Mutational impact of APOBEC3A and APOBEC3B in a human cell line and comparisons to breast cancer

Running title: Mutation signatures of APOBEC3A and APOBEC3B

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

[1] Department of Biochemistry, Molecular Biology and Biophysics, 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] Department of Biochemistry and Structural Biology, University of Texas Health San Antonio, San Antonio, TX 78229, USA
[5] Howard Hughes Medical Institute, University of Texas Health San Antonio, San Antonio, TX 78229, USA
[6] Institute for Health Informatics, University of Minnesota, Minneapolis, MN 55455, USA
[7] Department of Medicine, University of Minnesota Medical School, Minneapolis, Minnesota 55455, USA
[8] Correspondence to Reuben S. Harris, email: rsh@uthscsa.edu
[*] Equal contributions

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

Manuscript Information: 96 character main title; 44 character short title; 298 word abstract; 5900 word main text (excluding methods, references, and legends); 5 Main Figures, 14 Supplementary Figures, and 1 Supplementary Table.

Abstract

A prominent source of mutation in cancer is single-stranded DNA cytosine deamination by cellular APOBEC3 enzymes, which results in signature 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 protein(s) are responsible. Here we report the development of a selectable system to quantify genome mutation and demonstrate its utility by comparing the mutagenic activities of three leading candidates - APOBEC3A, APOBEC3B, and APOBEC3H. The human cell line, HAP1, is engineered to express the *thymidine kinase* (*TK*) gene of HSV-1, which confers sensitivity to ganciclovir. Expression of APOBEC3A and APOBEC3B, but not catalytic mutant controls or APOBEC3H, triggers increased frequencies of *TK* mutation and nearly indistinguishable TC-biased cytosine mutation profiles in the selectable *TK* reporter gene. Whole genome sequences from *TK* mutant clones enabled an analysis of thousands of single base substitution mutations and extraction of local sequence preferences with APOBEC3A preferring YTCW motifs over 70% of the time and APOBEC3B just under 50% of the time (Y=C/T; W=A/T). Signature comparisons with breast tumor whole genome sequences indicate that most malignancies manifest intermediate percentages of APOBEC3 signature mutations in YTCW motifs, mostly between 50 and 70%, suggesting that both enzymes are contributing in a combinatorial manner to the overall mutation landscape. Although the vast majority of APOBEC3A- and APOBEC3B-induced single base substitution mutations occur outside of predicted chromosomal DNA hairpin structures, whole genome sequence analyses and supporting biochemical studies also indicate that both enzymes are capable of deaminating the single-stranded loop regions of DNA hairpins at elevated rates relative to control conditions. These studies combine to help resolve a long-standing etiologic debate on the source of APOBEC3 signature mutations in cancer and indicate that future diagnostic and therapeutic efforts should focus on both APOBEC3A and APOBEC3B.

59 **Introduction**

60 Over the past decade, advances in DNA sequencing technologies and bioinformatics have
61 helped to deconvolute a multitude of mutational processes that contribute to the genesis and
62 evolution of cancer (see pan-cancer analysis by [1] and reviews by [2-4]). Through these
63 approaches and complementary wet lab experiments, the APOBEC3 family of single-stranded
64 (ss)DNA cytosine deaminases has emerged as one of the top three sources of single base
65 substitution (SBS) mutation in cancer with particularly large contributions to tumors of the bladder,
66 breast, cervix, lung, and head/neck. APOBEC3 signature mutations in cancer are defined as C-
67 to-T transitions and C-to-G transversions in 5'-TCW motifs (W = A or T; SBS2 and SBS13,
68 respectively) [1, 5-10]. This definition is conservative because several APOBEC3 enzymes can
69 also accommodate 5'-CG and 5'-methyl-CG ssDNA substrates [9-17], which can also lead to C-
70 to-T transition mutations and overlap with the mutation signature attributable to spontaneous
71 water-mediated deamination of cytosine and methyl-cytosine nucleotides. Spontaneous, water
72 mediated methyl-cytosine deamination is a clock-like mutation process that occurs predominantly
73 in 5'-methyl-CG motifs and associates positively with a patient's biological age (ageing signature
74 [18]), whereas APOBEC3-catalyzed deamination is absent in most normal tissues, not associated
75 with ageing, present in many primary tumors, and often enriched in metastases (APOBEC3
76 signature [1, 6, 8, 9, 18-23]).

77 The human APOBEC3 (A3) family is comprised of seven different enzymes with extensive
78 homology and overlapping activities (reviewed by [24-26]), and it is unclear how much (or little)
79 each contributes to the composite APOBEC3 mutation signature evident in tumor DNA
80 sequences. The bulk of evidence favors two enzymes, APOBEC3A (A3A) and APOBEC3B (A3B),
81 though to wildly different degrees depending on the study, and additional work has also implicated
82 APOBEC3H (A3H). Evidence for A3A includes an intrinsic preference for 5'-TC substrates, high
83 catalytic activity (highest of any human DNA deaminase), cell-wide localization, DNA damage
84 responses, induction by the tumor virus HPV, and positive correlations between mRNA levels and

85 tumor APOBEC3 mutation loads [11, 14, 16, 27-34]. Evidence for A3B includes all the same
86 points except this enzyme is several-fold less active, localizes constitutively to the nuclear
87 compartment, is induced by multiple DNA tumor viruses (herpesviruses, papillomaviruses, and
88 polyomaviruses), and associates at both mRNA and protein levels with clinical outcomes [9, 11,
89 16, 19-21, 33-44]. Importantly, however, the APOBEC3 mutation signature can still accumulate
90 in tumors that lack *A3B* due to a naturally occurring full gene deletion allele [45-47]. This
91 observation with *A3B*-null tumors helped implicate a cell-wide, 5'-TC preferring variant of A3H
92 (haplotype I) in contributing to the overall APOBEC3 mutation signature [46, 48, 49]. An important
93 additional consideration is that A3A and A3B have been reported to exhibit broader
94 tetranucleotide preferences upon expression in yeast, 5'-YTCW vs RTCW, respectively [50-52].

95 Here, we develop a human cellular system for mutation research and use it to compare
96 the mutagenic potential of A3A, A3B, and A3H. The human cell line HAP1 was engineered to
97 express a single copy of the HSV-1 *thymidine kinase* (*TK*) gene, which enables the drug
98 ganciclovir to be used to select rare *TK* mutants and quantify mutation frequencies. Moreover,
99 the *TK* gene can be amplified readily from ganciclovir-resistant (Gan^R) clones by high-fidelity PCR
100 and sequenced to provide initial assessments of mutation spectra (signatures) prior to
101 undertaking additional experiments such as whole genome sequencing (WGS). Using this
102 system, only expression of A3A and A3B cause significant increases in Gan^R mutation
103 frequencies. Sanger sequences from panels of individual *TK* mutant clones show a clear TC-
104 biased mutation pattern including two hotspots with no obvious hairpin structure. WGS of *TK*
105 mutant clones demonstrates that both A3A and A3B can generate the APOBEC3 mutation
106 signatures SBS2 and SBS13. These single base substitution mutations are mostly dispersed
107 (non-clustered) throughout the genomes and both enzymes exhibit similar frequencies of mutation
108 in TCW motifs in chromosomal DNA predicted to form non-hairpin versus hairpin structures.
109 However, in comparison to catalytic mutant controls, both enzymes exhibit higher frequencies of
110 APOBEC3 signature mutation in the single-stranded DNA loop regions of predicted hairpin

111 structures. WGS also shows that A3A has a strong (slightly over 70%) preference for triggering
112 cytosine mutations in YTCW sequence motifs, whereas A3B has a weaker (slightly under 50%)
113 preference for the same motif. In comparison, APOBEC3 mutation signature-enriched primary
114 breast cancers show predominantly intermediate frequencies of mutations in YTCW motifs,
115 between 50% and 70%, suggesting involvement from both enzymes.
116

117 **Results**

118 **HAP1-TK-M9 – a human cellular system to report DNA damage and mutagenesis**

119 Model organisms such as *E. coli* and yeast are powerful systems for studying mutagens
120 including DNA deaminases (*e.g.*, original studies with A3 enzymes [34, 53-57]). However, these
121 model organisms only recapitulate a subset of DNA repair and regulatory mechanisms found in
122 human cells. We therefore sought to combine strengths of both approaches by introducing a
123 single copy of a selectable reporter, the HSV-1 *thymidine kinase* (*TK*) gene, into the genome of
124 the human cell line HAP1. Expression of the thymidine kinase (TK) protein confers exquisite
125 sensitivity to the drug ganciclovir and, as for many antimicrobial agents, only *TK*-mutant,
126 ganciclovir-resistant (Gan^R) cells survive selection by this drug. Gan^R mutants can be
127 characterized rapidly by conventional Sanger DNA sequencing because *TK* is a single open
128 reading frame. Moreover, once informative Gan^R mutants are revealed by selection, secondary
129 analyses including WGS can be used to uncover additional and potentially global features of a
130 given mutation process.

131 The overall experimental workflow is shown in **Figure 1A**. To generate a "mother" clone
132 of the commercially available HAP1 cell line, Sleeping Beauty (SB)-mediated transposition was
133 used to introduce a single copy of a *TK-Neo* cassette into the genome [9]. Neo^R clones were
134 selected with G418, expanded into healthy clonal populations (ca. 10^6 cells/ml), and screened for
135 ganciclovir sensitivity (Gan^S). One mother clone, HAP1-TK-M9, was selected for further studies
136 because it is Gan^S, it is mostly diploid (apart from pre-existing chromosome aberrations), it

137 cultures, engineers, and clones well (below), and it has a favorable *A3* expression profile (**Figure**
138 **1B**; **Figure 1 – Figure Supplement 1**). In particular, RT-qPCR measurements showed that its
139 *A3A* and *A3B* mRNA levels are lower than those of the original parent line and that *A3H* mRNA
140 levels are very low and near the detection threshold (**Figure 1B**). Genomic DNA sequencing also
141 revealed that this cell line's only *A3H* allele is haplotype III (ΔAsn15), which is known to produce
142 an unstable protein [58-60].
143
144 **Functionality of human A3s expressed from MLV-based constructs**
145 Plasmid constructs were assembled for these studies in which human *A3* expression is
146 driven by a *MND* promoter from within an MLV-based retroviral construct. An additional feature
147 of these constructs is a downstream puromycin resistance cassette, which facilitates selection of
148 expressing cells. The functionality of each construct was assessed by transfection into 293T cells
149 and, following 24-48 hrs incubation, protein analysis by immunoblotting and ssDNA deaminase
150 activity assays (**Figure 1C-E**; custom rabbit anti-human A3A monoclonal antibody validation
151 described in **Figure 1 – Figure Supplement 2**). Each A3 expressed at the expected kilodalton
152 size and only the wildtype enzymes exhibited catalytic activity. A3H-I was expressed at lower
153 levels than A3H-II and, accordingly, had less ssDNA deaminase activity in soluble extracts
154 (**Figure 1E**), in agreement with prior reports [58-60].
155 We next used immunofluorescent (IF) microscopy to examine subcellular localization and
156 DNA damage responses triggered by expression of each enzyme 3 days post-transduction into
157 HAP1-TK-M9 cells (without puromycin selection). A3A appeared cell-wide and associated with a
158 large increase in overall staining of the DNA damage marker γ-H2AX as well as an increase in
159 individual γ-H2AX foci (representative images in **Figure 1F**, quantification in **Figure 1G-H**, and
160 additional images in **Figure 1 – Figure Supplement 3**). Expression of an enzymatically inactive
161 mutant, A3A-E72A, exhibited vector control levels of γ-H2AX staining. In comparison, neither
162 expression of wildtype A3B nor A3H (haplotype I or II) was able to trigger statistically significant

163 increases in γ-H2AX levels in HAP1-TK-M9 cells (representative images in **Figure 1F**,
164 quantification in **Figure 1G-H**, and additional images in **Figure 1 – Figure Supplement 3**). These
165 results were unexpected given that A3B localization is predominantly nuclear, and A3H
166 (haplotypes I and II) can also access the nuclear compartment with some accumulation in nucleoli
167 (**Figure 1F** and additional images in **Figure 1 – Figure Supplement 3**). These results differed
168 from prior reports on A3B and A3H overexpression causing elevated γ-H2AX levels [9, 43, 48,
169 61], perhaps because expression levels here are lower with the *MND* promoter (versus strong
170 Tet/Dox-inducible systems) and/or because the HAP1 system is somehow unique and more
171 tolerant/adaptable to expression of these ssDNA deaminases. However, HAP1-TK-M9 cells are
172 indeed capable of a canonical DNA damage response, as demonstrated by γ-H2AX accumulation
173 following treatment with cisplatin (*cis*-diamminedichloroplatinum II; representative images in
174 **Figure 1F**, quantification in **Figure 1G-H**, and additional images in **Figure 1 – Figure**
175 **Supplement 3**).

176 To extend these results to a different cell line, our MLV-based A3 expression plasmids
177 were transfected transiently into HeLa cells and, after 24 hrs incubation, subjected to additional
178 analysis by IF microscopy. As with HAP1-TK-M9 cells, A3A appeared cell-wide (except nucleoli),
179 A3B predominantly nuclear, and A3H haplotype II cytoplasmic with nucleolar accumulations
180 (**Figure 1 – Figure Supplement 4A**). A3H haplotype I was not analyzed here due to low
181 expression and weak ssDNA deaminase activity in HAP1-TK-M9 cells. Interestingly, A3A caused
182 a strong pan-nuclear increase in γ-H2AX without concomitant focus formation (representative
183 images in **Figure 1 – Figure Supplement 4A** and quantification in **Figure 1 – Figure**
184 **Supplement 4B**). Similar results have been reported for A3A overexpression in other cell types
185 [30, 62-64]. Interestingly, although a dose-responsive accumulation of γ-H2AX was expected
186 given a wide range of transient transfection efficiencies for individual cells in each reaction, only
187 a weak positive association was found with even low A3A staining cells exhibiting very high γ-
188 H2AX levels (**Figure 1 – Figure Supplement 4B-C**). In contrast, comparatively low levels of

189 nuclear γ-H2AX were observed in cells expressing A3B, A3H, and catalytic mutant derivative
190 proteins, although A3B expression uniquely triggered a modest elevation of nuclear γ-H2AX levels
191 in comparison to the background observed in cells expressing the corresponding catalytic mutant
192 E255A protein (**Figure 1 – Figure Supplement 4A-D**). Moreover, as expected from the strong
193 increase in γ-H2AX levels, only wildtype A3A caused significant increases in DNA breakage as
194 quantified with alkaline comet assays (representative images in **Figure 1 – Figure Supplement**
195 **4E** and quantification in **Figure 1 – Figure Supplement 4F**). Taken together, results with two
196 different cell lines demonstrate that A3A expression induces a strong DNA damage response
197 (high γ-H2AX and DNA breakage), A3B expression triggers a modest DNA damage response
198 (low γ-H2AX and no overt DNA breakage), and A3H or catalytic mutant A3A/B/H expression is
199 indistinguishable from background levels in negative vector control conditions.
200

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

202 To directly test which A3 enzymes cause genomic mutation in the HAP1-TK-M9 system,
203 24 independent single-cell derived daughter clones were obtained for A3A, A3B, A3H-I, and A3H-
204 II expressing conditions as well as catalytic mutant and vector controls (**Methods**). A classical
205 fluctuation analysis was performed by growing each single cell clone for 1 month to $>10^7$ cells,
206 subjecting each population to selection by ganciclovir, and allowing time for single Gan^R mutant
207 cells to grow into countable colonies. Vector control conditions yielded a median Gan^R mutation
208 frequency of 3 mutants per 5 million cells (mean = 3.6, SD = 2.8). In contrast, A3A- and A3B-
209 expressing clones caused median Gan^R mutation frequencies to rise above 10 mutants per 5
210 million cells (A3A WT: median = 16, mean = 14, SEM = 2.2; A3B WT: median = 11, mean = 12,
211 SEM = 1.8; **Figure 2A**). In comparison, expression of A3H-I or A3H-II or catalytic mutant
212 derivatives of any of these DNA deaminases or empty vector controls failed to trigger increased
213 Gan^R mutation frequencies (A3H-I WT: median = 3.0, mean = 3.6, SEM = 0.71; A3H-II WT:
214 median = 2.0, mean = 2.8, SEM = 0.58; **Figure 2A**).

