBSs resolved into 2 signatures - an APOBEC-like signature shown in red and
1188 a background signature in blue. APOBEC3 mutation enrichment scores are shown above each
1189 bar, with significantly enriched values shown in red (BH FDR corrected q-value < 0.05).
1190 (**B**) Trinucleotide profiles of the two signatures derived from NMF in panel A clearly showing
1191 APOBEC3 signature T(C>T/G)W mutations.
1192
1193 **Figure 4. Mesoscale properties of A3A and A3B *in vitro* and in the HAP1-TK-M9 system.**
1194 (**A-B**) DNA deamination kinetics of A3A and A3B using *SDHB* and *NUP93* hairpin substrates in
1195 comparison to corresponding linear controls made by scrambling the 5' or 3' portion of the stem,
1196 respectively. See text for full description and **Figure 4 – Figure Supplement 1** for a genome-

1197 wide analysis.

1198 (**C**) Rainfall plots of genome-wide intermutation distances showing APOBEC3 signature kataegic

1199 tracts (red arrows) in representative A3A- and A3B-expressing Gan^R clones (C>T mutations are

1200 red, C>G black, C>A blue, T>A purple, T>C yellow, and T>G green).

1201

1202 **Figure 4 – Figure Supplement 1. Hairpin mutation analysis in HAP1-TK-M9 WGS data.**

1203 Pie charts representing the fraction of APOBEC3 signature mutations and other types of single

1204 base substitution mutations found in predicted hairpin substrates (red) versus non-hairpin

1205 contexts (grey). The total number of APOBEC3 signature T(C>T/G)W mutations is shown in the

1206 center of each pie chart. Hairpin stems and loops were defined liberally as any feature between

1207 3 and 100 nucleotides, and a hairpin is considered mutated if it has a SBS mutation in any position

1208 of the loop region.

1209

1210 **Figure 5. A composite origin of APOBEC3 signature mutations in breast cancer.**

1211 A Euclidean distance analysis of similarity between the SBS profiles from WGSs of the A3A, A3B,

1212 and control Gan^R clones described here versus primary breast tumor whole-genome sequencing

1213 data (ICGC, n=794). The APOBEC3 signature is represented by both enrichment score and

1214 SBS2+13 (red), HRD signature as SBS3 (blue), and ageing signature as SBS1 (gray).

1215

Figure 1

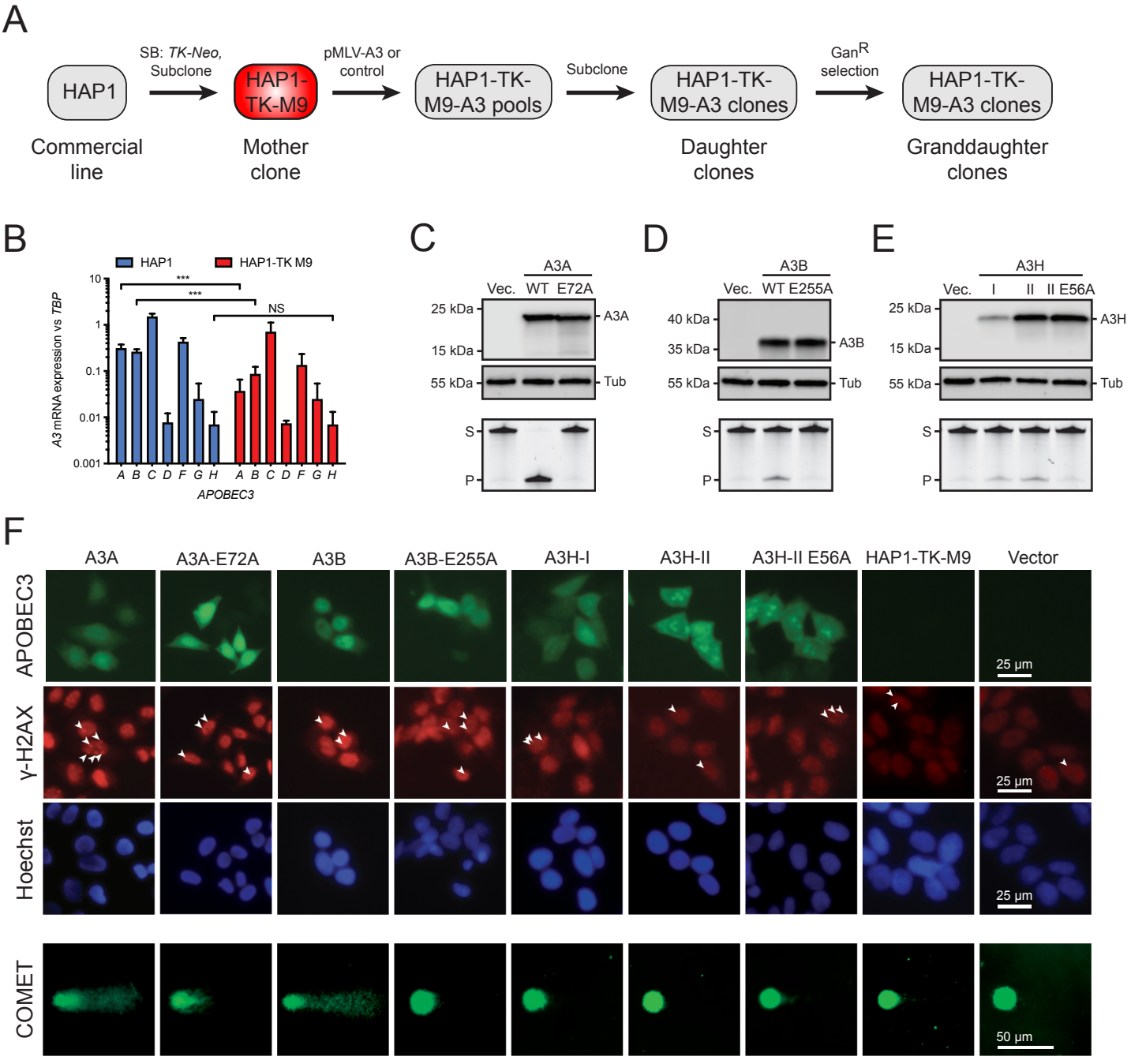

A
HAP1
SB: TK-Neo, Subclone
HAP1-TK-M9
pMLV-A3 or control
HAP1-TK-M9-A3 pools
Subclone
HAP1-TK-M9-A3 clones
GanR selection
HAP1-TK-M9-A3 clones
Commercial line
Mother clone
Daughter clones
Granddaughter clones
B
HAP1
HAP1-TK M9
A3 mRNA expression vs TBP
APOBEC3
C
A3A
Vec. WT E72A
25 kDa
15 kDa
55 kDa
Tub
S
P
D
A3B
Vec. WT E255A
40 kDa
35 kDa
E
A3H
Vec. I II II E56A
F
A3A
A3A-E72A
A3B
A3B-E255A
A3H-I
A3H-II
A3H-II E56A
HAP1-TK-M9
Vector
APOBEC3
γ-H2AX
Hoechst
COMET
25 µm
50 µm
G
γ-H2AX foci per cell
p < 0.001
NS
WT E72A WT E255A I II E56A Vector
A3A A3B A3H HAP1-TK-M9
H
Percent tail DNA (%)
p < 0.05
p < 0.001
H2O2

Figure 1 - Supp 1

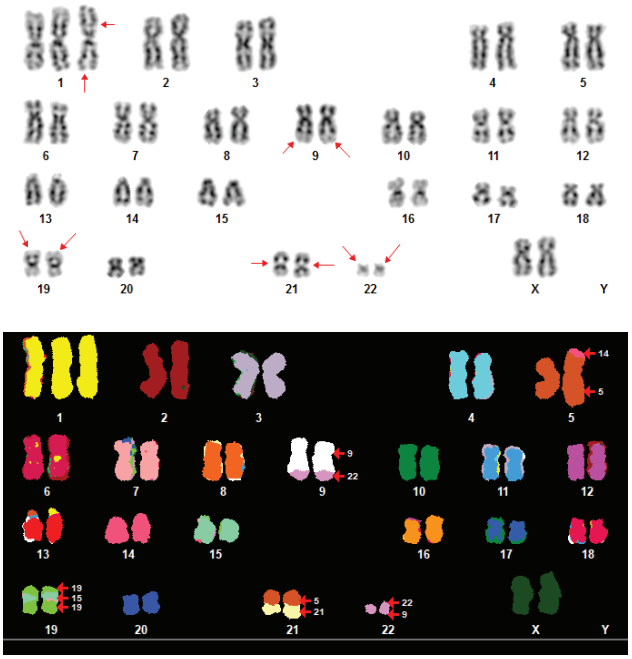


Figure 1 - Supp 2

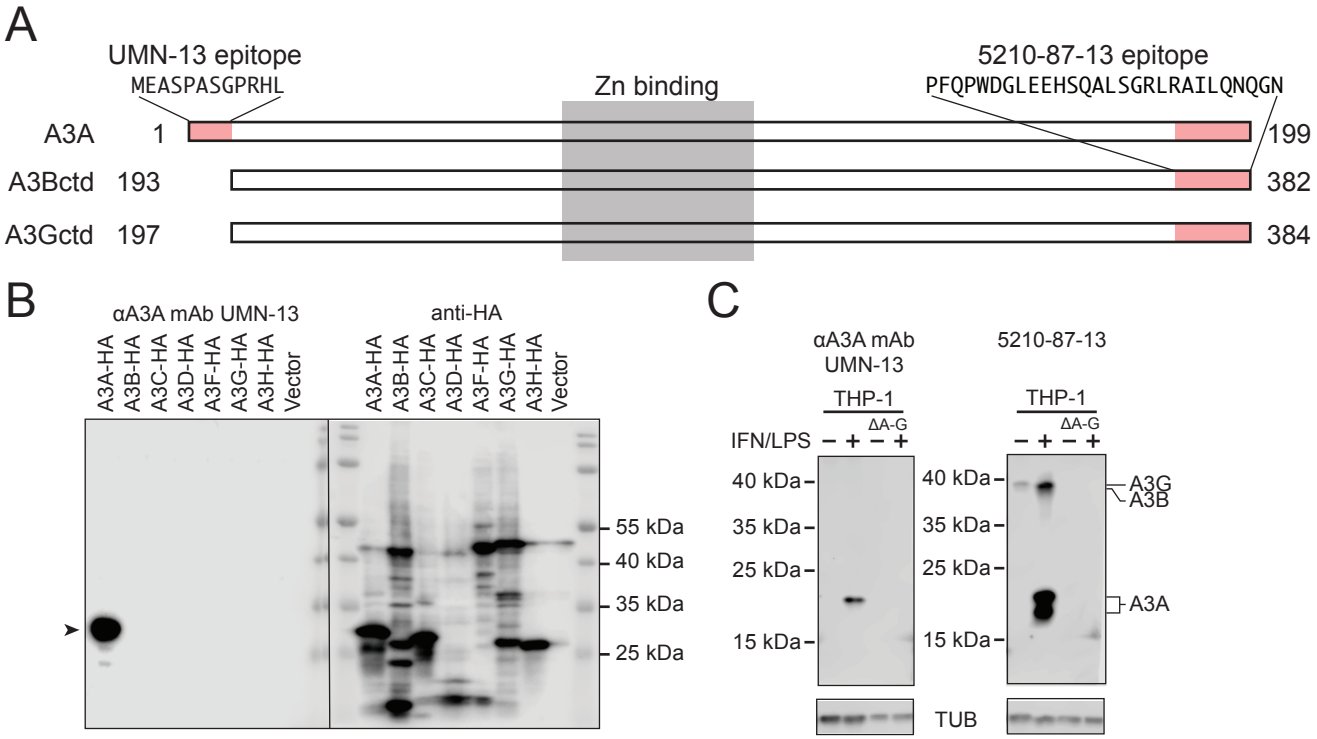


Figure 2

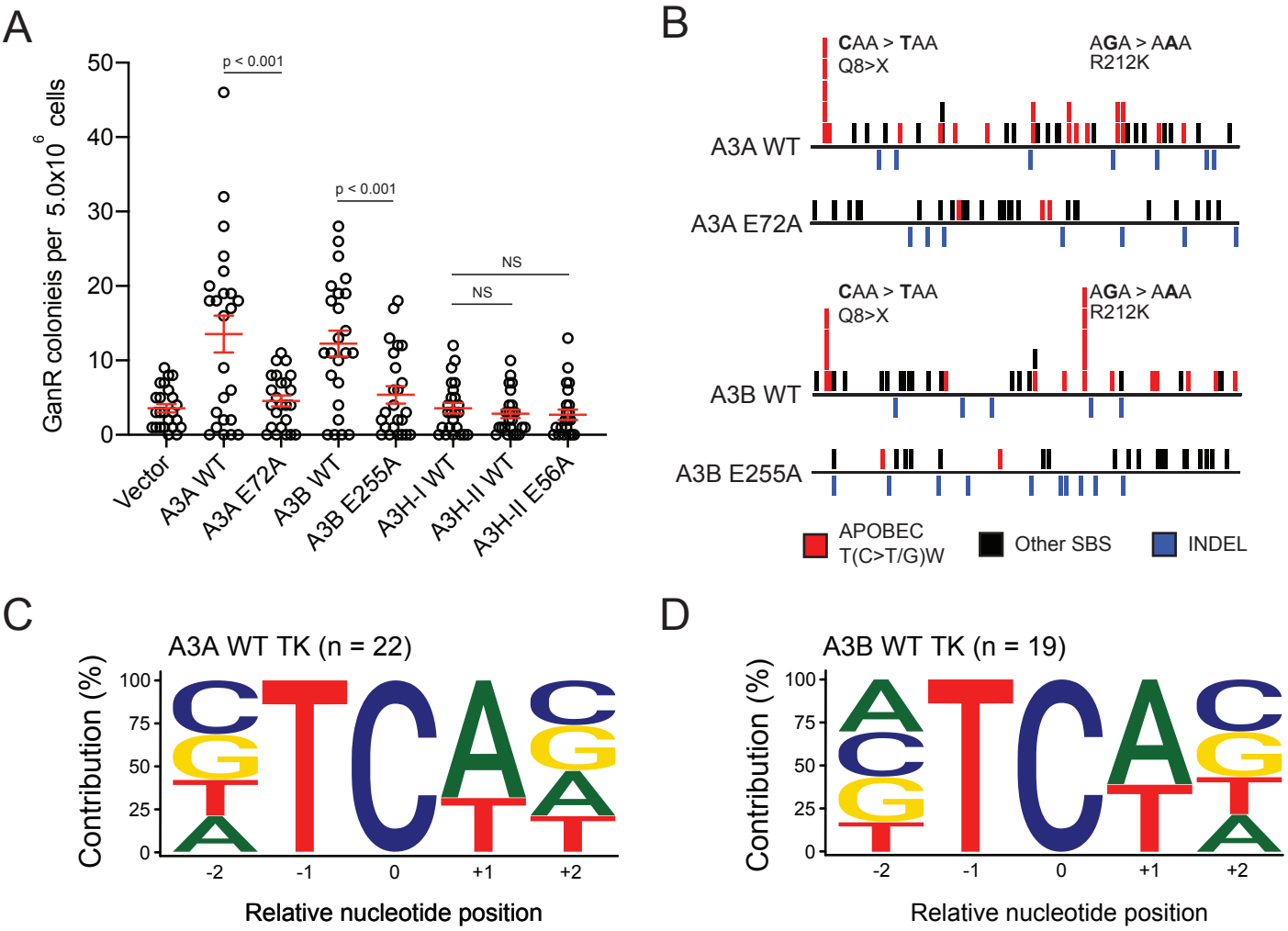
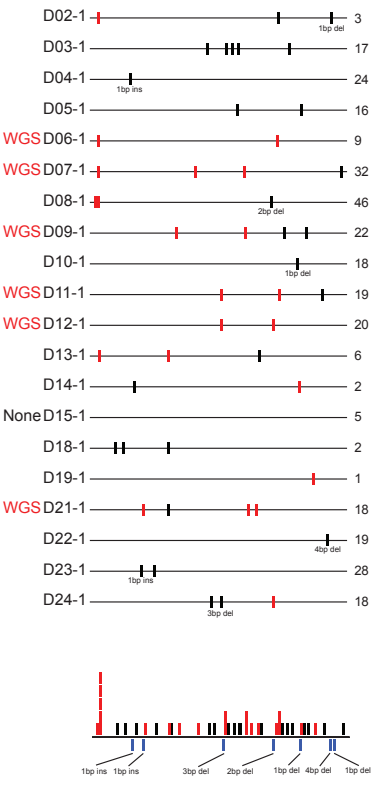
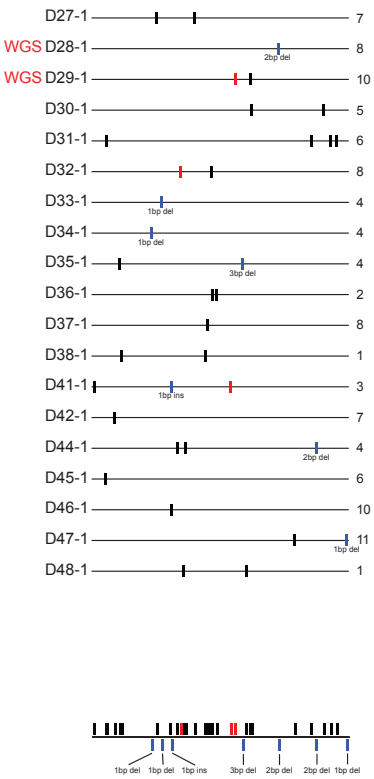

A
GanR colonieis per 5.0x10^6 cells
50
40
30
20
10
0
p < 0.001
p < 0.001
NS
NS
Vector
A3A WT
A3A E72A
A3B WT
A3B E255A
A3H-I WT
A3H-II WT
A3H-II E56A
B
CAA > TAA
Q8>X
AGA > AAA
R212K
A3A WT
A3A E72A
CAA > TAA
Q8>X
AGA > AAA
R212K
A3B WT
A3B E255A
APOBEC
T(C>T/G)W
Other SBS
INDEL
C
A3A WT TK (n = 22)
Contribution (%)
100
75
50
25
0
-2
-1
0
+1
+2
Relative nucleotide position
D
A3B WT TK (n = 19)
Contribution (%)
100
75
50
25
0
-2
-1
0
+1
+2
Relative nucleotide position