215 We next asked what types of genetic alterations led to inactivation of the *TK* gene in Gan^R
216 granddaughter clones derived from the different A3-expressing and control conditions. The *TK*
217 gene was PCR-amplified from genomic DNA of Gan^R granddaughter clones and Sanger
218 sequenced. C-to-T and C-to-G mutations in an APOBEC3-signature trinucleotide motif
219 (APOBEC), all other single base substitution mutations (other SBS), and all insertion/deletion
220 mutations (INDELs) were placed into groups for comparison (red, black, and blue tics in **Figure**
221 **2B**, respectively; individual sequence schematics in **Figure 2 – Figure Supplement 1**). *TK*
222 sequences derived from A3A-expressing granddaughter clones harbored a greater number of
223 APOBEC3 signature mutations [12/20 clones contained at least 1 APOBEC mutation, 22
224 T(C>T/G)W mutations total, range of 0 to 3 SBS per sequence] relative to catalytic mutant control
225 clones [3/19 clones contained 1 APOBEC mutation, 3 T(C>T/G)W mutations total, range of 0 to
226 1 SBS per sequence]. Similarly, *TK* sequences derived from A3B-expressing granddaughter
227 clones also harbored a greater number of APOBEC3 signature mutations [11/20 clones contained
228 at least 1 APOBEC mutation, 19 T(C>T/G)W mutations total, range of 0 to 3 SBS per sequence]
229 relative to catalytic mutant control clones [2/18 clones contained 1 APOBEC mutation, 2
230 T(C>T/G)W mutations total, range of 0 to 1 SBS per sequence] (**Figure 2B**; **Figure 2 – Figure**
231 **Supplement 1**). Additionally, *TK* cytosine bases 22 (Q8X) and 635 (R212K) emerged as
232 candidate mutational hotspots in both the A3A- and A3B-expressing clones (**Figure 2B**;
233 considered further in **Discussion**). No significant differences were found in the number of other
234 SBS or INDEL mutations between A3A- or A3B-expressing clones and controls.
235 We next examined the broader sequence context of the 22 A3A- and 19 A3B-induced
236 APOBEC3 signature mutations that occurred at 5'-TC dinucleotides in *TK* (**Figure 2C-D**). In both
237 instances, A was preferred over T at the +1 nucleobase position relative to the mutated C, and
238 this bias was not significantly different between the two enzymes (68% for A3A and 74% for A3B;
239 p=0.367 by Fisher's exact test). Similarly, no obvious biases were evident at the +2 or -2
240 nucleobase positions with all four nucleotides observed at similar frequencies for both enzymes.

241 Moreover, even when pyrimidines and purines were grouped for comparison, A3A did not show
242 an overt preference for C/T (Y) or A/G (R) at the -2 or +2 nucleobase positions (51% vs 49% and
243 52% vs 48%, respectively). Likewise, A3B also failed to show an overt preference for C/T (Y) or
244 A/G (R) at the -2 or +2 nucleobase positions (45% vs 55% and 53% vs 47%, respectively). These
245 similarities underscore the fact that small mutation numbers are primarily useful for delineating
246 major signature differences such as the shifts described above from a heterogeneous pattern in
247 catalytic mutant- or vector control-expressing cells towards a predominantly 5'-TC focused SBS
248 mutation pattern in A3A- and A3B-expressing cells.
249

250 **WGS shows SBS2 and SBS13 signatures reflecting the intrinsic biochemical preferences**
251 **of A3A and A3B**

252 To prepare for WGS of Gan^R granddaughter clones, RNA sequencing was done to confirm
253 expression of each exogenously expressed *A3* construct and compare mRNA levels relative to
254 established breast cell lines and primary breast tumors. All expression values were determined
255 relative to those of the conserved housekeeping gene *TBP* to be able to compare RNAseq data
256 from different sources (*e.g.*, unrelated cell lines and tumors). Interestingly, the average *A3A* and
257 *A3A-E72A* mRNA levels in Gan^R granddaughter clones was over 5-fold higher than the average
258 endogenous *A3A* expression levels of APOBEC3 signature-enriched breast cancer cell lines
259 BT474 and MDA-MD-453, breast cancer cell lines of the CCLE, or breast tumors of TCGA (**Figure**
260 **3 – Figure Supplement 1A**). It is possible that A3A levels are higher in this system due to difficulty
261 identifying a sufficiently weak promoter for constitutive expression. In comparison, *A3B* and *A3B-*
262 *E255A* mRNA levels in Gan^R granddaughter clones are similar to the averages reported for breast
263 cancer cell lines of the CCLE and breast tumors of the TCGA, and approximately 2-fold lower
264 than those of BT474 and MDA-MD-453 (**Figure 3 – Figure Supplement 1A**). *A3H* (haplotype I)
265 mRNA levels showed greater variance but only two clones were analyzed by RNAseq and WGS
266 due to negative results above with IF microscopy experiments and *TK* mutation analysis.

267 The mRNA expression levels of the other four *A3* genes, as well as *AICDA* (*AID*),
268 *APOBEC1*, *APOBEC2*, and *APOBEC4*, were also quantified and compared with those of *A3A*,
269 *A3B*, and *A3H* (**Figure 3 – Figure Supplement 1B**). Endogenous *A3C* was expressed at similarly
270 high levels in all granddaughter clones, providing a stable relatively internal control. Endogenous
271 *A3F* and *A3G* were expressed at lower but still detectable levels, and endogenous *AICDA*,
272 *APOBEC1*, *APOBEC2*, *APOBEC4*, and *A3D* were expressed at very low or undetectable levels.
273 As expected, levels of ectopically expressed *A3A*, *A3B*, and *A3H* mRNA exceeded those in the
274 HAP1-TK-M9 parent clone as well as those in vector expressing granddaughter controls. In
275 addition, protein expression of A3A, A3B, and A3H was confirmed in *TK* mutant granddaughter
276 clones by immunoblotting and activity by ssDNA deamination assays (**Figure 3 – Figure**
277 **Supplement 2**). Finally, the integration sites of the *TK* reporter and MLV-based A3 expression
278 constructs were determined using WGS reads. Data from multiple clones demonstrated a single
279 *TK-Neo* integration site in chromosome 3 between nucleotides 143996100-143996500, which is
280 ~3.5 kbp downstream of the nearest gene (*DIPK2A*), and representative *A3* expressing clones
281 additionally have single MLV-A3 insertion sites on different chromosomes as expected from the
282 independent reactions and low multiplicities of infection that used initially to establish these clones
283 (**Figure 3 – Figure Supplement 3**).
284 To investigate mutational differences genome-wide, Illumina short-read WGS was done
285 for randomly selected Gan^R granddaughter clones (specified by "WGS" in **Figure 2 – Figure**
286 **Supplement 1**). Mutations unique to each granddaughter were identified by calling SBS
287 variations versus the genomic DNA sequence of the HAP1-TK-M9 mother clone. This approach
288 eliminated any somatic variation that accumulated in the Gan^S mother clone prior to transduction
289 with each A3 or control expression construct. Thus, all new SBS mutations had to be present in
290 a significant proportion of reads from the Gan^R granddaughter clones and absent from the reads
291 from the Gan^S mother clone and, as such, must have occurred in the presence of an active A3
292 enzyme or a catalytic mutant control.

293 In A3A-expressing Gan^R granddaughter clones, the total number of unique SBSs ranged
294 from 2057 to 5256 (n=6, median=3129, mean=3494, SD=1358; **Figure 3 – Figure Supplement**
295 **4A**). The total number of SBSs in A3B-expressing clones was lower, ranging from 1920 to 2652
296 (n=5, median=2346, mean=2334, SD=278). In comparison, the total number of SBSs in catalytic
297 mutant and GFP control Gan^R granddaughter clones ranged from 1230 to 2204 (n=6;
298 median=1748, mean=1729, SD=388). SBS mutations are further broken down into those
299 occurring within NCN and TCW motifs in dot plots in **Figure 3 – Figure Supplement 4B-E**. Most
300 importantly, analyses of the trinucleotide contexts of all unique SBS mutations revealed strong C-
301 to-T and C-to-G mutation biases in 5'-TCA and 5'-TCT motifs in A3A-expressing clones and
302 weaker, but still significant, mutation biases in the same motifs in A3B-expressing clones (**Figure**
303 **3A**; **Figure 3 – Figure Supplement 4C and E**; **Figure 3 – Figure Supplement 5**). In other words,
304 only A3A- and A3B-expressing granddaughter clones exhibited significant amounts of APOBEC3
305 signature single base substitution mutations.

306 This result was confirmed by assessing APOBEC3 signature enrichment scores [50, 65],
307 which indicated that 6/6 A3A-expressing clones and 4/5 A3B-expressing clones have significant
308 enrichments of APOBEC3 signature mutations, whereas clones expressing catalytically inactive
309 A3A or A3B, as well as clones expressing A3H-I or vector control have none (**Figure 3 – Figure**
310 **Supplement 6A**). A complementary bioinformatics analysis, non-negative matrix factorization
311 (NMF [66]), yielded similar results with "signature A" resembling SBS2 and SBS13 in A3A- and
312 A3B-expressing clones (*i.e.*, APOBEC3 signature) and "signature B" occurring in all clones
313 regardless of A3 presence or functionality (**Figure 3 – Figure Supplement 6B**). In comparison,
314 genome-wide patterns of insertion/deletion mutations (INDELs) appeared largely unaffected by
315 A3A or A3B (**Figure 3 – Figure Supplement 7**), in agreement with aforementioned *TK* mutation
316 data where the majority of Gan^R mutations are single base substitution mutations and the INDEL
317 frequencies were similar for A3A, A3B, and catalytic mutant conditions.

318 We next analyzed the broader contexts of the 5'-TC-focused C-to-T and C-to-G single

319 base substitution mutations that accumulated in A3A- and A3B-expressing clones (n=7172 and
320 n=2033, respectively) in comparison to those that accumulated in aggregate control clones
321 (n=680) as well as the overall distribution of 5'-TC in the human genome (n=339619283) (**Figure**
322 **3B** and **Supplementary Table S1**). First, a bias for +1 A over +1 T emerged in A3A-expressing
323 clones (48.3% > 36.3%), whereas the opposite bias was evident in A3B-expressing clones (33.9%
324 < 41.2%). However, for both enzymes, the percentage of +1 A and T (W) was similar (84.6% and
325 75.1%, respectively). Second, no significant bias was noted at the +2 position except that guanine
326 is slightly over-represented in pentanucleotide motifs derived from A3A-expressing conditions
327 compared to motifs derived from A3B-expressing clones. Third and most importantly, an
328 exceptionally strong bias for a pyrimidine nucleobase (C or T) occurred at the -2 position in A3A-
329 expressing clones (72.9% YTC/NTC), which was also evident in the broader APOBEC3
330 tetranucleotide context (72.7% YTCW/NTCW). This cytosine mutation preference resembles the
331 strong -2 pyrimidine bias reported for human A3A in murine hepatocellular carcinomas [67], the
332 chicken B cell line DT40 [73], and yeast [50, 51]. Fourth, in contrast, a slight bias for purine
333 nucleobases (A or G) was apparent at the -2 position in A3B-expressing clones (51.1%
334 RTC/NTC), which was also reflected in the broader APOBEC3 tetranucleotide context (52.8%
335 RTCW/NTCW). This latter result also agreed with prior data from yeast [50, 51], but contrasted
336 slightly with the mutation signature detected in tumors derived from human A3B-expresssing mice
337 (47% RTCW [68]). Possible explanations for this variability are considered in **Discussion**.
338

339 **Features of A3A and A3B mutagenesis in the HAP1-TK-M9 system**

340 X-ray structures have revealed a U-shaped bend in ssDNA substrates bound by A3A and
341 A3B [16, 29], and other studies have indicated that similarly bent ssDNA loop regions of hairpins
342 (*i.e.*, DNA cruciform or stem-loop structures) may be preferred substrates for deamination by A3A
343 [52, 69, 70]. To ask whether this preference extends to the HAP1-TK-M9 system described here,
344 we analyzed our A3A and A3B *TK* PCR sequences and granddaughter clone WGS data for

345 evidence of mutagenesis in the single-stranded loop regions of DNA hairpin structures. First,
346 neither of the two A3A/B mutation hotspots in the *TK* gene reported above appeared to be part of
347 predicted stem-loop structures. Second, none of the top-100 cruciform structures reported
348 previously to harbor recurring APOBEC3 signature mutations in tumors [52] were mutated in our
349 HAP1-TK-M9 WGS data sets. Third, in global comparisons of base substitution mutations, the
350 frequency of APOBEC3 signature TCW mutations was similar in ssDNA loop regions of predicted
351 hairpin structures versus non-hairpin regions (*i.e.*, APOBEC3 signature mutation events were not
352 enriched in the loop regions of stem-loop structures over those occurring in canonical ssDNA
353 substrates; **Figure 4 – Figure Supplement 1**). Moreover, the frequency of APOBEC3 signature
354 mutations in hairpin or non-hairpin structures appeared similar (not statistically distinguishable) in
355 A3A- and A3B-expressing conditions. However, interestingly, the frequency of APOBEC3
356 signature mutations in predicted loop regions of hairpin structures appeared higher in A3A and
357 A3B expressing clones in comparison to all non-catalytic control conditions (**Figure 4 – Figure**
358 **Supplement 1**; P = 0.0655 and P = 0.0367 by Welch's t-test, respectively). The A3A versus
359 control comparison likely failed to reach statistical significance due to the small number of clones
360 with WGS (n = 6) and the large variance in numbers of TCW mutations in hairpin loop regions in
361 the different clones (range = 1-16 TCW mutations). Nevertheless, these results combined to
362 indicate that single-stranded loop regions of hairpin structures in human chromosomal DNA may
363 be similarly susceptible to deamination by both A3A and A3B.

364 To assess relative rates of A3A and A3B-catalyzed deamination of experimental hairpin
365 versus non-hairpin substrates, we used purified enzymes and ssDNA substrates representing two
366 previously reported RNA editing hotspots of A3A – *SDHB* and *NUP93* [52, 71]. The *SDHB* hairpin
367 is predicted to have a 5 bp stem and a 4 nt loop, and the *NUP93* hairpin a 7 bp stem and a 4 nt
368 loop (**Figure 4A-B**). The control oligonucleotides have the same loop region sequences and a
369 randomization of one-half of the nucleobases in the hairpin stem to reduce base-pairing potential.
370 In each case, the linear substrates migrated similarly on native and denaturing PAGE while the

371 hairpin substrates migrated faster by native PAGE and similar to the linear substrates when
372 denatured, thereby confirming the integrity of both hairpins (**Figure 4 – Figure Supplement 2**).
373 A3A and A3B were affinity-purified from human cells and incubated under single hit conditions
374 with these oligonucleotide substrates over time. First, both A3A and A3B showed a strong
375 preference for deaminating the *SDHB* hairpin substrate in comparison to a linear control with the
376 same nucleobase content scrambled (~4- and ~8-fold preference, respectively; **Figure 4A**). A3A
377 showed higher rates of deamination than A3B on both the hairpin and the linear substrate in
378 agreement with prior studies [52]. Thus, the relative deamination rates for *SDHB* substrates were:
379 A3A/hairpin > A3B/hairpin > A3A/linear > A3B linear (120, 41, 31, and 5.1 nM/min, respectively).
380 However, a different picture emerged from analyses of deamination of *NUP93*-based
381 substrates **Figure 4B**). Rates of A3A-catalyzed deamination were similarly high for the *NUP93*
382 hairpin and linear control with the same nucleobase content scrambled. In contrast, A3B showed
383 higher rates of deamination of the linear substrate and was only able to deaminate the hairpin
384 substrate with low efficiencies and linear kinetics. Thus, the relative deamination rates for *NUP93*
385 substrates are: A3A/hairpin = A3A/linear >> A3B linear > A3B/hairpin (93, 83, 7.7, and 1.3 nM/min,
386 respectively). These results combined to show that <u>*both*</u> A3A and A3B can deaminate hairpin and
387 linear ssDNA substrates and, further, that it is possible to identify sites such as the *NUP93* hairpin
388 (here with DNA and previously shown with RNA and DNA [52]) that are strongly (and perhaps
389 even exclusively, in a few instances) preferred by a single A3 enzyme.
390 Another feature of APOBEC3 mutagenesis in human cancer is clusters of strand-
391 coordinated cytosine SBS mutations in TCA and TCT motifs most likely caused by processive
392 deamination of exposed tracts of ssDNA (aka. *kataegis*; here defined as $\geq$2 strand-coordinated
393 APOBEC3 signature mutations within a 10 kbp window) [5, 7, 66, 72]. No APOBEC3 signature
394 *kataegic* events were observed in control conditions. However, several APOBEC3 signature
395 *kataegic* events were evident in the genomic DNA of both A3A and A3B expressing
396 granddaughter clones (*e.g.*, **Figure 4C**). For instance, one A3A-attributable *kataegic* event was

397 comprised of 7 T(C>T/G)W mutations, and an A3B-attributable *kataegic* event included 8
398 T(C>T/G)W mutations. Interestingly, however, the frequency of *kataegic* events did not differ
399 significantly between A3A- and A3B-expressing granddaughter clones (A3A: median = 30, mean
400 = 35 events, SD = 8.3; A3B: median = 12, mean = 15, SD = 6.9; p = 0.28 by Welch's t-test). These
401 results show that both A3A and A3B can cause *kataegis* in a human cell line and, surprisingly, at
402 similar overall frequencies.
403
404 **APOBEC signature etiology in primary breast tumors**
405 Sequencing data from model systems such as HAP1-TK-M9 are powerful because the
406 resulting mutation signatures can help to establish cause-and-effect relationships for comparison
407 to more complex tumor WGS data sets to identify similarities and, potentially, to infer the precise
408 source of an observed mutation signature in individual tumors. We therefore performed an
409 unsupervised clustering analysis to compare the pentanucleotide cytosine mutation signatures
410 derived from sequencing A3A, A3B, and A3H expressing HAP1 clones and those from primary
411 breast tumors with WGS available through the ICGC data portal resource. This analysis revealed
412 three distinct tumor groups with respect to APOBEC3 signature mutations: 1) a group that showed
413 similarity to A3A-expressing HAP1-TK-M9 clones, 2) a group that showed similarity to A3B-
414 expressing HAP1-TK-M9 clones, and 3) a group that showed little to no significant APOBEC3
415 mutation signature (**Figure 5**). As expected from a lack of substantial APOBEC3 signature
416 mutations above in the *TK* reporter or in representative WGS, A3H-I expressing clones and
417 catalytic-inactive clones clustered with non-APOBEC3 signature tumors.
418 Both the A3A-like and A3B-like groups were comprised of tumors that show significant
419 levels of APOBEC3 signature mutations and correspondingly high enrichment scores (**Figure 5**
420 and **Supplementary Table S1**). However, only a small proportion of the A3A-like group of breast
421 tumors showed ≥71.7% APOBEC3 signature mutations in YTCW motifs, which is the overall
422 preference of A3A observed here in the HAP1-TK-M9 system. Similarly strong YTCW preferences

423 have also been reported recently for A3A expression in the chicken B cell line DT40 [73] and for
424 A3A-driven murine hepatocellular carcinomas [67]. These observations therefore suggest that
425 A3A alone may not account for the observed composite APOBEC3 mutation signature in
426 individual breast tumors and, moreover, that many tumors in A3A-like group may include
427 contributions from another A3 enzyme, most likely A3B. As regards the A3B-like group, most
428 tumors have higher percentages of C-to-T and C-to-G mutations in YTCW motifs than can be
429 explained by A3B alone. For instance, most tumors manifest a greater proportion of YTCW
430 mutations than observed here for A3B in the HAP1-TK-M9 system (47.2%) and for A3B-driven
431 tumors in mice (53%) [68]. These intrinsic preferences therefore combined to suggest that the
432 observed percentages of APOBEC3 signature mutations in most breast tumors may be a
433 composite resulting from the combinatorial activities of both A3A and A3B activity. In comparison,
434 the homologous recombination repair deficiency (HRD) signature (SBS3) appeared
435 underrepresented in both the A3A- and A3B-like tumor groups, and the ageing signature (SBS1)
436 occurred in all three groups regardless of the presence or absence of an APOBEC3 mutation
437 signature enrichment (**Figure 5**).
438

439 **Discussion**

440 Here we have reported the development and implementation of a genetic system to
441 investigate mutational processes in the human HAP1 cell line. Like many bacterial and yeast
442 model systems, the HAP1-TK-M9 system enabled a uniform cytotoxic selection with ganciclovir
443 such that only *TK* mutant cells survive. An analysis of clonally derived, A3A and A3B expressing
444 *TK* mutants by Sanger sequencing of high-fidelity PCR amplicons demonstrated a strong shift in
445 the mutational pattern from a variety of different base substitution mutations in control conditions
446 to a strongly 5'TCW-biased pattern (**Figure 2**). Interestingly, the *TK* mutation spectra inflicted by
447 A3A and A3B were very similar including two shared hotspots (Q8X, R212K) and no obvious
448 mesoscale features such as palindromic sequences capable of hairpin formation. This result may

449 be due to the limited number of mutable cytosines and TC motifs in *TK* that confer resistance to
450 ganciclovir (local base composition) and/or to selective pressure. Regardless of the precise
451 molecular explanation, an analogy can be drawn with the mutational spectrum of the *PIK3CA*
452 gene in breast, head/neck, cervical, and others cancers, which has two prominent APOBEC3
453 mutation hotspots (E542K, E545K) and no obvious hairpin structures [52, 74]. These observations
454 combined to suggest that selective pressure has the potential to overshadow the intrinsic
455 preferences of individual APOBEC3 enzymes and complicate assignment of direct cause-and-
456 effect relationships.

457 Drawing direct connections between A3A and/or A3B and a given mutation, even a
458 prominent hotspot, has been additionally challenging due to the fact that both enzymes can
459 deaminate DNA cytosines in linear substrates as well as single-stranded loop regions of stem-
460 loop structures and, importantly, rates can vary dramatically between different substrates (*e.g.*,
461 **Figure 4** and prior biochemical studies [52, 69, 70]). Thus, the *TK*-based system described here
462 is capable of yielding informative, rapid, and inexpensive mutation data sets with positive results
463 motivating genome-wide analyses where mutations are unselected and larger mutation data sets
464 enable broader analyses. For instance, our unbiased WGS analysis of A3A- and A3B-expressing
465 clones indicated that both enzymes are capable of deaminating the ssDNA loop regions of hairpin
466 substrates at similar frequencies (**Figure 4 – Figure Supplement 1**). This result was somewhat
467 anticipated by structural studies where both A3A and a mutant A3B catalytic domain were shown
468 to bind to ssDNA in a U-shaped conformation [16, 29], but it was also unexpected due to recent
469 reports describing strong RNA and DNA hairpin biases for A3A [52, 69]. Additional biochemical
470 and WGS studies will be required to confirm these results as well as extend them to additional
471 cell lines and experimental systems. One potential drawback of the HAP1-TK-M9 system is that
472 an exclusive focus on Gan^R clones might overestimate the mutational impact of a given process.
473 However, earlier work also used HAP1 cells to successfully characterize a variety of different
474 mutation sources, but these studies did leverage the power of a lethal genetic selection nor did

475 they address the enzyme(s) responsible for APOBEC3 signature mutations [75, 76].

476 Taken together, the HAP1-TK-M9 studies here have demonstrated unambiguously that
477 both A3A and A3B can inflict an APOBEC3 mutation signature in human genomic DNA with, in
478 both instances, ssDNA deamination events immortalizing predominantly as C-to-T and C-to-G
479 mutations in TCA and TCT trinucleotide motifs (**Figure 2**, **Figure 3**, and **Supplementary Table**
480 **S1**). Over identical month-long timeframes, A3A causes 4-fold more APOBEC3 signature
481 mutations in comparison to A3B (6070 vs 1528 mutations from 6 and 5 subclone WGSs,
482 respectively). This difference may be explained in part by super-physiological A3A expression
483 levels in the HAP1-TK-M9 system (at least 5-fold higher than levels in breast tumors or cell lines)
484 and in part by the higher intrinsic activity of this enzyme in comparison to A3B. Regardless, both
485 A3A and A3B inflicted thousands of TC-focused APOBEC3 signature mutations, which enabled
486 comparisons between extended intrinsic preferences. Most importantly, A3A has a strong
487 preference for a pyrimidine at the -2 position relative to the target cytosine (72.7% YTCW). This -
488 2 pyrimidine bias mirrors original results from human A3A expression in yeast [50, 51], as well as
489 recent WGS results from human A3A expression in the chicken B cell line DT40 (~70% YTCW
490 [73]) and from human A3A-induced murine hepatocellular carcinomas (70% YTCW [67]). Also,
491 similar to original studies in yeast [50, 51], human A3B showed a slight negative preference for
492 YTCW motifs (47.2% YTCW) and a corresponding enrichment for RTCW motifs (52.8% RTCW;
493 **Figure 3B**, **Figure 3 – Figure Supplement 4**, and **Supplementary Table S1**). In comparison,
494 recent studies showed that human A3B driven tumors in mice (hepatocellular carcinomas and B
495 cell lymphomas) exhibit an opposite APOBEC3 mutation signature bias with 53% YTCW and 47%
496 RTCW [68]. These differences in APOBEC3B local preferences in the different systems may be
497 due to genetic and/or epigenetic factors including but not limited to different base content,
498 chromatin states, DNA repair processes, and/or post-translational regulatory mechanisms. They
499 could also be due to simple stochastic variation and, accordingly, we hypothesize that A3B is non-
500 discriminatory at the -2 position relative to the target cytosine and that this may relate to the amino

501 acid composition of catalytic domain loop 1 residues relative to those of A3A (loop 3, loop 7, and
502 most other active site residues are identical or nearly identical).

503 Regardless of the precise molecular explanation(s) for differences between experimental
504 systems, the fact that both A3A and A3B can inflict YTCW mutations in human cells helps to
505 inform interpretations of the APOBEC3 enzyme responsible for the overall APOBEC3 mutation
506 program in cancer. For instance, based on comparisons of A3A- and A3B-attributable single base
507 substitution mutation signatures observed here in the HAP1-TK-M9 system and extracted
508 APOBEC3 mutation signatures from 784 breast cancer WGS, it is likely that neither enzyme's
509 preferred motif fully explains the composite signature in most individual tumors (**Figure 5** and
510 **Supplementary Table S1**). As regards A3A, the majority of individual breast tumors has ≤72.7%
511 APOBEC3 signature mutations in YTCW motifs, which is the overall preference of A3A here in
512 the HAP1-TK-M9 system suggesting the involvement of at least one other A3 enzyme. As regards
513 A3B, most individual breast tumors have a larger proportion of mutations in YTCW motifs than
514 can be explained by A3B alone (≥47.2%). Thus, it is likely that both A3A and A3B contribute to
515 the composite single base substitution mutation signature observed in individual APOBEC3
516 signature enriched breast tumors. However, we cannot exclude the possibility that some breast
517 tumors may be mutated exclusively by A3A or A3B due to factors listed above, or that some breast
518 tumors may have small mutagenic contributions from other APOBEC3 enzymes. Of course, A3B
519 is not a direct factor in *A3B*-null breast tumors, although the inherited deletion that removes all
520 *A3B* coding sequences may dysregulate *A3A* expression [45, 46]. In addition, the WGS data sets
521 for A3A- and A3B-expressing HAP1-TK-M9 clones described here may be useful for comparing
522 with APOBEC3 attributable events in other tumor types.

523 In addition to selective pressures and mesoscale features, additional factors are likely to
524 influence the APOBEC3-attributable fraction of an overall tumor mutational landscape including
525 whether A3A and/or A3B is expressed, expression levels, duration of expression, intrinsic activity,
526 and accessibility of chromosomal DNA (replication stress, R-loop levels, chromatin state, *etc.*).

527 With regards to studies here with the HAP1-TK-M9 system, both A3A and A3B were expressed
528 constitutively from the same promoter/construct for identical durations prior to ganciclovir
529 selection, A3A is intrinsically more active than A3B, A3A is cell-wide and A3B predominantly
530 nuclear, and yet these enzymes and cellular factors combined to yield remarkably similar *TK*
531 mutation frequencies and only a 4-fold difference in overall genome-wide SBS mutation level.
532 With respect to cancer, the *A3A* gene is expressed at lower levels than *A3B* in almost all cell lines
533 and tumors, A3A is cell-wide or predominantly cytoplasmic where A3B is constitutively nuclear,
534 and A3A has higher enzymatic activity that can vary from 2- to 100-fold above that of A3B
535 depending on substrate (*e.g.*, **Figure 4** and prior biochemical studies [11, 52]). It is therefore
536 notable here that the overall genome-wide level of APOBEC3 signature mutation from A3A is only
537 4-fold higher than that attributable to A3B. Endogenous *A3A* and *A3B* also have both distinct and
538 overlapping transcription programs, and both genes can be induced by a variety of conditions
539 including viral infection and inflammation [28, 33, 35-37, 39, 42, 61, 77-82]. *In vivo*, *A3A* and *A3B*
540 gene expression is also likely to be affected by the local tumor microenvironment, which can vary
541 both between and within cancer types, as well as by a patient's global state of health. Taken
542 together with unknown and likely lengthy multi-year durations of pre-cancer and early cancer
543 development prior to clinical manifestation, deducing the exact fractions of mutations attributable
544 to A3A and/or A3B may be a fruitless endeavor (except in *A3B*-null tumors). Rather, it may be
545 more prudent to focus on developing strategies to simultaneously diagnose and treat the
546 mutagenic contributions of both enzymes.

547 Independent whole genome sequencing experiments have provided additional information
548 on the APOBEC3 mutation process. Initial studies induced overexpression of A3B in 293-derived
549 cell lines, documented the resulting DNA damage responses, and performed WGS to assess
550 genome-wide associations [83, 84]. However, an unambiguous APOBEC3 mutation signature
551 was difficult to extract from these whole genome sequences due to large numbers of mutations
552 attributable to defective mismatch repair [83, 84]. A more recent study compared *de novo*

553 mutations occurring in APOBEC3 signature positive cell lines during multiple generations of clonal
554 outgrowth [85]. An intriguing finding from this work is that APOBEC3 signature mutations may be
555 able to occur in an episodic manner, accumulating in some generations and not others, consistent
556 with evidence discussed above that A3A and A3B expression can be induced by multiple signal
557 transduction pathways. Episodic mutagenesis, however, is unexpected in cell-based systems in
558 which continuous and relatively stochastic mutagenesis should predominate given defined media
559 and well-controlled growth conditions. These studies were followed-up more recently by WGS
560 comparisons of subclones of the same cancer cell lines CRISPR-engineered to lack *A3A*, *A3B*,
561 or both genes [86]. The results of over 250 WGS combined to indicate that A3A may be the source
562 of a significant fraction of observed APOBEC3 signature SBS mutations, A3B a smaller fraction,
563 and another as-yet-undefined APOBEC3 enzyme an additional minor fraction. These data are
564 complementary to the major results here, with both A3A and A3B proving capable of generating
565 genome-wide APOBEC3 signature single base substitution mutations. Differences in the overall
566 magnitude of A3A vs A3B mutagenesis may be due to the factors described above including
567 differential intrinsic activity, protein expression levels, genomic DNA accessibility, cell culture
568 conditions, and importantly durations of mutagenesis. Both studies were necessarily done in
569 model cellular systems, each with obvious strengths, but neither capable of fully recapitulating
570 the wide repertoire of factors that impact the actual pre- and post-transformation environments *in*
571 *vivo*, including anti-tumor immune responses, which are further likely to vary between different
572 tissue types, tumor types, and patients.