Figure 2 - Supp 1

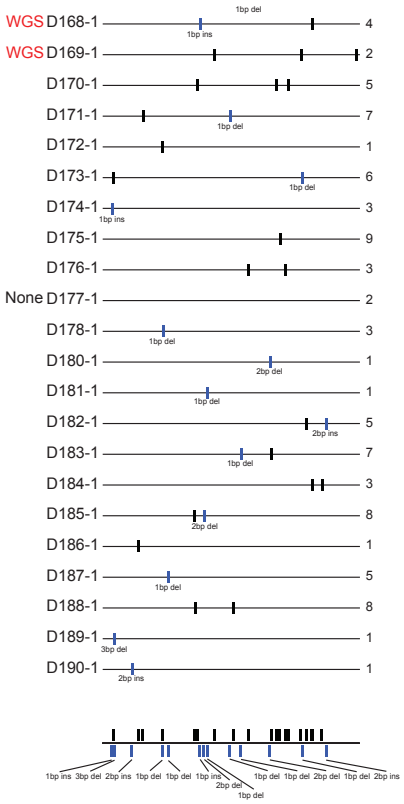
A3A WT



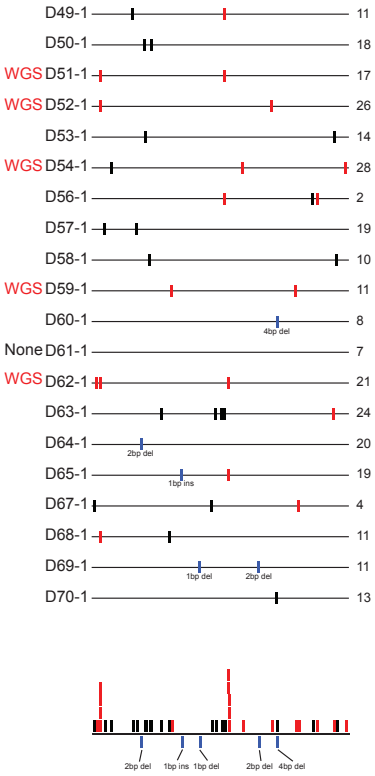
A3A E72A



Vector



A3B WT



A3B E255A

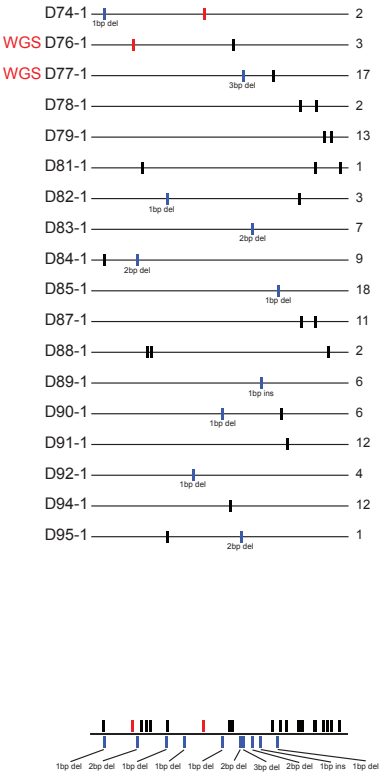


Figure 3

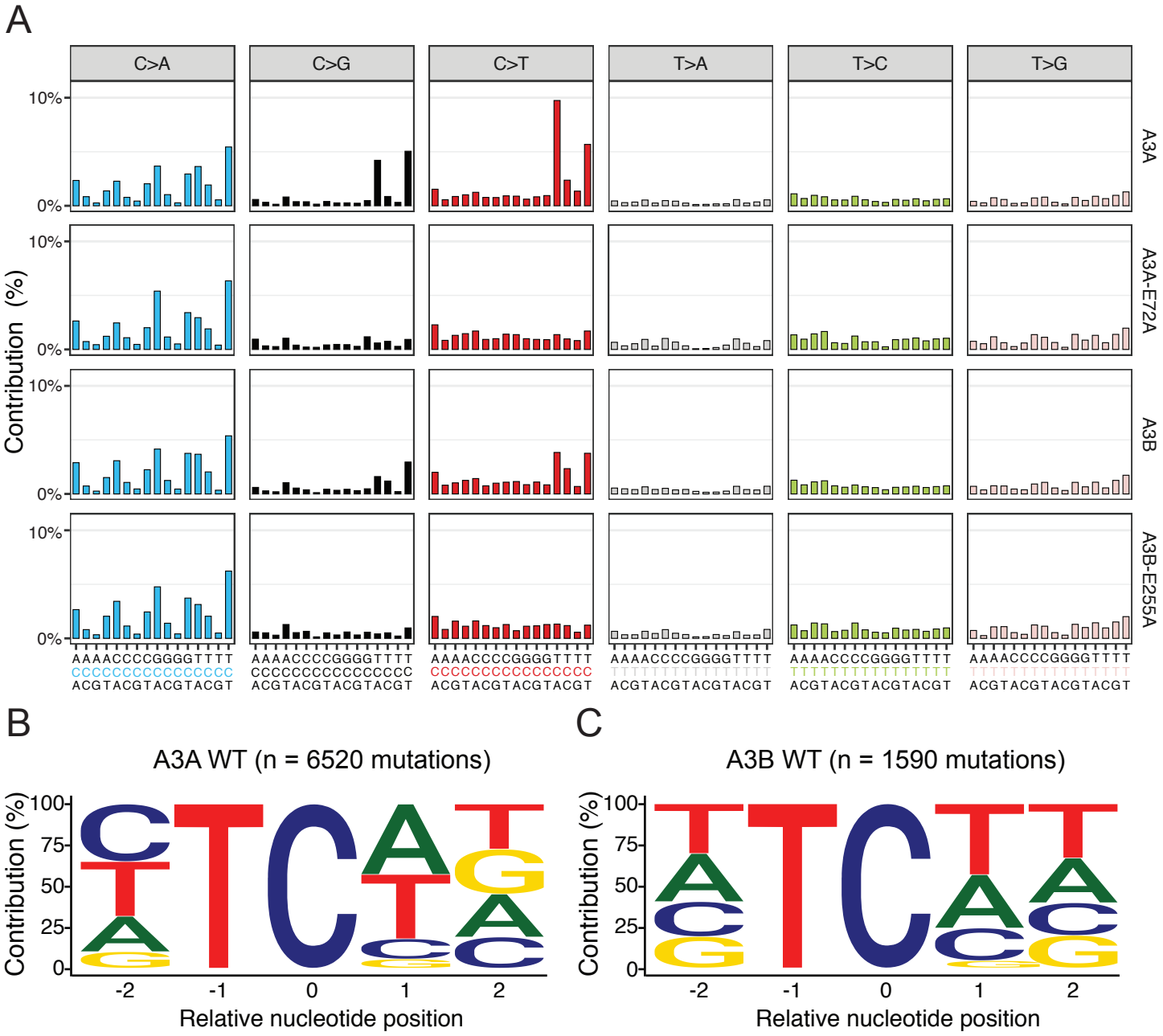
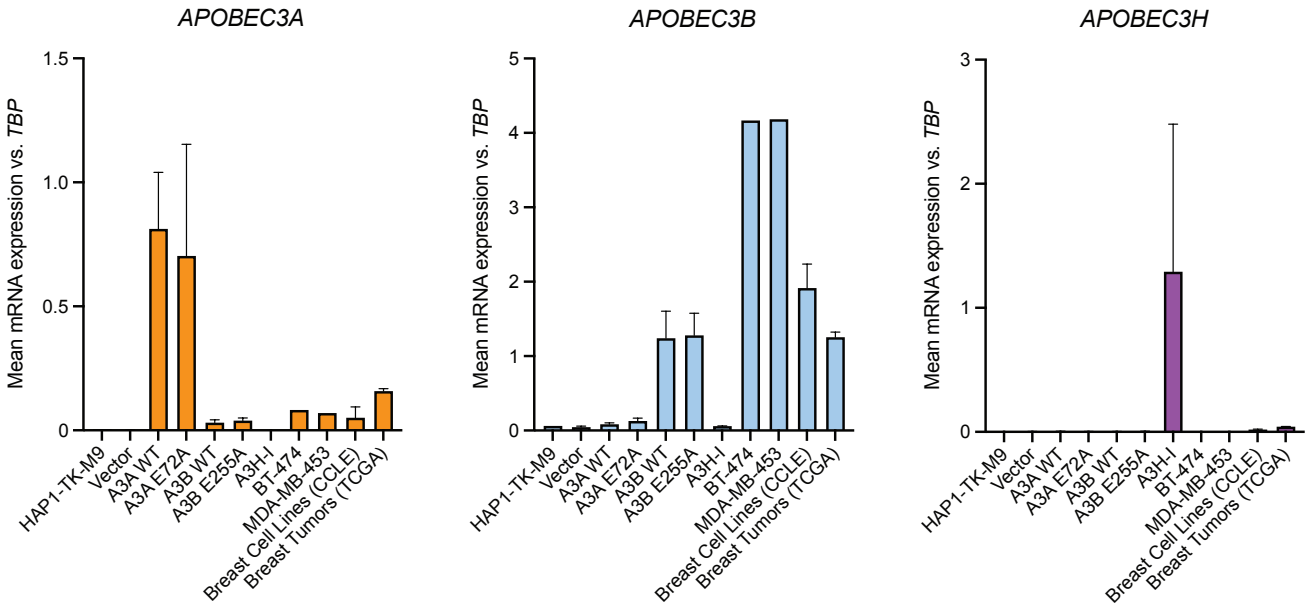