573 A role for A3H, haplotypes I or II, in cancer is disfavored by our results here showing that
574 these variants are incapable of eliciting DNA damage responses or increasing the *TK* mutation
575 frequency. Two A3H-I expressing *TK* mutant clones were subjected to WGS and no APOBEC3
576 mutation signature was evident. In addition, no specific evidence for A3H emerged from
577 sequencing clonally-derived cancer cell lines [86]. However, all of these cell-based studies have
578 limitations as discussed above and have yet to fully eliminate A3H as a source of APOBEC3

579 signature mutations in cancer. For instance, A3H-I may take more time to inflict detectable levels
580 of mutation, it may be subject to different transcriptional and post-transcriptional regulatory
581 processes, and/or it may only be mutagenic in a subset of cancer types subject to different
582 stresses and different selective pressures.

583 Ultimately, the studies here showed that A3A and A3B are each individually capable of
584 inflicting a robust APOBEC3 mutation signature in human cells and, taken together with other
585 work summarized above, support a model in which both of these enzymes contribute to the
586 composite APOBEC3 mutation signature reported in many different tumor types. This conclusion
587 is supported by clinical studies implicating A3A and/or A3B in a variety of different tumor
588 phenotypes including drug resistance/susceptibility, metastasis, and immune responsiveness [9,
589 20, 21, 27, 32, 42, 47, 87-93]. Thus, efforts to diagnose and treat APOBEC3 signature-positive
590 tumors should take both enzymes into account, not simply one or the other. Such longer-term
591 goals are not trivial given the high degree of identity between A3A and the A3B catalytic domain
592 (>90%), the related difficulty of developing specific and versatile antibodies for detecting each
593 enzyme, and the fact that each can be regulated differentially by a wide variety of common factors
594 including virus infection and inflammation. Thus, we are hopeful that the HAP1-TK-M9 system,
595 the whole genome sequences, and the A3A-specific rabbit monoclonal antibody described here
596 will help to expedite the achievement of these goals.

597

598 **Materials and Methods**

599 **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 that expresses HSV-1 *TK* and is Gan^{S}; request by contacting RSH
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	
Cell line (*Homo sapiens*, female)	HeLa	NIH ARP	ARP-153	
Antibody	Anti-β-actin (mouse monoclonal)	Millipore Sigma	Cat #: A1978	WB (1:5000)
Antibody	Anti-β-actin (rabbit monoclonal)	Cell Signaling Technology	Cat#:4970	WB (1:5000)
Antibody	Anti-A3A (rabbit monoclonal)	This study	UMN-13	WB (1:1500); request by contacting RSH
Antibody	Anti-A3B (rabbit monoclonal)	[94]	RRID:AB_2721202	WB (1:1,000) IF (1:300); request by contacting RSH

Antibody	Anti-A3H (mouse monoclonal)	[58]	Clone P1-D8	WB (1:1,000); request by contacting RSH
Antibody	Anti-A3H (rabbit polyclonal)	Novus ARP10	Cat#: NBP1-91682 RRID: AB_2057523	IF (1:300)
Antibody	Anti-γ-H2AX (S139) (rabbit polyclonal)	Millipore Sigma JW301	Cat#: H5912 RRID: AB_310406	IF (1:500)
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

600

601

602 **Cell lines and culture conditions**

603 All cell lines were cultured at 37°C under 5% CO_2. HAP1 cells (Horizon) and derivatives
604 (this study) were grown in IMDM (Invitrogen) supplemented with 10% fetal bovine serum (Sigma),
605 penicillin (100 U/mL), and streptomycin (100 µg/mL). 293T and HeLa cells were cultured in DMEM

606 (Invitrogen) with 10% fetal bovine serum (Sigma), penicillin (100 U/mL), and streptomycin (100
607 µg/mL). THP1 cells were cultured in RPMI (ThermoFisher) supplemented with 10% fetal bovine
608 serum (Sigma), penicillin (100 U/mL), and streptomycin (100 µg/mL). TransIT-LT1 (Mirus) was
609 used for all transfections. All parent and clonal lineage cell lines tested negative for mycoplasma
610 using a PCR-based assay [95]. Puromycin (ThermoFisher) and G418 (Gold Biotechnology) were
611 used at 1 µg/mL and 1 mg/mL, respectively. Ganciclovir (ThermoFisher) was used at 5 µM to
612 select *TK* mutant clones.
613
614 **HAP1-TK-M9 system**
615 The HAP1-TK-M9 system was generated by co-transfecting HAP1 parent cells with a
616 plasmid expressing the Sleeping Beauty transposase and a separate plasmid with *TK-Neo* coding
617 sequences flanked by SB recognition sites [9, 96]. Semi-confluent cells in 6 well plates were
618 transfected, treated 24 hrs later with G418 (1 mg/mL), and subcloned by limiting dilution in 96 well
619 plates to create single cell derivatives. Single cell clones were then expanded and characterized
620 as described in the main text. The *A3H* genotype was determined by Sanger sequencing exon-
621 specific PCR amplicons [97, 98] and further confirmed by WGS (below).
622 Standard molecular cloning procedures were used to create derivatives of MLV pQCXIP
623 for expressing each A3 protein. First, pQCXIP was cut with MluI and PacI to excise the strong
624 CMV promoter and replace it with a weaker MND promoter (a synthetic promoter containing
625 regions of both the MLV LTR and the myeloproliferative sarcoma virus enhancer). Second, this
626 new construct was cut with SfiI and BsiWI to insert intron-containing A3 coding sequences [67].
627 This was done for *A3A*, *A3B*, *A3H* haplotype-I, *A3H* haplotype-II, and appropriate catalytic mutant
628 derivatives (E-to-A). An eGFP expressing construct was generated in parallel to use as a control
629 in various experiments. All new constructs were confirmed by Sanger sequencing and functional
630 assays as described in the main text.

631 Each MLV-based construct was co-transfected into 293T cells with appropriate packaging
632 vectors and 48 hrs later the resulting viral supernatants were filtered (0.2 μm) and used to
633 transduce semi-confluent HAP1-TK-M9 cells (MOI < 0.1). After 48 hrs incubation, transduced
634 cells were selected with puromycin (1 μg/mL) and subcloned by limiting dilution to obtain A3
635 expressing daughter clones. Each daughter clone expressed only a single, integrated construct,
636 which was anticipated by low MOI transduction and verified by mapping insertion sites for 5
637 representative daughter clones (**Figure 3 – Figure Supplement 3**). These A3 expressing and
638 control daughter clones were expanded for 1 month and characterized as described in the main
639 text. No overt growth/proliferation defects were noted, and all granddaughter clones expanded at
640 similar rates. Mutation frequencies were determined by plating 5 x 10^6 cells in 96 well flat bottom
641 plates, treating with 5 μM ganciclovir (ThermoFisher), and after 14 days incubation counting the
642 number of *TK* mutant colonies that survived selection. Single Gan^R granddaughter clones were
643 counted using a light microscope and expanded and characterized as described in the main text
644 including immunoblotting, DNA deaminase activity assays, *TK* sequencing, and WGS.
645
646 **Immunoblots (IB)**
647 Cells were treated with trypsin EDTA and collected, washed in 1X PBS, and re-suspended
648 in 100 μL of reducing sample buffer per one million cells [0.5 M Tris-HCl pH 6.8, 1% 2-
649 mercaptoethanol, 10% sodium dodecyl sulfate (SDS), 50% glycerol]. Proteins were denatured by
650 boiling samples for 20 min and resolved by SDS-polyacrylamide gel electrophoresis. Proteins
651 were transferred to a PVDF-FL membrane (Millipore Sigma) and blocked in 5% milk in 1X PBS.
652 Primary and secondary antibodies were incubated in blocking buffer, with the addition of 0.2%
653 SDS for fluorescent antibodies. The primary antibodies used were anti-A3A (UMN-13, IB: 1:1500,
654 IF: 1:300 [this study]), anti-A3A/B (5210-87-13, IB: 1:1,000, IF: 1:300 [94]), anti-A3H (P1-D8, IB:
655 1:300 [58]), anti-Tubulin (Sigma Aldrich, 1:20,000), and anti-β-Actin (Millipore Sigma, mouse mAb,
656 1:5000 or Cell Signaling Technology, rabbit mAb, 1:5000, as indicated). The secondary antibodies

657 used were anti-rabbit HRP-linked (CST 7074, 1:2,000), IRDye 800CW goat anti-mouse (LI-COR
658 Biosciences, 1:10,000), Alexa Fluor 680 goat anti-rabbit (Molecular Probes, Eugene, OR, USA,
659 1:10,000), and Alexa Fluor 680 goat anti-mouse (Molecular Probes, Eugene, OR, USA, 1:10,000).
660 Membranes were imaged using a LI-COR Odyssey instrument or LI-COR Odyssey-Fc instrument
661 for HRP visualization (LI-COR Biosciences).
662
663 **DNA deaminase activity assays**
664 Whole cell extract (WCE) assays: ssDNA deamination activities were measured using
665 WCE prepared using 100 µL HED lysis buffer per 1 million cells (25 mM HEPES, 15 mM EDTA,
666 10% Glycerol, 1 mM DTT, and 1 protease inhibitor tablet [Roche]). Samples were sonicated in a
667 water bath sonicator to ensure complete lysis. A3-containing lysates were incubated at 37°C for
668 1 hr (A3A and A3B) or 4 hrs (A3H) with purified human UNG2 and a ssDNA substrate containing
669 either a single TCA or a single TCT trinucleotide motif (5'-
670 ATTATTATTATTCAAATGGATTTATTTATTTATTTATTTATTT-FAM [A3A and A3B]; 5'-
671 ATTATTATTATTCTAATGGATTTATTTATTTATTTATTTATTT-FAM [A3H]) following established
672 protocols [9, 99]. After this initial incubation, the reaction was treated with 100 mM NaOH for 5
673 min at 95°C. The reaction was run out on a 15% TBE-urea acrylamide gel to separate substrate
674 oligo from cleaved product oligo and imaged on a Typhoon FLA 7000 with ImageQuant TL 8.2.0
675 software (GE Healthcare).
676 Recombinant enzyme assays: A3A- and A3B-mycHis were prepared from transfected
677 293T cells as reported [14, 16, 100-102]. Single hit kinetics were ensured by incubating 25 nM of
678 each protein with 800 nM substrate in reaction buffer (25 mM HEPES pH 7.4, 50 mM NaCl, 5 mM
679 imidazole) for the indicated times at 37°C. Reactions were stopped by freezing in liquid nitrogen
680 and then were heated to 95°C to denature the enzymes. Reactions were then treated with 0.5
681 U/reaction uracil DNA glycosylase (NEB, USA) for 10 min at 37°C. The resulting abasic sites were
682 cleaved by incubation with 100 mM NaOH and heating to 95°C for 5 min. Products were separated

683 by 20% TBE-Urea PAGE, imaged on a Typhoon FLA-7000 (GE Healthcare, USA), and quantified
684 using ImageQuant TL 8.2.0 software (GE Healthcare). Deamination of the target cytosine was
685 calculated by dividing the total reaction product by the total amount of starting substrate. The
686 oligonucleotide substrates were analyzed by both denaturing and native 20% PAGE to determine
687 the extent of hairpin formation. The oligos (and a previously reported NUP93-noHP oligo) were
688 heated to 70°C and slowly cooled to 37°C in HEPES buffer as above. 1 pmol of each oligo was
689 then mixed with agarose gel loading dye (30% Ficoll 400 in 1x TAE, xylene cyanol, bromophenol
690 blue) or with DNA PAGE loading dye (80% formamide in 1x TBE, xylene cyanol, bromophenol
691 blue) and separated by native and denaturing PAGE, respectively.
692
693 **DNA content by flow cytometry**
694 Propidium iodide (PI) staining was used to assess the ploidy of HAP1 clones relative to
695 THP1 (ATCC, Cat#: TIB-202) as a confirmed diploid control [103]. Cells were trypsinized and
696 suspended in 100 µL of 1X PBS per 1 million cells. Then, 500 µL ice-cold ethanol was added to
697 cell suspensions and incubated at -20°C for 1 hr to fix the cells. After fixation, cells were pelleted
698 and washed in 1X PBS. Cells were finally suspended in 500 µL of FxCycle PI stain (Invitrogen)
699 and incubated for 30 min at RT in the dark to stain the cells. Cells were spun down and
700 resuspended in 300 µL of the PI stain solution and placed in a 96 well round bottom plate for flow
701 cytometry analysis using a BD LSRFortessa flow cytometer (with high-throughput 96 well adapter
702 system). A minimum of ten thousand events were acquired for each condition.
703
704 **RT-qPCR**
705 Total RNA was extracted using the High Pure RNA isolation kit (Roche). cDNA was
706 synthesized using SuperScript First-Strand RT (ThermoFisher). Quantification of mRNA was
707 done using validated primer sets for all human *A3* genes relative to the housekeeping gene *TBP*
708 [9, 78, 96, 97]. All RT-qPCR reaction were performed using SsoFast SYBR Green mastermix

709 (Bio-Rad) in 384-well plates on a LightCycler 480 (Roche) following the manufacturer's protocol.
710 Statistical analyses were done using GraphPad Prism 6 and R.
711
712 **Immunofluorescent (IF) microscopy**
713 IF microscopy was conducted as described [104, 105]. As a positive control for DNA
714 damage, 4 μM cisplatin (*cis*-diamminedichloroplatinum II, Selleck Chemical) was incubated with
715 cells for 24 hrs. Cells were grown at low density in 4-chamber, tissue culture-treated glass slides
716 (Falcon) prior to fixation. The cells were then fixed with 4% paraformaldehyde in PBS for 15 min
717 at room temperature. Permeabilization followed using PBS containing 0.2% Triton X-100 (Sigma
718 Aldrich), before rinsing with PBS. The cells were blocked using an IF blocking solution (0.1% triton
719 X-100, 5% goat serum in PBS) for 1 hr at room temperature, then incubated with primary antibody
720 overnight at 4°C. The primary antibodies used were anti-A3A/B/G (5210-87-13, 1:300) [94], anti-
721 A3A (UMN-13, 1:300 [this study]), anti-A3H (Novus, 1:300) [58], and anti-γ-H2AX (JBW301,
722 Millipore Sigma, 1:500). Following this, cells were washed with PBS and incubated with a
723 fluorophore-conjugated secondary antibody for 1 hr at room temperature in the dark. The
724 secondary antibodies used were Alexa Fluor 488 goat anti-mouse (Invitrogen 1:1,000) and Alexa
725 Fluor 594 goat anti-rabbit (Invitrogen, 1:1,000). Both primary and secondary antibodies were
726 diluted in blocking buffer. Hoechst 33342 (Mirus) was used at a final concentration of 1 μg/mL to
727 stain nuclei for another 15 min, before the slides were washed three times with PBS and mounted
728 using antifade mounting media (Cell Signaling 9071). Images were captured at 60x magnification
729 using a Nikon ECLIPSE Ti2 microscope. The number of γ-H2AX foci per cell were counted for at
730 least 50 cells per condition for each experiment. The nuclear intensity of the γ-H2AX signal and
731 the cellular intensity of each A3 were measured by ImageJ2 software (2.9.0/1.53t). Statistical
732 analyses were conducted in GraphPad Prism 9.
733
734 **Alkaline comet assays**