A
C>A
C>G
C>T
T>A
T>C
T>G
A3A
A3A-E72A
A3B
A3B-E255A
Contribution (%)
10%
0%
AAAACCCCGGGGTTTT
CCCCCCCCCCCCCCCC
ACGTACGTACGTACGT
B
A3A WT (n = 6520 mutations)
C
A3B WT (n = 1590 mutations)
Contribution (%)
100
75
50
25
0
-2
-1
0
1
2
Relative nucleotide position

Figure 3 - Supp 1

A



B

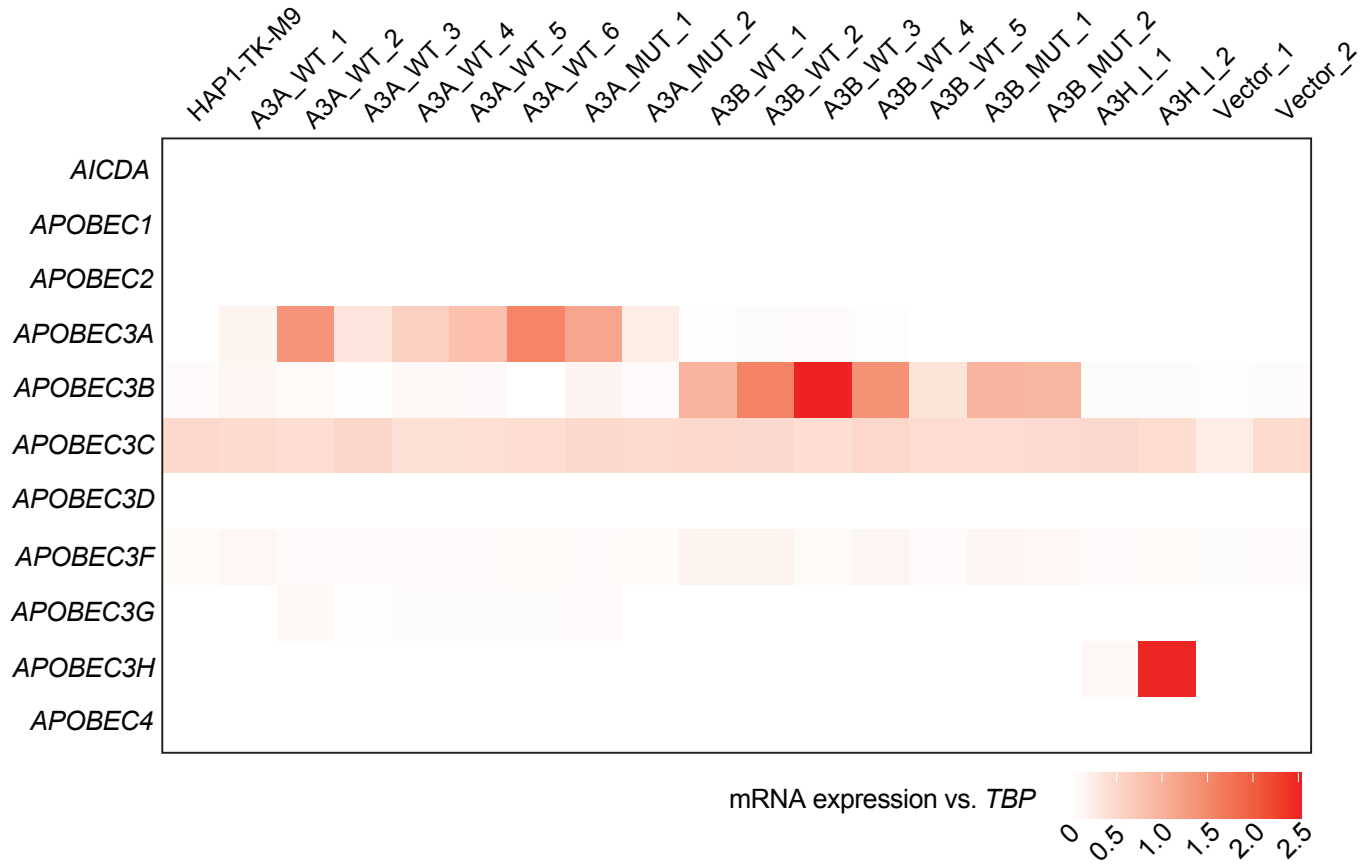


Figure 3 - Supp 2

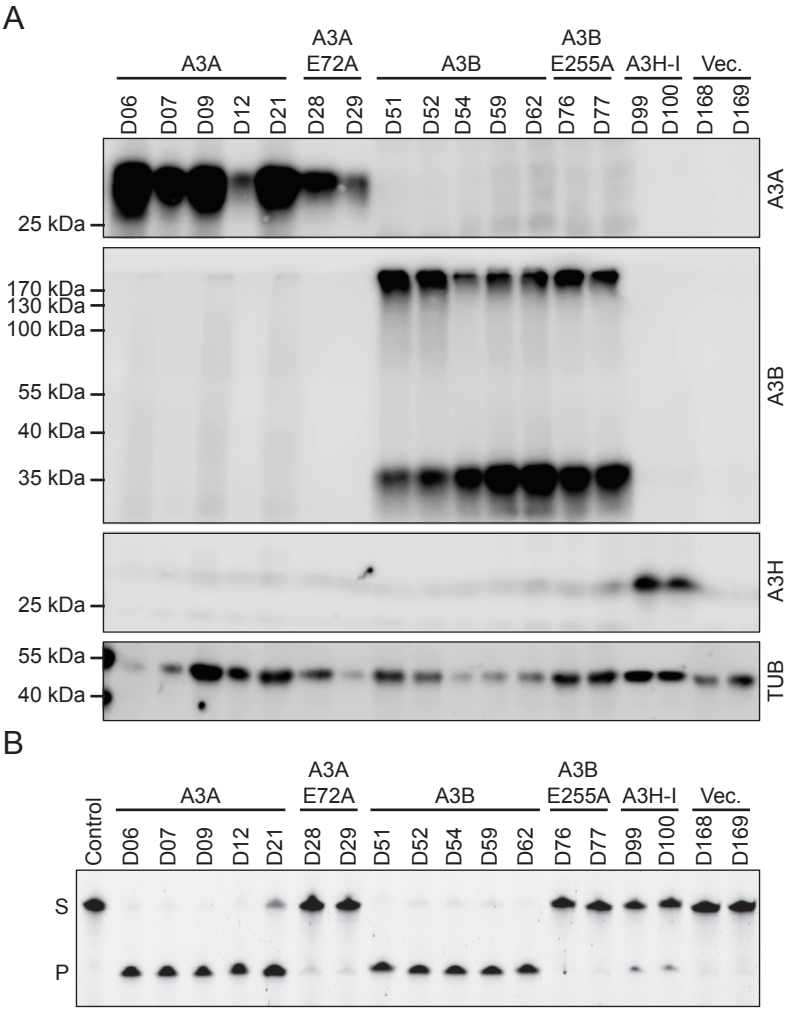


Figure 3 - Supp 3

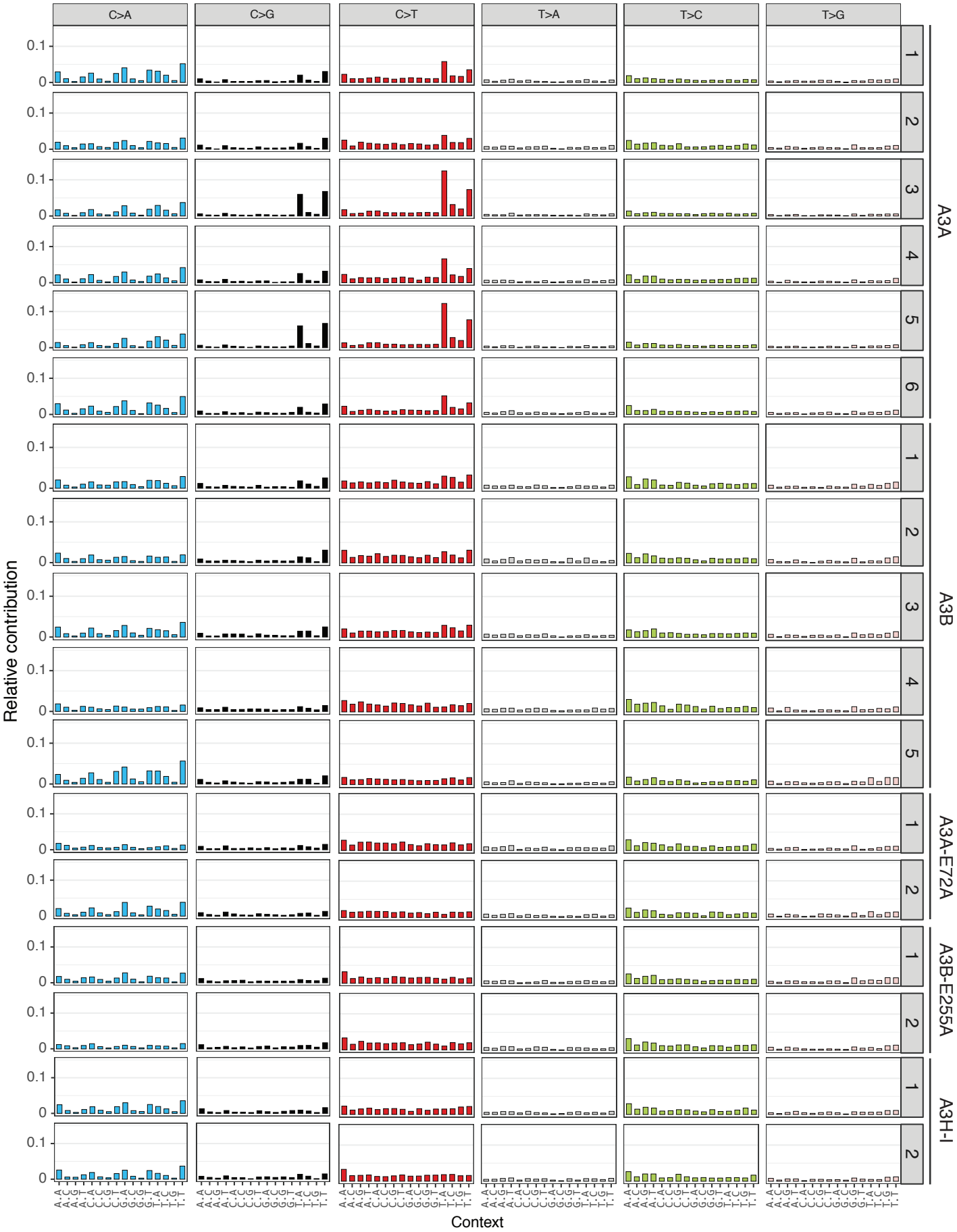
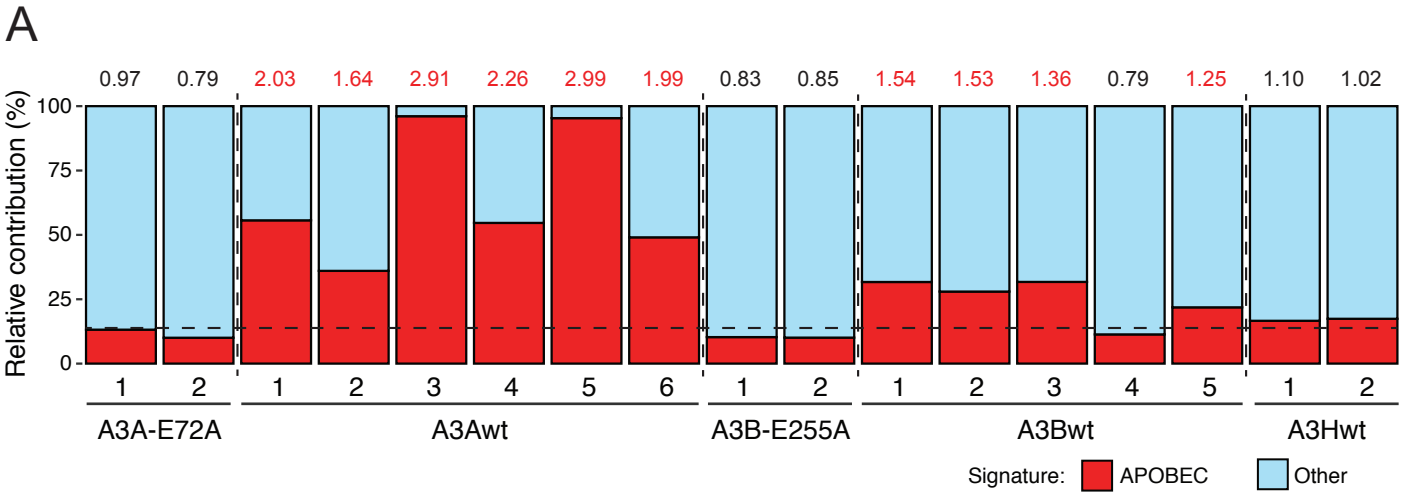


Figure 3 - Supp 4



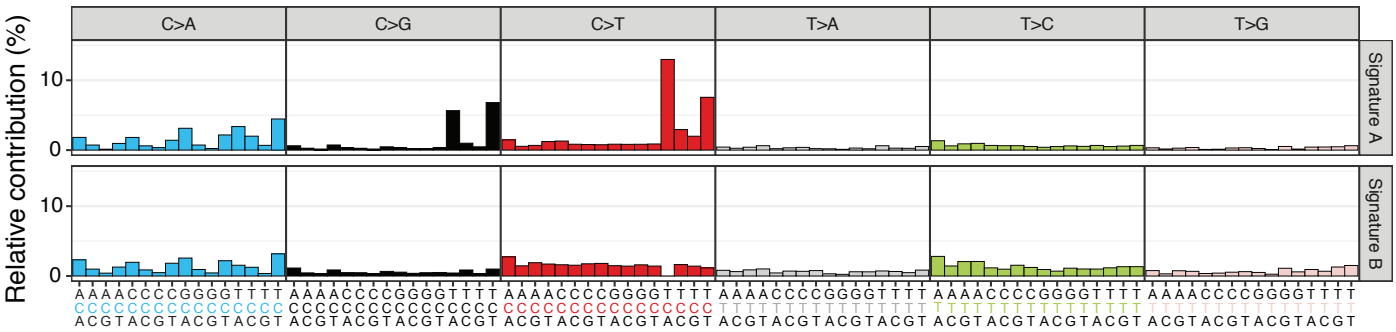


Figure 4

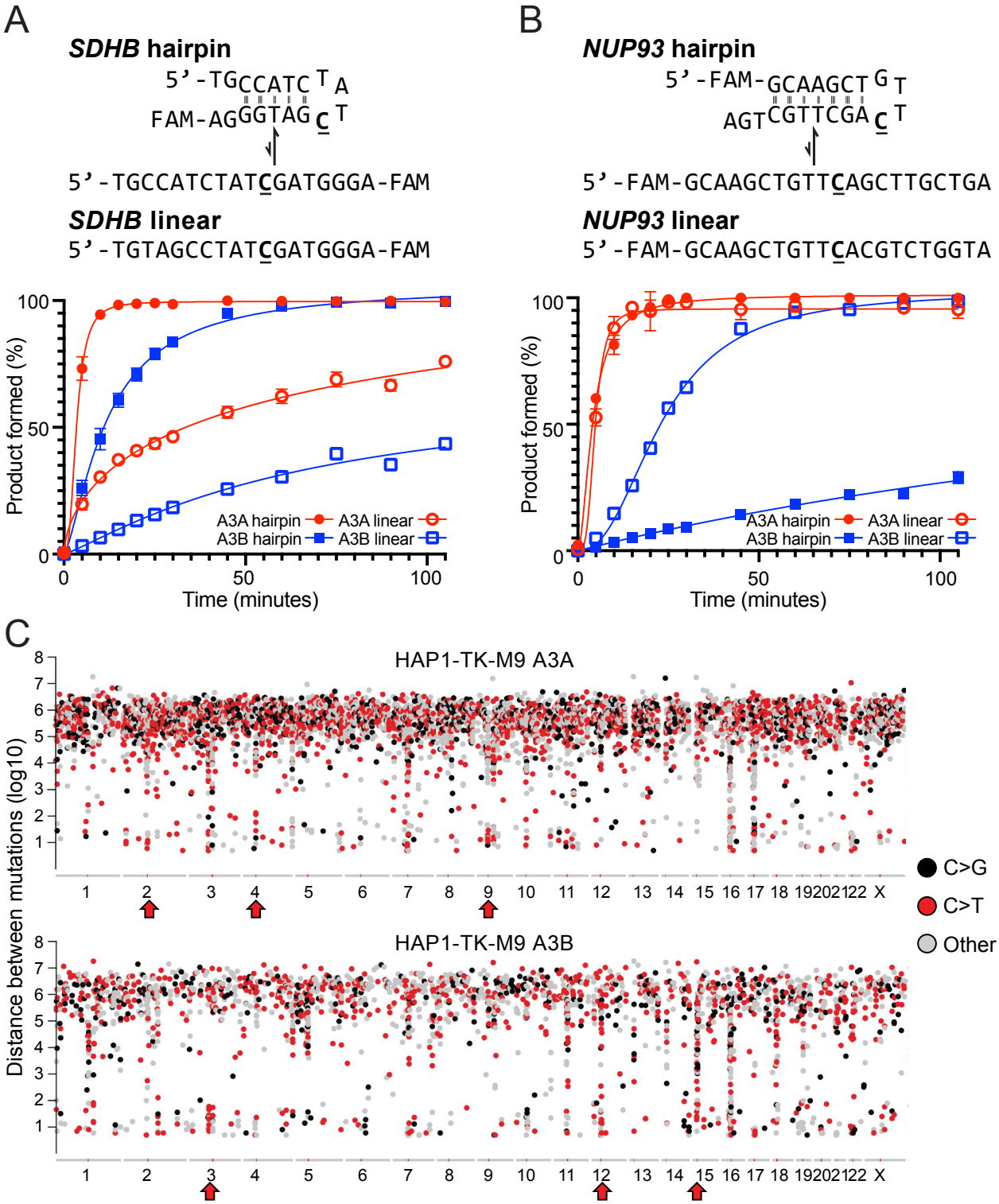

A
SDHB hairpin
5'-TGCCATCTA
FAM-AGGGTAGCT
5'-TGCCATCTATCGATGGGA-FAM
SDHB linear
5'-TGTAGCCTATCGATGGGA-FAM
B
NUP93 hairpin
5'-FAM-GCAAGCTGT
AGTCGTTCGACT
5'-FAM-GCAAGCTGTTCAGCTTGCTGA
NUP93 linear
5'-FAM-GCAAGCTGTTCACGTCTGGTA
Product formed (%)
Time (minutes)
A3A hairpin
A3A linear
A3B hairpin
A3B linear
C
HAP1-TK-M9 A3A
HAP1-TK-M9 A3B
Distance between mutations (log10)
C>G
C>T
Other