735 HeLa cells transfected with individual A3 or control constructs were harvested 24 hrs post-
736 transfection and resuspended in ice-cold 1X phosphate-buffered saline (PBS, Ca^{2+} and Mg^{2+} free,
737 at a density of 10^5 cells/mL). As a positive control, mock-transfected cells were treated with 2 μM
738 camptothecin (CPT, Millipore Sigma) for 2 hrs prior to harvesting as above. The CometAssay ESII
739 kit was used for all alkaline Comet assays, following the manufacturer's protocol (BioTechne).
740 Cells were resuspended in low-melt agarose and spread at low density on a glass slide to secure
741 the cells in place for lysis. DNA unwinding and electrophoresis were performed according to the
742 manufacturer's alkaline protocol. Comet tail moments were measured for at least 50 cells per
743 condition using the OpenComet plugin for Image J [106]. Statistical analyses were done using
744 GraphPad Prism 9.
745
746 ***TK* sequencing**
747 Cells were harvested and genomic DNA was isolated using the Puregene DNA isolation
748 protocol. The *TK* cassette was amplified from genomic DNA using 5'-
749 ATCTTGGTGGCGTGAAACTC and 5'-CTTCCGGTATTGTCTCCTTCC. PCR products were
750 cleaned-up using the GeneJet Gel Extraction Kit (Thermo Scientific) and Sanger sequenced with
751 4 different primers to cover the full open reading frame (5'-ATCTTGGTGGCGTGAAACTC, 5'-
752 GGTCATGCTGCCCATAAGGTA, 5'-CCGTTCTGGCTCCTCATATC, and 5'-
753 CTTCCGGTATTGTCTCCTTCC.
754
755 **Whole genome sequencing (WGS) and analyses**
756 Genomic DNA was prepared from cell pellets (1 million cells) using Allprep DNA/RNA mini
757 kit (Qiagen). Whole genome libraries were sequenced 150x2 bp on a NovaSeq 6000 (Illumina) to
758 a target read depth of 30X coverage for all granddaughter clones as well as the parental HAP1-
759 TK-M9 mother clone. Resulting sequences were aligned to the human genome (hg38) using
760 SpeedSeq [107], which relies on the Burrows-Wheeler Aligner, BWA (version 0.7.17). PCR

761 duplicates were removed using Picard (version 2.18.16). Reads were locally realigned around
762 INDELs using GATK3 (version 3.6.0) tools RealignerTargetCreator to create intervals, followed
763 by IndelRealigner on the aligned bam files. Single base substitutions and small INDELs were
764 called in each clone relative to the bam file generated from the HAP1-TK-M9 mother clone using
765 Mutect2 from GATK3 (version 3.6.0). SBSs that passed the internal GATK3 filter with minimum 4
766 reads supporting each variant, minimum 20 total reads at each variant site and a variant allele
767 frequency over 0.05 were used for downstream analysis. SBSs were analyzed in R (version 4.0.5)
768 using the MutationalPatterns [108] and deconstructSigs R packages (version 1.8.0 [109]). All
769 visualizations were generated using the ggplot2 package (version 3.3.5). The indel landscapes
770 were generated using the MutationalPatterns R package [108] following PCAWG definitions [1].
771 All individual clone data from each condition were pooled for presentation.
772 COSMIC single base substitution mutation signatures (v3 – May 2019
773 https://cancer.sanger.ac.uk/cosmic/signatures/SBS/) were obtained from
774 https://www.synapse.org/#!Synapse:syn11738319. *De-novo* non-negative matrix factorization of
775 mutational signatures was performed with the "extract_signature" command from the
776 MutationalPatterns package, with a rank of 2 and 100 iterations. TCW mutation enrichment scores
777 were calculated as described [50, 65]. Sequence logos of -2 to +2 sequence surrounding C-to-T
778 mutations were created using the ggseqlogo (version 0.1) package. Putative hairpin structures
779 were predicted using the nBMST tool [110] and human genome GRCH38 with a minimum stem
780 length of 6 bases. The loop criteria established by Buisson *et al*. with ssDNA loop lengths of 3 to
781 11 nucleotides were used to search for APOBEC3 signature mutations in these regions [52].
782

783 **APOBEC expression and mutation signature analyses in TCGA and ICGC data sets**

784 TCGA primary breast tumors represented by both RNA-seq and whole exome sequencing
785 were downloaded from the Firehose GDAC resource through the Broad Institute pipeline
786 (http://gdac.broadinstitute.org/) for multiple tumor tissue types. APOBEC3 mutation signatures

787 were determined as described [5, 65] using the deconstructSigs R package [109]. APOBEC3
788 mutation enrichment scores were calculated using the hg19 reference genome and published
789 methods [50]. Enrichment score significance was assessed using a Fisher exact test with
790 Benjamini-Hochberg false discovery rate (FDR) correction. All data analyses and visualizations
791 were conducted using R and the ggplot2 package (https://www.R-project.org/).
792
793 ***TK* integration site determination**
794 To determine the integration site of the single copy *TK-Neo* construct, a *TK* reference
795 sequence was provided as an additional chromosome during alignment of the WGS reads to the
796 reference genome (hg38). Reads that mapped to this region were then categorized as discordant
797 and realigned to hg38 using GRIDSS (v2.2) [111] to determine the site of integration.
798
799 **Clustering analysis**
800 All primary breast tumor whole genome sequencing variant information from International
801 Cancer Genome Consortium (ICGC) was downloaded from the ICGC data portal
802 (https://dcc.icgc.org). SBSs used in these analyses included only C-to-T variants in TC
803 dinucleotide contexts (TCA, TCC, and TCT) and excluded all mutations in CG motifs due to
804 potential overlap with spontaneous water-mediated methyl-C deamination. SBSs meeting these
805 inclusion criteria from all clones expressing A3A, A3B, A3A-E72A, A3B-E255A, and A3H-I were
806 pooled per condition for this analysis. A matrix comprised of the number of mutations within a
807 pentanucleotide across all samples within a cancer type was generated, and counts were
808 normalized to frequency within each cancer type. The resulting matrix was then clustered using
809 the hclust function in R with the classical Euclidean distance as the distance method for clustering,
810 which was then plotted as dendrograms. Mutation signatures were calculated using
811 deconstructSigs as described above.
812

813 **Data and code availability**

814 All HAP1-TK-M9 clone alignment files (FASTQ and BAM format) are available through the
815 Sequence Read Archive under the BioProject accession number PRJNA832427.

816

817 **Author contributions**

818 R.S. Harris conceptualized and designed the overall project. M.A. Carpenter, M.A.
819 Ibrahim, M.C. Jarvis, M.R. Brown, P.P. Argyris, and W.L. Brown performed experiments. M.C.
820 Jarvis, M.A. Carpenter, M.A. Ibrahim, N.A. Temiz, M.R. Brown, P.P. Argyris, and W.L. Brown did
821 formal data analysis. R.S. Harris, M.C. Jarvis, and D. Yee contributed to funding acquisition. R.S.
822 Harris and M.C. Jarvis drafted the manuscript, and all authors except M.C. Jarvis contributed to
823 manuscript proofing and revision.

824

825 **Additional contributions**

826 We thank Justin Leung, Eloïse Dray, Weixing Zhao, and Wenjing Li for expert assistance
827 with IF microscopy and comet assays, Arad Moghadasi and Sofia Moraes for initial technical
828 support, and Kyle Richards for helping to set-up spectral karyotype analysis. We also thank the
829 UMN Cytogenetics Core for spectral karyotyping and Scott McIvor for sharing a plasmid construct
830 with the *MND* promoter. The results presented here are in part based upon data generated by the
831 TCGA Research Network: http://www.cancer.gov/tcga. Additionally, data from the International
832 Cancer Genome Consortium (ICGC) were used in these analyses: https://dcc.icgc.org/.

833

834 **Additional information**

835 **Funding**

836 These studies were supported by NCI P01-CA234228, NCI P50-CA247749, and a
837 Recruitment of Established Investigators Award from the Cancer Prevention and Research
838 Institute of Texas (CPRIT RR220053). Salary support for MCJ was provided in part by NCI T32-

839 CA009138 and subsequently NCI F31-CA243306. RSH is an Investigator of the Howard Hughes
840 Medical Institute and the Ewing Halsell President's Council Distinguished Chair at University of
841 Texas Health San Antonio.

842

843 **Compliance with ethical standards**

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

845

846 **References**

847 1. Alexandrov LB, Kim J, Haradhvala NJ, Huang MN, Tian Ng AW, Wu Y, et al. The repertoire
848 of mutational signatures in human cancer. Nature. 2020;578(7793):94-101. Epub
849 2020/02/07. doi: 10.1038/s41586-020-1943-3. PubMed PMID: 32025018; PubMed
850 Central PMCID: PMCPMC7054213.
851 2. Venkatesan S, Rosenthal R, Kanu N, McGranahan N, Bartek J, Quezada SA, et al.
852 Perspective: APOBEC mutagenesis in drug resistance and immune escape in HIV and
853 cancer evolution. Annals of Oncology. 2018;29(3):563-72. doi:
854 10.1093/annonc/mdy003. PubMed PMID: 29324969; PubMed Central PMCID:
855 PMC5888943.
856 3. Petljak M, Maciejowski J. Molecular origins of APOBEC-associated mutations in cancer.
857 DNA Repair (Amst). 2020;94:102905. Epub 2020/08/21. doi:
858 10.1016/j.dnarep.2020.102905. PubMed PMID: 32818816; PubMed Central PMCID:
859 PMCPMC7494591.
860 4. Koh G, Degasperi A, Zou X, Momen S, Nik-Zainal S. Mutational signatures: emerging
861 concepts, caveats and clinical applications. Nature reviews. 2021;21(10):619-37.
862 Epub 2021/07/29. doi: 10.1038/s41568-021-00377-7. PubMed PMID: 34316057.
863 5. Alexandrov LB, Nik-Zainal S, Wedge DC, Aparicio SA, Behjati S, Biankin AV, et al.
864 Signatures of mutational processes in human cancer. Nature. 2013;500(7463):415-

865 21. Epub 2013/08/16. doi: 10.1038/nature12477 [pii]. PubMed PMID: 23945592.
866 6. Roberts SA, Lawrence MS, Klimczak LJ, Grimm SA, Fargo D, Stojanov P, et al. An APOBEC
867 cytidine deaminase mutagenesis pattern is widespread in human cancers. Nat Genet.
868 2013;45(9):970-6. doi: 10.1038/ng.2702. PubMed PMID: 23852170; PubMed Central
869 PMCID: PMC3789062.
870 7. Nik-Zainal S, Alexandrov LB, Wedge DC, Van Loo P, Greenman CD, Raine K, et al.
871 Mutational processes molding the genomes of 21 breast cancers. Cell.
872 2012;149(5):979-93. Epub 2012/05/23. doi: S0092-8674(12)00528-4 [pii]
873 10.1016/j.cell.2012.04.024. PubMed PMID: 22608084.
874 8. Nik-Zainal S, Davies H, Staaf J, Ramakrishna M, Glodzik D, Zou X, et al. Landscape of
875 somatic mutations in 560 breast cancer whole-genome sequences. Nature.
876 2016;534(7605):47-54. doi: 10.1038/nature17676. PubMed PMID: 27135926;
877 PubMed Central PMCID: PMC4910866.
878 9. Burns MB, Lackey L, Carpenter MA, Rathore A, Land AM, Leonard B, et al. APOBEC3B is
879 an enzymatic source of mutation in breast cancer. Nature. 2013;494(7437):366-70.
880 Epub 2013/02/08. doi: 10.1038/nature11881 [pii]. PubMed PMID: 23389445.
881 10. Leonard B, Hart SN, Burns MB, Carpenter MA, Temiz NA, Rathore A, et al. APOBEC3B
882 upregulation and genomic mutation patterns in serous ovarian carcinoma. Cancer
883 Res. 2013;73(24):7222-31. Epub 2013/10/25. doi: 10.1158/0008-5472.CAN-13-1753.
884 PubMed PMID: 24154874; PubMed Central PMCID: PMC3867573.
885 11. Ito F, Fu Y, Kao SA, Yang H, Chen XS. Family-wide comparative analysis of cytidine and
886 methylcytidine deamination by eleven human APOBEC proteins. J Mol Biol.
887 2017;429(12):1787-99. doi: 10.1016/j.jmb.2017.04.021. PubMed PMID: 28479091;
888 PubMed Central PMCID: PMC5530319.
889 12. Ito F, Yang H, Xiao X, Li SX, Wolfe A, Zirkle B, et al. Understanding the structure,
890 multimerization, subcellular localization and mC selectivity of a genomic mutator and

891 anti-HIV factor APOBEC3H. Sci Rep. 2018;8(1):3763. doi: 10.1038/s41598-018-
892 21955-0. PubMed PMID: 29491387; PubMed Central PMCID: PMC5830531.
893 13. Hultquist JF, Lengyel JA, Refsland EW, LaRue RS, Lackey L, Brown WL, et al. Human and
894 rhesus APOBEC3D, APOBEC3F, APOBEC3G, and APOBEC3H demonstrate a
895 conserved capacity to restrict Vif-deficient HIV-1. J Virol. 2011;85(21):11220-34. Epub
896 2011/08/13. doi: JVI.05238-11 [pii] 10.1128/JVI.05238-11. PubMed PMID: 21835787;
897 PubMed Central PMCID: PMC3194973.
898 14. Carpenter MA, Li M, Rathore A, Lackey L, Law EK, Land AM, et al. Methylcytosine and
899 normal cytosine deamination by the foreign DNA restriction enzyme APOBEC3A. J
900 Biol Chem. 2012;287(41):34801-8. Epub 2012/08/17. doi: 10.1074/jbc.M112.385161
901 [pii]. PubMed PMID: 22896697; PubMed Central PMCID: PMC3464582.
902 15. Suspène R, Aynaud MM, Vartanian JP, Wain-Hobson S. Efficient deamination of 5-
903 methylcytidine and 5-substituted cytidine residues in DNA by human APOBEC3A
904 cytidine deaminase. PLoS One. 2013;8(6):e63461. Epub 2013/07/11. doi:
905 10.1371/journal.pone.0063461 PONE-D-13-02127 [pii]. PubMed PMID: 23840298;
906 PubMed Central PMCID: PMC3688686.
907 16. Shi K, Carpenter MA, Banerjee S, Shaban NM, Kurahashi K, Salamango DJ, et al. Structural
908 basis for targeted DNA cytosine deamination and mutagenesis by APOBEC3A and
909 APOBEC3B. Nat Struct Mol Biol. 2017;24(2):131-9. doi: 10.1038/nsmb.3344. PubMed
910 PMID: 27991903; PubMed Central PMCID: PMC5296220.
911 17. Schutsky EK, Nabel CS, Davis AKF, DeNizio JE, Kohli RM. APOBEC3A efficiently
912 deaminates methylated, but not TET-oxidized, cytosine bases in DNA. Nucleic Acids
913 Res. 2017;45(13):7655-65. Epub 2017/05/05. doi: 10.1093/nar/gkx345. PubMed
914 PMID: 28472485; PubMed Central PMCID: PMCPMC5570014.
915 18. Alexandrov LB, Jones PH, Wedge DC, Sale JE, Campbell PJ, Nik-Zainal S, et al. Clock-like
916 mutational processes in human somatic cells. Nat Genet. 2015;47(12):1402-7. Epub

917 2015/11/10. doi: 10.1038/ng.3441. PubMed PMID: 26551669; PubMed Central
918 PMCID: PMCPMC4783858.
919 19. Burns MB, Temiz NA, Harris RS. Evidence for APOBEC3B mutagenesis in multiple human
920 cancers. Nat Genet. 2013;45(9):977-83. Epub 2013/07/16. doi: 10.1038/ng.2701 [pii].
921 PubMed PMID: 23852168.
922 20. Sieuwerts AM, Schrijver WA, Dalm SU, de Weerd V, Moelans CB, Ter Hoeve N, et al.
923 Progressive APOBEC3B mRNA expression in distant breast cancer metastases.
924 PLoS One. 2017;12(1):e0171343. doi: 10.1371/journal.pone.0171343. PubMed
925 PMID: 28141868; PubMed Central PMCID: PMC5283735 sharing data and materials.
926 21. Sieuwerts AM, Willis S, Burns MB, Look MP, Meijer-Van Gelder ME, Schlicker A, et al.
927 Elevated APOBEC3B correlates with poor outcomes for estrogen-receptor-positive
928 breast cancers. Hormones & Cancer. 2014;5(6):405-13. doi: 10.1007/s12672-014-
929 0196-8. PubMed PMID: 25123150; PubMed Central PMCID: PMC4228172.
930 22. Angus L, Smid M, Wilting SM, van Riet J, Van Hoeck A, Nguyen L, et al. The genomic
931 landscape of metastatic breast cancer highlights changes in mutation and signature
932 frequencies. Nat Genet. 2019;51(10):1450-8. Epub 2019/10/02. doi: 10.1038/s41588-
933 019-0507-7. PubMed PMID: 31570896; PubMed Central PMCID: PMCPMC6858873.
934 23. Moore L, Cagan A, Coorens THH, Neville MDC, Sanghvi R, Sanders MA, et al. The
935 mutational landscape of human somatic and germline cells. Nature.
936 2021;597(7876):381-6. Epub 2021/08/27. doi: 10.1038/s41586-021-03822-7. PubMed
937 PMID: 34433962.
938 24. Harris RS, Dudley JP. APOBECs and virus restriction. Virology. 2015;479-480C:131-45.
939 doi: 10.1016/j.virol.2015.03.012. PubMed PMID: 25818029; PubMed Central PMCID:
940 PMC4424171.
941 25. Simon V, Bloch N, Landau NR. Intrinsic host restrictions to HIV-1 and mechanisms of viral
942 escape. Nat Immunol. 2015;16(6):546-53. doi: 10.1038/ni.3156. PubMed PMID:

943 25988886.
944 26. Silvas TV, Schiffer CA. APOBEC3s: DNA-editing human cytidine deaminases. Protein Sci.
945 2019;28(9):1552-66. Epub 2019/06/27. doi: 10.1002/pro.3670. PubMed PMID:
946 31241202; PubMed Central PMCID: PMCPMC6699113.
947 27. Cortez LM, Brown AL, Dennis MA, Collins CD, Brown AJ, Mitchell D, et al. APOBEC3A is a
948 prominent cytidine deaminase in breast cancer. PLoS Genet. 2019;15(12):e1008545.
949 Epub 2019/12/17. doi: 10.1371/journal.pgen.1008545. PubMed PMID: 31841499.
950 28. Westrich JA, Warren CJ, Klausner MJ, Guo K, Liu CW, Santiago ML, et al. Human
951 papillomavirus 16 E7 stabilizes APOBEC3A protein by inhibiting Cullin 2-dependent
952 protein degradation. J Virol. 2018;92(7). Epub 2018/01/26. doi: 10.1128/JVI.01318-17.
953 PubMed PMID: 29367246; PubMed Central PMCID: PMCPMC5972886.
954 29. Kouno T, Silvas TV, Hilbert BJ, Shandilya SMD, Bohn MF, Kelch BA, et al. Crystal structure
955 of APOBEC3A bound to single-stranded DNA reveals structural basis for cytidine
956 deamination and specificity. Nat Commun. 2017;8:15024. Epub 2017/04/30. doi:
957 10.1038/ncomms15024. PubMed PMID: 28452355; PubMed Central PMCID:
958 PMCPMC5414352.
959 30. Green AM, Landry S, Budagyan K, Avgousti DC, Shalhout S, Bhagwat AS, et al.
960 APOBEC3A damages the cellular genome during DNA replication. Cell Cycle.
961 2016;15(7):998-1008. doi: 10.1080/15384101.2016.1152426. PubMed PMID:
962 26918916.
963 31. Chen H, Lilley CE, Yu Q, Lee DV, Chou J, Narvaiza I, et al. APOBEC3A is a potent inhibitor
964 of adeno-associated virus and retrotransposons. Curr Biol. 2006;16(5):480-5. PubMed
965 PMID: 16527742.
966 32. Faden DL, Thomas S, Cantalupo PG, Agrawal N, Myers J, DeRisi J. Multi-modality analysis
967 supports APOBEC as a major source of mutations in head and neck squamous cell
968 carcinoma. Oral Oncology. 2017;74:8-14. doi: 10.1016/j.oraloncology.2017.09.002.

969 PubMed PMID: 29103756.
970 33. Warren CJ, Xu T, Guo K, Griffin LM, Westrich JA, Lee D, et al. APOBEC3A functions as a
971 restriction factor of human papillomavirus. J Virol. 2015;89(1):688-702. doi:
972 10.1128/JVI.02383-14. PubMed PMID: 25355878; PubMed Central PMCID:
973 PMC4301161.
974 34. Taylor BJ, Nik-Zainal S, Wu YL, Stebbings LA, Raine K, Campbell PJ, et al. DNA
975 deaminases induce break-associated mutation showers with implication of
976 APOBEC3B and 3A in breast cancer kataegis. Elife. 2013;2:e00534. Epub
977 2013/04/20. doi: 10.7554/eLife.00534 [pii]. PubMed PMID: 23599896; PubMed
978 Central PMCID: PMC3628087.
979 35. Verhalen B, Starrett GJ, Harris RS, Jiang M. Functional upregulation of the DNA cytosine
980 deaminase APOBEC3B by polyomaviruses. J Virol. 2016;90(14):6379-86. doi:
981 10.1128/JVI.00771-16. PubMed PMID: 27147740.
982 36. Vieira VC, Leonard B, White EA, Starrett GJ, Temiz NA, Lorenz LD, et al. Human
983 papillomavirus E6 triggers upregulation of the antiviral and cancer genomic DNA
984 deaminase APOBEC3B. MBio. 2014;5(6). doi: 10.1128/mBio.02234-14. PubMed
985 PMID: 25538195; PubMed Central PMCID: PMC4278539.
986 37. Starrett GJ, Serebrenik AA, Roelofs PA, McCann JL, Verhalen B, Jarvis MC, et al.
987 Polyomavirus T antigen induces APOBEC3B expression using an LXCXE-dependent
988 and TP53-independent mechanism. MBio. 2019;10(1). doi: 10.1128/mBio.02690-18.
989 PubMed PMID: 30723127.
990 38. Walker BA, Wardell CP, Murison A, Boyle EM, Begum DB, Dahir NM, et al. APOBEC family
991 mutational signatures are associated with poor prognosis translocations in multiple
992 myeloma. Nat Commun. 2015;6:6997. doi: 10.1038/ncomms7997. PubMed PMID:
993 25904160; PubMed Central PMCID: PMC4568299.
994 39. Roelofs PA, Goh CY, Chua BH, Jarvis MC, Stewart TA, McCann JL, et al. Characterization

995 of the mechanism by which the RB/E2F pathway controls expression of the cancer
996 genomic DNA deaminase APOBEC3B. Elife. 2020;9. Epub 2020/09/29. doi:
997 10.7554/eLife.61287. PubMed PMID: 32985974; PubMed Central PMCID:
998 PMCPMC7553775.
999 40. Salamango DJ, McCann JL, Demir O, Brown WL, Amaro RE, Harris RS. APOBEC3B
1000 nuclear localization requires two distinct N-terminal domain surfaces. J Mol Biol.
1001 2018;430(17):2695-708. doi: 10.1016/j.jmb.2018.04.044. PubMed PMID: 29787764.
1002 41. de Bruin EC, McGranahan N, Mitter R, Salm M, Wedge DC, Yates L, et al. Spatial and
1003 temporal diversity in genomic instability processes defines lung cancer evolution.
1004 Science. 2014;346(6206):251-6. doi: 10.1126/science.1253462. PubMed PMID:
1005 25301630.
1006 42. Argyris PP, Wilkinson PE, Jarvis MC, Magliocca KR, Patel MR, Vogel RI, et al. Endogenous
1007 APOBEC3B overexpression characterizes HPV-positive and HPV-negative oral
1008 epithelial dysplasias and head and neck cancers. Modern Pathology 2020. Epub
1009 2020/07/08. doi: 10.1038/s41379-020-0617-x. PubMed PMID: 32632179.
1010 43. Yamazaki H, Shirakawa K, Matsumoto T, Hirabayashi S, Murakawa Y, Kobayashi M, et al.
1011 Endogenous APOBEC3B overexpression constitutively generates DNA substitutions
1012 and deletions in myeloma cells. Sci Rep. 2019;9(1):7122. Epub 2019/05/11. doi:
1013 10.1038/s41598-019-43575-y. PubMed PMID: 31073151; PubMed Central PMCID:
1014 PMCPMC6509214.
1015 44. Yamazaki H, Shirakawa K, Matsumoto T, Kazuma Y, Matsui H, Horisawa Y, et al.
1016 APOBEC3B reporter myeloma cell lines identify DNA damage response pathways
1017 leading to APOBEC3B expression. PLoS One. 2020;15(1):e0223463. Epub
1018 2020/01/09. doi: 10.1371/journal.pone.0223463. PubMed PMID: 31914134; PubMed
1019 Central PMCID: PMCPMC6948746.
1020 45. Nik-Zainal S, Wedge DC, Alexandrov LB, Petljak M, Butler AP, Bolli N, et al. Association of

1021 a germline copy number polymorphism of APOBEC3A and APOBEC3B with burden
1022 of putative APOBEC-dependent mutations in breast cancer. Nat Genet.
1023 2014;46(5):487-91. doi: 10.1038/ng.2955. PubMed PMID: 24728294; PubMed Central
1024 PMCID: PMC4137149.
1025 46. Starrett GJ, Luengas EM, McCann JL, Ebrahimi D, Temiz NA, Love RP, et al. The DNA
1026 cytosine deaminase APOBEC3H haplotype I likely contributes to breast and lung
1027 cancer mutagenesis. Nat Commun. 2016;7:12918. doi: 10.1038/ncomms12918.
1028 PubMed PMID: 27650891.
1029 47. Chen TW, Lee CC, Liu H, Wu CS, Pickering CR, Huang PJ, et al. APOBEC3A is an oral
1030 cancer prognostic biomarker in Taiwanese carriers of an APOBEC deletion
1031 polymorphism. Nat Commun. 2017;8(1):465. doi: 10.1038/s41467-017-00493-9.
1032 PubMed PMID: 28878238; PubMed Central PMCID: PMC5587710.
1033 48. Hix MA, Wong L, Flath B, Chelico L, Cisneros GA. Single-nucleotide polymorphism of the
1034 DNA cytosine deaminase APOBEC3H haplotype I leads to enzyme destabilization and
1035 correlates with lung cancer. NAR Cancer. 2020;2(3):zcaa023. Epub 2020/09/29. doi:
1036 10.1093/narcan/zcaa023. PubMed PMID: 32984821; PubMed Central PMCID:
1037 PMCPMC7503452.
1038 49. Feng Y, Love RP, Ara A, Baig TT, Adolph MB, Chelico L. Natural polymorphisms and
1039 oligomerization of human APOBEC3H contribute to single-stranded DNA scanning
1040 ability. J Biol Chem. 2015;290(45):27188-203. doi: 10.1074/jbc.M115.666065.
1041 PubMed PMID: 26396192; PubMed Central PMCID: PMC4646388.
1042 50. Chan K, Roberts SA, Klimczak LJ, Sterling JF, Saini N, Malc EP, et al. An APOBEC3A
1043 hypermutation signature is distinguishable from the signature of background
1044 mutagenesis by APOBEC3B in human cancers. Nat Genet. 2015;47(9):1067-72. doi:
1045 10.1038/ng.3378. PubMed PMID: 26258849.
1046 51. Hoopes JI, Cortez LM, Mertz TM, Malc EP, Mieczkowski PA, Roberts SA. APOBEC3A and

1047 APOBEC3B preferentially deaminate the lagging strand template during DNA
1048 replication. Cell Rep. 2016;14(6):1273-82. doi: 10.1016/j.celrep.2016.01.021. PubMed
1049 PMID: 26832400; PubMed Central PMCID: PMC4758883.
1050 52. Buisson R, Langenbucher A, Bowen D, Kwan EE, Benes CH, Zou L, et al. Passenger
1051 hotspot mutations in cancer driven by APOBEC3A and mesoscale genomic features.
1052 Science. 2019;364(6447). Epub 2019/06/30. doi: 10.1126/science.aaw2872. PubMed
1053 PMID: 31249028; PubMed Central PMCID: PMCPMC6731024.
1054 53. Petersen-Mahrt SK, Harris RS, Neuberger MS. AID mutates *E. coli* suggesting a DNA
1055 deamination mechanism for antibody diversification. Nature. 2002;418(6893):99-103.
1056 PubMed PMID: 12097915.
1057 54. Harris RS, Petersen-Mahrt SK, Neuberger MS. RNA editing enzyme APOBEC1 and some
1058 of its homologs can act as DNA mutators. Mol Cell. 2002;10(5):1247-53. PubMed
1059 PMID: 12453430.
1060 55. Mayorov VI, Rogozin IB, Adkison LR, Frahm C, Kunkel TA, Pavlov YI. Expression of human
1061 AID in yeast induces mutations in context similar to the context of somatic
1062 hypermutation at G-C pairs in immunoglobulin genes. BMC Immunol. 2005;6(1):10.
1063 PubMed PMID: 15949042.
1064 56. Schumacher AJ, Nissley DV, Harris RS. APOBEC3G hypermutates genomic DNA and
1065 inhibits Ty1 retrotransposition in yeast. Proc Natl Acad Sci U S A. 2005;102(28):9854-
1066 9. PubMed PMID: 16000409.
1067 57. Chan K, Sterling JF, Roberts SA, Bhagwat AS, Resnick MA, Gordenin DA. Base damage
1068 within single-strand DNA underlies *in vivo* hypermutability induced by a ubiquitous
1069 environmental agent. PLoS Genet. 2012;8(12):e1003149. doi:
1070 10.1371/journal.pgen.1003149. PubMed PMID: 23271983; PubMed Central PMCID:
1071 PMC3521656.
1072 58. Refsland EW, Hultquist JF, Luengas EM, Ikeda T, Shaban NM, Law EK, et al. Natural

1073 polymorphisms in human APOBEC3H and HIV-1 Vif combine in primary T
1074 lymphocytes to affect viral G-to-A mutation levels and infectivity. PLoS Genet.
1075 2014;10(11):e1004761. doi: 10.1371/journal.pgen.1004761. PubMed PMID:
1076 25411794; PubMed Central PMCID: PMC4238949.
1077 59. Ebrahimi D, Richards CM, Carpenter MA, Wang J, Ikeda T, Becker JT, et al. Genetic and
1078 mechanistic basis for APOBEC3H alternative splicing, retrovirus restriction, and
1079 counteraction by HIV-1 protease. Nat Commun. 2018;9(1):4137. doi: 10.1038/s41467-
1080 018-06594-3. PubMed PMID: 30297863; PubMed Central PMCID: PMC6175962.
1081 60. Chesarino NM, Emerman M. Polymorphisms in human APOBEC3H differentially regulate
1082 ubiquitination and antiviral activity. Viruses. 2020;12(4). Epub 2020/04/03. doi:
1083 10.3390/v12040378. PubMed PMID: 32235597; PubMed Central PMCID:
1084 PMCPMC7232234.
1085 61. Periyasamy M, Patel H, Lai CF, Nguyen VT, Nevedomskaya E, Harrod A, et al. APOBEC3B-
1086 mediated cytidine deamination is required for estrogen receptor action in breast
1087 cancer. Cell Rep. 2015;13(1):108-21. doi: 10.1016/j.celrep.2015.08.066. PubMed
1088 PMID: 26411678; PubMed Central PMCID: PMC4597099.
1089 62. Green AM, Budagyan K, Hayer KE, Reed MA, Savani MR, Wertheim GB, et al. Cytosine
1090 deaminase APOBEC3A sensitizes leukemia cells to inhibition of the DNA replication
1091 checkpoint. Cancer Res. 2017;77(17):4579-88. doi: 10.1158/0008-5472.CAN-16-
1092 3394. PubMed PMID: 28655787; PubMed Central PMCID: PMC5581702.
1093 63. Landry S, Narvaiza I, Linfesty DC, Weitzman MD. APOBEC3A can activate the DNA
1094 damage response and cause cell-cycle arrest. EMBO Rep. 2011;12(5):444-50. Epub
1095 2011/04/05. doi: embor201146 [pii] 10.1038/embor.2011.46. PubMed PMID:
1096 21460793.
1097 64. Land AM, Law EK, Carpenter MA, Lackey L, Brown WL, Harris RS. Endogenous
1098 APOBEC3A DNA cytosine deaminase is cytoplasmic and nongenotoxic. J Biol Chem.