Figure 4 - Supp 1

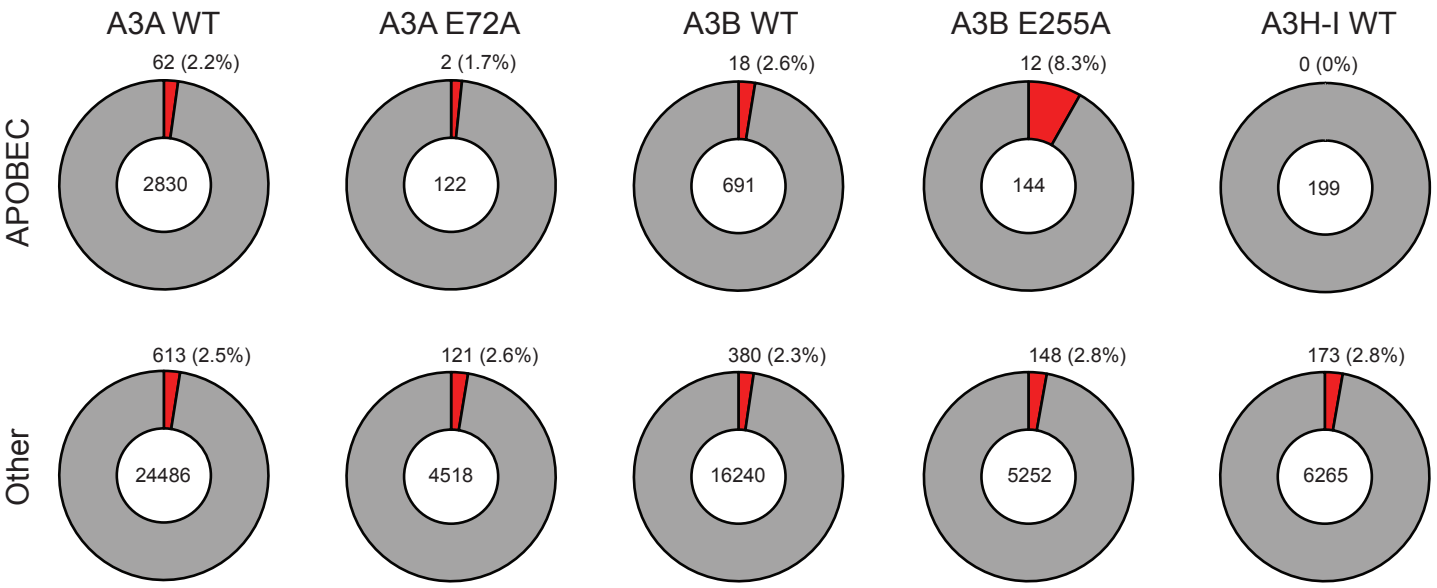
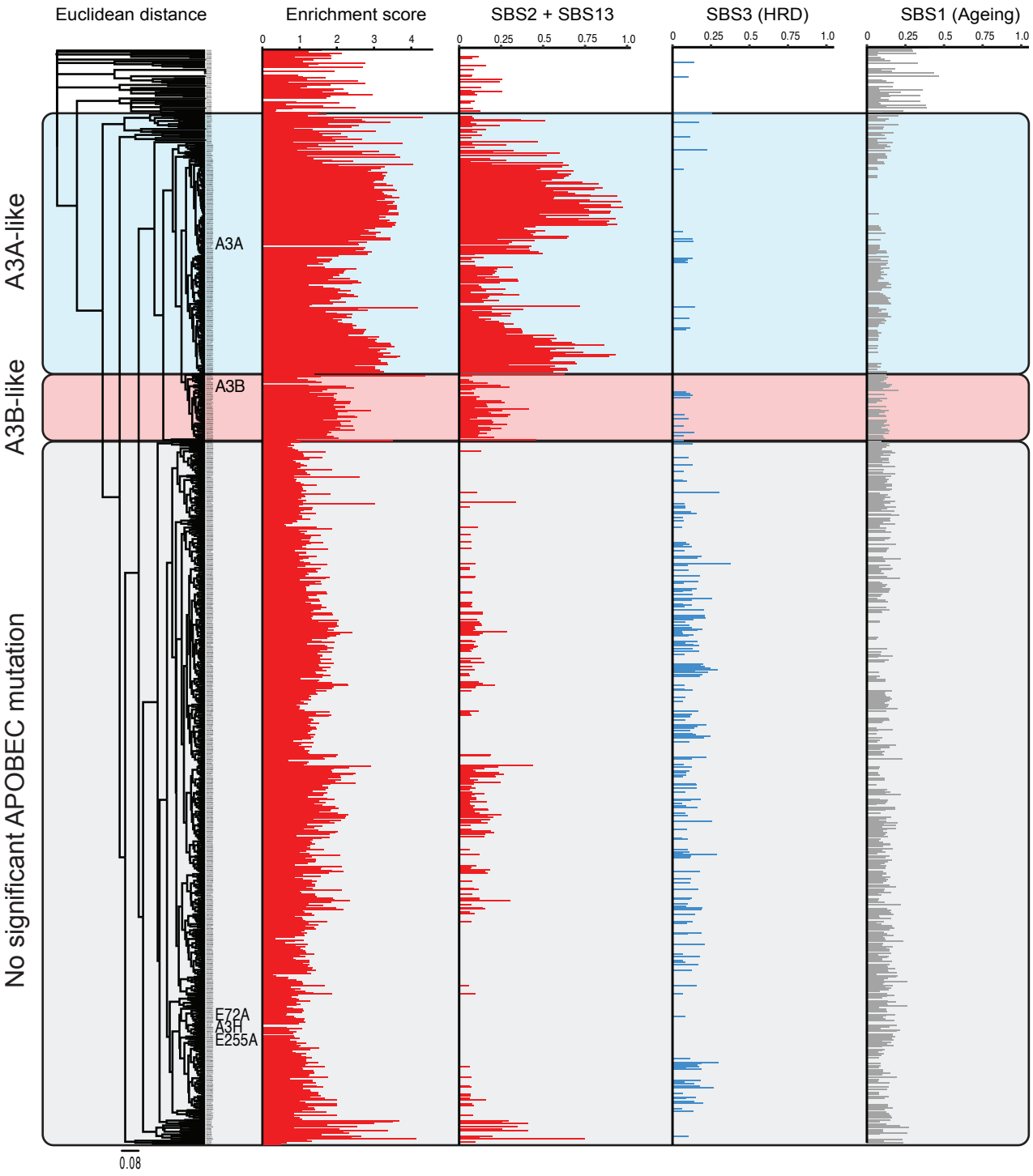


Figure 5


Euclidean distance
Enrichment score
SBS2 + SBS13
SBS3 (HRD)
SBS1 (Ageing)
0
1
2
3
4
0
0.25
0.5
0.75
1.0
A3A-like
A3B-like
No significant APOBEC mutation
A3A
A3B
E72A
A3H
E255A
0.08