1099 2013;288(24):17253-60. Epub 2013/05/04. doi: 10.1074/jbc.M113.458661 [pii].
1100 PubMed PMID: 23640892; PubMed Central PMCID: PMC3682529.
1101 65. Jarvis MC, Ebrahimi D, Temiz NA, Harris RS. Mutation signatures including APOBEC in
1102 cancer cell lines. JNCI Cancer Spectr. 2018;2(1). doi: 10.1093/jncics/pky002. PubMed
1103 PMID: 29888758; PubMed Central PMCID: PMC5993214.
1104 66. Alexandrov LB, Nik-Zainal S, Wedge DC, Campbell PJ, Stratton MR. Deciphering signatures
1105 of mutational processes operative in human cancer. Cell Rep. 2013;3(1):246-59. Epub
1106 2013/01/16. doi: 10.1016/j.celrep.2012.12.008. PubMed PMID: 23318258; PubMed
1107 Central PMCID: PMCPMC3588146.
1108 67. Law EK, Levin-Klein R, Jarvis MC, Kim H, Argyris PP, Carpenter MA, et al. APOBEC3A
1109 catalyzes mutation and drives carcinogenesis *in vivo*. J Exp Med. 2020;217(12). Epub
1110 2020/09/02. doi: 10.1084/jem.20200261. PubMed PMID: 32870257.
1111 68. Durfee C, Temiz NA, Levin-Klein R, Argyris PP, Alsøe L, Carracedo S, et al. Human
1112 APOBEC3B promotes tumor heterogeneity in vivo including signature mutations and
1113 metastases. *bioRxiv*. 2023;
1114 https://www.biorxiv.org/content/10.1101/2023.02.24.529970v1
1115 69. Jalili P, Bowen D, Langenbucher A, Park S, Aguirre K, Corcoran RB, et al. Quantification of
1116 ongoing APOBEC3A activity in tumor cells by monitoring RNA editing at hotspots. Nat
1117 Commun. 2020;11(1):2971. Epub 2020/06/14. doi: 10.1038/s41467-020-16802-8.
1118 PubMed PMID: 32532990; PubMed Central PMCID: PMCPMC7293259.
1119 70. Langenbucher A, Bowen D, Sakhtemani R, Bournique E, Wise JF, Zou L, et al. An extended
1120 APOBEC3A mutation signature in cancer. Nat Commun. 2021;12(1):1602. Epub
1121 2021/03/13. doi: 10.1038/s41467-021-21891-0. PubMed PMID: 33707442; PubMed
1122 Central PMCID: PMCPMC7952602.
1123 71. Sharma S, Patnaik SK, Taggart RT, Kannisto ED, Enriquez SM, Gollnick P, et al.
1124 APOBEC3A cytidine deaminase induces RNA editing in monocytes and

1125 macrophages. Nat Commun. 2015;6:6881. doi: 10.1038/ncomms7881. PubMed
1126 PMID: 25898173; PubMed Central PMCID: PMC4411297.
1127 72. Roberts SA, Sterling J, Thompson C, Harris S, Mav D, Shah R, et al. Clustered mutations
1128 in yeast and in human cancers can arise from damaged long single-strand DNA
1129 regions. Mol Cell. 2012;46(4):424-35. doi: 10.1016/j.molcel.2012.03.030. PubMed
1130 PMID: 22607975; PubMed Central PMCID: PMC3361558.
1131 73. DeWeerd RA, Nemeth E, Poti A, Petryk N, Chen CL, Hyrien O, et al. Prospectively defined
1132 patterns of APOBEC3A mutagenesis are prevalent in human cancers. Cell Rep.
1133 2022;38(12):110555. Epub 2022/03/24. doi: 10.1016/j.celrep.2022.110555. PubMed
1134 PMID: 35320711.
1135 74. Henderson S, Chakravarthy A, Su X, Boshoff C, Fenton TR. APOBEC-mediated cytosine
1136 deamination links PIK3CA helical domain mutations to human papillomavirus-driven
1137 tumor development. Cell Rep. 2014;7(6):1833-41. doi: 10.1016/j.celrep.2014.05.012.
1138 PubMed PMID: 24910434.
1139 75. Zou X, Koh GCC, Nanda AS, Degasperi A, Urgo K, Roumeliotis TI, et al. A systematic
1140 CRISPR screen defines mutational mechanisms underpinning signatures caused by
1141 replication errors and endogenous DNA damage. Nat Cancer. 2021;2(6):643-57. Epub
1142 2021/06/25. doi: 10.1038/s43018-021-00200-0. PubMed PMID: 34164627; PubMed
1143 Central PMCID: PMCPMC7611045.
1144 76. Zou X, Owusu M, Harris R, Jackson SP, Loizou JI, Nik-Zainal S. Validating the concept of
1145 mutational signatures with isogenic cell models. Nat Commun. 2018;9(1):1744. Epub
1146 2018/05/03. doi: 10.1038/s41467-018-04052-8. PubMed PMID: 29717121; PubMed
1147 Central PMCID: PMCPMC5931590.
1148 77. Warren CJ, Westrich JA, Doorslaer KV, Pyeon D. Roles of APOBEC3A and APOBEC3B in
1149 human papillomavirus infection and disease progression. Viruses. 2017;9(8). doi:
1150 10.3390/v9080233. PubMed PMID: 28825669; PubMed Central PMCID:

1151 PMC5580490.
1152 78. Leonard B, McCann JL, Starrett GJ, Kosyakovsky L, Luengas EM, Molan AM, et al. The
1153 PKC/NF-kappaB signaling pathway induces APOBEC3B expression in multiple
1154 human cancers. Cancer Res. 2015;75(21):4538-47. doi: 10.1158/0008-5472.CAN-15-
1155 2171-T. PubMed PMID: 26420215; PubMed Central PMCID: PMC4631676.
1156 79. Periyasamy M, Singh AK, Gemma C, Kranjec C, Farzan R, Leach DA, et al. p53 controls
1157 expression of the DNA deaminase APOBEC3B to limit its potential mutagenic activity
1158 in cancer cells. Nucleic Acids Res. 2017;45(19):11056-69. doi: 10.1093/nar/gkx721.
1159 PubMed PMID: 28977491; PubMed Central PMCID: PMC5737468.
1160 80. Cho RJ, Alexandrov LB, den Breems NY, Atanasova VS, Farshchian M, Purdom E, et al.
1161 APOBEC mutation drives early-onset squamous cell carcinomas in recessive
1162 dystrophic epidermolysis bullosa. Sci Transl Med. 2018;10(455). doi:
1163 10.1126/scitranslmed.aas9668. PubMed PMID: 30135250.
1164 81. Oh S, Bournique E, Bowen D, Jalili P, Sanchez A, Ward I, et al. Genotoxic stress and viral
1165 infection induce transient expression of APOBEC3A and pro-inflammatory genes
1166 through two distinct pathways. Nat Commun. 2021;12(1):4917. Epub 2021/08/15. doi:
1167 10.1038/s41467-021-25203-4. PubMed PMID: 34389714; PubMed Central PMCID:
1168 PMCPMC8363607.
1169 82. Maruyama W, Shirakawa K, Matsui H, Matsumoto T, Yamazaki H, Sarca AD, et al. Classical
1170 NF-kappaB pathway is responsible for APOBEC3B expression in cancer cells.
1171 Biochem Biophys Res Commun. 2016;478(3):1466-71. doi:
1172 10.1016/j.bbrc.2016.08.148. PubMed PMID: 27577680.
1173 83. Akre MK, Starrett GJ, Quist JS, Temiz NA, Carpenter MA, Tutt AN, et al. Mutation processes
1174 in 293-based clones overexpressing the DNA cytosine deaminase APOBEC3B. PloS
1175 One. 2016;11(5):e0155391. doi: 10.1371/journal.pone.0155391. PubMed PMID:
1176 27163364; PubMed Central PMCID: PMC4862684.

1177 84. Nikkila J, Kumar R, Campbell J, Brandsma I, Pemberton HN, Wallberg F, et al. Elevated
1178 APOBEC3B expression drives a kataegic-like mutation signature and replication
1179 stress-related therapeutic vulnerabilities in p53-defective cells. British Journal of
1180 Cancer. 2017;117(1):113-23. doi: 10.1038/bjc.2017.133. PubMed PMID: 28535155;
1181 PubMed Central PMCID: PMC5520199.
1182 85. Petljak M, Alexandrov LB, Brammeld JS, Price S, Wedge DC, Grossmann S, et al.
1183 Characterizing mutational signatures in human cancer cell lines reveals episodic
1184 APOBEC mutagenesis. Cell. 2019;176(6):1282-94 e20. Epub 2019/03/09. doi:
1185 10.1016/j.cell.2019.02.012. PubMed PMID: 30849372; PubMed Central PMCID:
1186 PMCPMC6424819.
1187 86. Petljak M, Dananberg A, Chu K, Bergstrom EN, Striepen J, von Morgen P, et al.
1188 Mechanisms of APOBEC3 mutagenesis in human cancer cells. Nature.
1189 2022;607(7920):799-807. Epub 20220720. doi: 10.1038/s41586-022-04972-y.
1190 PubMed PMID: 35859169; PubMed Central PMCID: PMCPMC9329121.
1191 87. Law EK, Sieuwerts AM, LaPara K, Leonard B, Starrett GJ, Molan AM, et al. The DNA
1192 cytosine deaminase APOBEC3B promotes tamoxifen resistance in ER-positive breast
1193 cancer. Sci Adv. 2016;2(10):e1601737. doi: 10.1126/sciadv.1601737. PubMed PMID:
1194 27730215.
1195 88. Chen Z, Wen W, Bao J, Kuhs KL, Cai Q, Long J, et al. Integrative genomic analyses of
1196 APOBEC-mutational signature, expression and germline deletion of APOBEC3 genes,
1197 and immunogenicity in multiple cancer types. BMC Medical Genomics.
1198 2019;12(1):131. Epub 2019/09/20. doi: 10.1186/s12920-019-0579-3. PubMed PMID:
1199 31533728; PubMed Central PMCID: PMCPMC6751822.
1200 89. Glaser AP, Fantini D, Wang Y, Yu Y, Rimar KJ, Podojil JR, et al. APOBEC-mediated
1201 mutagenesis in urothelial carcinoma is associated with improved survival, mutations
1202 in DNA damage response genes, and immune response. Oncotarget. 2018;9(4):4537-

1203 48. Epub 2018/02/13. doi: 10.18632/oncotarget.23344. PubMed PMID: 29435122;
1204 PubMed Central PMCID: PMCPMC5796993.
1205 90. Middlebrooks CD, Banday AR, Matsuda K, Udquim KI, Onabajo OO, Paquin A, et al.
1206 Association of germline variants in the APOBEC3 region with cancer risk and
1207 enrichment with APOBEC-signature mutations in tumors. Nat Genet.
1208 2016;48(11):1330-8. Epub 2016/10/28. doi: 10.1038/ng.3670. PubMed PMID:
1209 27643540; PubMed Central PMCID: PMCPMC6583788.
1210 91. Cannataro VL, Gaffney SG, Sasaki T, Issaeva N, Grewal NKS, Grandis JR, et al. APOBEC-
1211 induced mutations and their cancer effect size in head and neck squamous cell
1212 carcinoma. Oncogene. 2019;38(18):3475-87. Epub 2019/01/17. doi: 10.1038/s41388-
1213 018-0657-6. PubMed PMID: 30647454; PubMed Central PMCID: PMCPMC6499643.
1214 92. Driscoll CB, Schuelke MR, Kottke T, Thompson JM, Wongthida P, Tonne JM, et al.
1215 APOBEC3B-mediated corruption of the tumor cell immunopeptidome induces
1216 heteroclitic neoepitopes for cancer immunotherapy. Nat Commun. 2020;11(1):790.
1217 Epub 2020/02/09. doi: 10.1038/s41467-020-14568-7. PubMed PMID: 32034147;
1218 PubMed Central PMCID: PMCPMC7005822.
1219 93. Tsuboi M, Yamane A, Horiguchi J, Yokobori T, Kawabata-Iwakawa R, Yoshiyama S, et al.
1220 APOBEC3B high expression status is associated with aggressive phenotype in
1221 Japanese breast cancers. Breast cancer. 2016;23(5):780-8. Epub 2015/10/20. doi:
1222 10.1007/s12282-015-0641-8. PubMed PMID: 26476745.
1223 94. Brown WL, Law EK, Argyris PP, Carpenter MA, Levin-Klein R, Ranum AN, et al. A rabbit
1224 monoclonal antibody against the antiviral and cancer genomic DNA mutating enzyme
1225 APOBEC3B. Antibodies (Basel). 2019;8(3). Epub 2019/09/24. doi:
1226 10.3390/antib8030047. PubMed PMID: 31544853.
1227 95. Uphoff CC, Drexler HG. Detecting mycoplasma contamination in cell cultures by polymerase
1228 chain reaction. Methods Mol Biol. 2011;731:93-103. Epub 2011/04/26. doi:

1229 10.1007/978-1-61779-080-5_8. PubMed PMID: 21516400.
1230 96. Stenglein MD, Burns MB, Li M, Lengyel J, Harris RS. APOBEC3 proteins mediate the
1231 clearance of foreign DNA from human cells. Nat Struct Mol Biol. 2010;17(2):222-9.
1232 PubMed PMID: 20062055.
1233 97. Refsland EW, Stenglein MD, Shindo K, Albin JS, Brown WL, Harris RS. Quantitative profiling
1234 of the full *APOBEC3* mRNA repertoire in lymphocytes and tissues: implications for
1235 HIV-1 restriction. Nucleic Acids Res. 2010;38:4274-84. PubMed PMID: 20308164.
1236 98. Wang J, Becker JT, Shi K, Lauer KV, Salamango DJ, Aihara H, et al. The role of RNA in
1237 HIV-1 Vif-mediated degradation of APOBEC3H. J Mol Biol. 2019;431(24):5019-31.
1238 Epub 2019/10/20. doi: 10.1016/j.jmb.2019.09.014. PubMed PMID: 31628948;
1239 PubMed Central PMCID: PMCPMC6948013.
1240 99. Lackey L, Demorest ZL, Land AM, Hultquist JF, Brown WL, Harris RS. APOBEC3B and AID
1241 have similar nuclear import mechanisms. J Mol Biol. 2012;419(5):301-14. Epub
1242 2012/03/27. doi: 10.1016/j.jmb.2012.03.011 S0022-2836(12)00266-5 [pii]. PubMed
1243 PMID: 22446380; PubMed Central PMCID: PMC3368237.
1244 100. Shi K, Carpenter MA, Kurahashi K, Harris RS, Aihara H. Crystal structure of the DNA
1245 deaminase APOBEC3B catalytic domain. J Biol Chem. 2015;290(47):28120-30. doi:
1246 10.1074/jbc.M115.679951. PubMed PMID: 26416889; PubMed Central PMCID:
1247 PMC4653671.
1248 101. Shaban NM, Yan R, Shi K, Moraes SN, Cheng AZ, Carpenter MA, et al. Cryo-EM structure
1249 of the EBV ribonucleotide reductase BORF2 and mechanism of APOBEC3B inhibition.
1250 Science Advances. 2022;8(17):eabm2827. doi: 10.1126/sciadv.abm2827.
1251 102. McCann JL, Cristini A, Law EK, Lee SY, Tellier M, Carpenter MA, et al. R-loop homeostasis
1252 and cancer mutagenesis promoted by the DNA cytosine deaminase APOBEC3B.
1253 BioRxiv. 2021. https://www.biorxiv.org/content/10.1101/2021.08.30.458235v1
1254 103. Ikeda T, Shimizu R, Nasser H, Carpenter MA, Cheng AZ, Brown WL, et al. APOBEC3

1255 degradation is the primary function of HIV-1 Vif for virus replication in the myeloid cell
1256 line THP-1. bioRxiv. 2023. Epub 20230329. doi: 10.1101/2023.03.28.534666.
1257 PubMed PMID: 37034786; PubMed Central PMCID: PMCPMC10081227.
1258 104. McNamara RP, McCann JL, Gudipaty SA, D'Orso I. Transcription factors mediate the
1259 enzymatic disassembly of promoter-bound 7SK snRNP to locally recruit P-TEFb for
1260 transcription elongation. Cell Rep. 2013;5(5):1256-68. Epub 2013/12/10. doi:
1261 10.1016/j.celrep.2013.11.003. PubMed PMID: 24316072; PubMed Central PMCID:
1262 PMCPMC3882317.
1263 105. Lackey L, Law EK, Brown WL, Harris RS. Subcellular localization of the APOBEC3 proteins
1264 during mitosis and implications for genomic DNA deamination. Cell Cycle.
1265 2013;12(5):762-72. Epub 2013/02/08. doi: 10.4161/cc.23713 [pii]. PubMed PMID:
1266 23388464; PubMed Central PMCID: PMC3610724.
1267 106. Gyori BM, Venkatachalam G, Thiagarajan PS, Hsu D, Clement MV. OpenComet: an
1268 automated tool for comet assay image analysis. Redox Biol. 2014;2:457-65. Epub
1269 2014/03/14. doi: 10.1016/j.redox.2013.12.020. PubMed PMID: 24624335; PubMed
1270 Central PMCID: PMCPMC3949099.
1271 107. Chiang C, Layer RM, Faust GG, Lindberg MR, Rose DB, Garrison EP, et al. SpeedSeq:
1272 ultra-fast personal genome analysis and interpretation. Nat Methods.
1273 2015;12(10):966-8. Epub 2015/08/11. doi: 10.1038/nmeth.3505. PubMed PMID:
1274 26258291; PubMed Central PMCID: PMCPMC4589466.
1275 108. Blokzijl F, Janssen R, van Boxtel R, Cuppen E. MutationalPatterns: comprehensive
1276 genome-wide analysis of mutational processes. Genome Medicine. 2018;10(1):33.
1277 Epub 2018/04/27. doi: 10.1186/s13073-018-0539-0. PubMed PMID: 29695279;
1278 PubMed Central PMCID: PMCPMC5922316.
1279 109. Rosenthal R, McGranahan N, Herrero J, Taylor BS, Swanton C. DeconstructSigs:
1280 delineating mutational processes in single tumors distinguishes DNA repair

1281 deficiencies and patterns of carcinoma evolution. Genome Biol. 2016;17:31. Epub
1282 20160222. doi: 10.1186/s13059-016-0893-4. PubMed PMID: 26899170; PubMed
1283 Central PMCID: PMCPMC4762164.
1284 110. Cer RZ, Bruce KH, Donohue DE, Temiz NA, Mudunuri US, Yi M, et al. Searching for non-B
1285 DNA-forming motifs using nBMST (non-B DNA motif search tool). Curr Protoc Hum
1286 Genet. 2012;Chapter 18:Unit 18 7 1-22. doi: 10.1002/0471142905.hg1807s73.
1287 PubMed PMID: 22470144; PubMed Central PMCID: PMCPMC3350812.
1288 111. Cameron DL, Schroder J, Penington JS, Do H, Molania R, Dobrovic A, et al. GRIDSS:
1289 sensitive and specific genomic rearrangement detection using positional de Bruijn
1290 graph assembly. Genome Res. 2017;27(12):2050-60. Epub 2017/11/04. doi:
1291 10.1101/gr.222109.117. PubMed PMID: 29097403; PubMed Central PMCID:
1292 PMCPMC5741059.
1293

1294 **Figure Legends – Main and Figure Supplements**

1295

1296 **Figure 1. A3 activity in the HAP1-TK-M9 mutation reporter system.**

1297 (**A**) Schematic of the construction of the HAP1-TK-M9 system and overall experimental workflow

1298 (see **Figure 1 – Figure Supplement 1** for HAP1-TK-M9 karyotype).

1299 (**B**) *A3* mRNA levels in parental HAP1 cells in comparison to the HAP1-TK-M9 daughter clone by

1300 RT-qPCR (p-values from Welch's t-test).

1301 (**C-E**) Immunoblots of A3A, A3B, and A3H in WCE 24 hrs post-transfection of 293T cells. Primary

1302 antibodies are UMN-13, 5210-87-13, and P1-D8, respectively. Mouse anti-β-actin (ACT) is a

1303 loading control for the A3A and A3B blots, and rabbit anti-β-actin (ACT) for the A3H blot (see

1304 **Figure 1 – Figure Supplement 2** for anti-A3A UMN-13 mAb validation). The lower images show

1305 ssDNA deaminase activity of extracts from the same cell populations (S, substrate; P, product).

1306 (**F**) IF microscopy images of HAP1-TK-M9 cells expressing the indicated A3 enzymes or treated

1307 with 4 μM cisplatin for 24 hrs. A3 staining is red, γ-H2AX staining is green, and nuclei are blue

1308 from Hoechst (scale bar = 20 μm).

1309 (**G-H**) Quantification of pan-nuclear γ-H2AX intensity and discrete γ-H2AX foci, respectively, for

1310 each condition described in panel F (each data point represents an independent cell; n>25 cells

1311 per condition; p-values from one-way ANOVA test; red bar indicates mean for each group;

1312 representative of 2 biologically independent experiments).

1313

1314 **Figure 1 – Figure Supplement 1. HAP1-TK-M9 karyotype analysis.**

1315 Representative G-band and spectral karyotype (SKY) images of HAP1-TK-M9 M-phase

1316 chromosomes showing a near diploid DNA content and previously reported aberrations including

1317 the reciprocal chromosome 9:22 translocation (Philadelphia chromosome) characteristic of CML

1318 tumor cells. The Y-chromosome is missing, as reported for the KBM7 parent line of HAP1.

1319

1320 **Figure 1 – Figure Supplement 2. Validation of UMN-13 as a custom rabbit anti-human**
1321 **APOBEC3A mAb.**
1322 (**A**) Schematic of human A3A, A3B, and A3G indicating the unique N-terminal epitope used here
1323 to generate the A3A-specific mAb UMN-13. The schematic also shows the C-terminal epitope
1324 used previously to generate the versatile 5210-87-13 mAb that recognizes these three enzymes.
1325 (**B**) Comparative immunoblots of whole cell extracts from 293T cells expressing each of the 7
1326 human A3 family members with C-terminal HA tags. The blot was probed first with our custom
1327 rabbit anti-human A3A mAb UMN-13 (left) and, after stripping, a commercial anti-HA mAb as an
1328 expression control (right). The positions of the full-length proteins are indicated by red asterisks.
1329 (**C**) Comparative immunoblots of whole cell extracts from the monocytic cell line THP-1 and a
1330 clonal derivative lacking *A3A*-through-*A3G* (Δ*A*-*G*), each treated with DMSO as a control or
1331 LPS/IFN-α to induce expression of multiple A3s including *A3A* and *A3G*. The UMN-13 mAb blot
1332 on the left shows a single band representing full-length A3A (starting at Met1), which is absent in
1333 the deletion mutant, and the 5210-87-13 mAb blot on the right shows A3G (strong top band), A3B
1334 (weak band just below A3G), and both A3A translation products (strong band for full-length A3A
1335 starting at Met1 and a faster-migrating band for the shorter isoform starting at Met13), which are
1336 all absent in the deletion mutant.
1337 (**D**) IF microscopy images of 293T cells expressing A3A-mCherry, A3B-mCherry, or A3G-
1338 mCherry. Only the A3A construct is detected by the UMN-13 mAb as indicated by green signal in
1339 the same cells and cellular compartments as the A3A-mCherry signal.
1340
1341 **Figure 1 – Figure Supplement 3. Additional IF microscopy images of HAP1-TK-M9 cells**
1342 **expressing A3A, A3B, and catalytic mutant derivatives.**
1343 (**A-B**) Additional IF-images of HAP1-TK-M9 cells mock-transduced and DMF or cisplatin treated,
1344 respectively, and stained as indicated in parallel with cells in panels C-F (scale = 20 μm).
1345 (**C-F**) Additional IF-images of HAP1-TK-M9 cells transduced with the indicated A3 expression

1346 constructs and stained with the 5210-87-13 rabbit anti-human A3A/B mAb, γ-H2AX, and Hoechst
1347 (scale = 20 μm).
1348
1349 **Figure 1 – Figure Supplement 4. A3 localization, γ-H2AX staining, and DNA breakage in**
1350 **HeLa cells.**
1351 (**A**) IF-images of HeLa cells transfected with the indicated A3 expression constructs and stained
1352 for A3A/B (5210-87-13 rabbit anti-human A3A/B mAb) or A3H (Novus ARP10 rabbit anti-human
1353 A3H pAb), γ-H2AX, and Hoechst (scale = 20 μm).
1354 (**B**) Quantification of nuclear γ-H2AX staining intensity in the different A3 expressing conditions in
1355 panel A (n>50 cells per condition; red bars indicate mean expression levels; p-values by Welch's
1356 t-test).
1357 (**C**) Dot plot of nuclear γ-H2AX staining intensity versus A3A staining levels for A3A and A3A-
1358 E72A expressing cells shown in panel B. Correlation coefficients (R-values) indicate that A3A
1359 expression levels fail to correlate with nuclear γ-H2AX staining intensity. The p-value by Welch's
1360 t-test reflects the difference between that A3A and A3A-E72A data sets.
1361 (**D**) Dot plot of nuclear γ-H2AX staining intensity versus A3B staining levels for A3B and A3B-
1362 E255A expressing cells shown in panel B. Correlation coefficients (R-values) indicate that A3B
1363 but not A3B-E255A expression levels associate with nuclear γ-H2AX staining intensity. The p-
1364 value by Welch's t-test reflects the difference between that A3B and A3B-E255A data sets.
1365 (**E**) Representative comets from HeLa cells mock transduced, treated with 2 μM camptothecin
1366 (CPT), or transduced with expression constructs for A3A, A3B, or catalytic mutant derivatives
1367 (scale = 50 μm).
1368 (**F**) Quantification of tail moment for >50 cells per condition indicated in panel E (red bars indicate
1369 mean tail moment; p-values by one-way ANOVA).
1370
1371 **Figure 2. Characterization of *TK* mutations in ganciclovir-resistant clones.**

1372 (**A**) A dot plot of Gan^R colonies generated under the indicated A3 expression or control conditions.
1373 Each data point represents the number of Gan^R mutants in a single clonal culture (mean +/- SD
1374 shown with p-values determined using Welch's t-test).
1375 (**B**) Schematics representing all *TK* mutations observed under the indicated A3 expression
1376 conditions (APOBEC3 signature T(C>T/G)W mutations in red, other SBSs in black, and INDELs
1377 in blue; see **Figure 2 – Figure Supplement 1** for schematics of individual *TK* mutants for these
1378 and vector control conditions). Q8X and R212K mutation hotspots are labeled.
1379 (**C-D**) Pentanucleotide logos depicting the -2 and +2 sequence preferences flanking all
1380 T(C>T/G)W mutations that accumulated during A3A- or A3B-expression.
1381
1382 **Figure 2 – Figure Supplement 1. Mutations in *TK* genes derived from ganciclovir-resistant**
1383 **clones.**
1384 Schematics of individual *TK* mutations in ganciclovir-resistant clones. T[C>G/T]W mutations are
1385 shown in red, other SBSs in black, and INDELs in blue. Composite mutation schematics are
1386 shown below for each condition.
1387
1388 **Figure 3. Single base substitution mutation signatures in genomes of ganciclovir-resistant**
1389 **clones.**
1390 (**A**) Trinucleotide profiles of pooled SBSs across all clones sequenced for each listed experimental
1391 condition (A3A n=6; A3A-E72A n=2; A3B n=5; A3B-E255A n=2). See **Figure 3 – Figure**
1392 **Supplement 1 and 2** for mRNA and protein level expression confirmation, respectively, and
1393 **Figure 3 – Figure Supplement 5** for SBS profiles from each WGS.
1394 (**B**) Pentanucleotide logos depicting -2, +1 and +2 sequence preferences flanking all C-to-T and
1395 C-to-G mutated TC motifs in WGS from HAP1-TK-M9 cells expressing A3A or A3B in comparison
1396 to aggregate controls (catalytic mutants and GFP only conditions, which do not show evidence
1397 for APOBEC3 signature mutations; **Figure 3 – Figure Supplement 5**). The distribution of

1398 nucleobases flanking all TC motifs in the human genome is shown for comparison.

1399

1400 **Figure 3 – Figure Supplement 1. *A3* mRNA expression in the HAP1-TK-M9 system.**

1401 (**A**) *A3A*, *A3B*, and *A3H* mRNA expression levels relative to those of the housekeeping gene *TBP*

1402 for the indicated HAP1-TK-M9 conditions (RNA-seq FKPM from n$\geq$2 Gan^R clones for each

1403 condition; mean +/- SD shown). RNA-seq data from A3 signature-high breast cancer cell lines

1404 (BT-474 and MDA-MB-453), CCLE breast-derived cell lines (n=52), and TGCA primary breast

1405 cancers (n=1093) for comparison (mean +/- SD).

1406 (**B**) A heatmap depicting mean expression levels of all 7 human *APOBEC3* family members, in

1407 addition to *AICDA*, *APOBEC1*, *APOBEC2*, and *APOBEC4*, relative to those of the housekeeping

1408 gene *TBP* (RNA-seq values are FKPM; n$\geq$2 for each condition to provide matching data sets for

1409 the Gan^R clones subjected to WGS). Endogenous *A3C* provides a consistent internal control.

1410

1411 **Figure 3 – Figure Supplement 2. A3 protein expression in Gan^R clones.**

1412 (**A**) Immunoblots of A3A, A3B, and A3H in the indicated Gan^R clones. Tubulin (TUB) is a loading

1413 control.

1414 (**B**) Deaminase activity of WCE on ssDNA from the same clones (S, substrate; P, product).

1415

1416 **Figure 3 – Figure Supplement 3. Representative MLV-A3 insertion sites in HAP1-TK-M9**

1417 **clones.**

1418 A table indicating chromosomal locations of representative MLV-A3 insertion sites in HAP-1-TK-

1419 M9 clones. In each clone, a single MLV-A3 insertion is positioned in the window between the

1420 indicated nucleotides.

1421

1422 **Figure 3 – Figure Supplement 4. SBS mutation summary.**

1423 (**A-C**) Dot plots showing total numbers of SBS mutations and cytosine mutations in NCN and

1424 TCW motifs, respectively, in WGS from individual Gan^R granddaughter clones (p-values using
1425 Welch's t-test).
1426 (**D-E**) Dot plots showing total numbers of cytosine mutations in NCN and TCW motifs,
1427 respectively, in WGS from individual A3A and A3B expressing Gan^R granddaughter clones in
1428 comparison to all non-catalytic controls combined (p-values using Welch's t-test).
1429
1430 **Figure 3 – Figure Supplement 5. Single base substitution mutation profiles of individual**
1431 **ganciclovir-resistant clones by whole genome sequencing.**
1432 Trinucleotide profiles of all SBS mutations in WGSs from the indicated Gan^R clones (conditions,
1433 clone names, and total SBS numbers are indicated to the right in each profile). Aggregate profiles
1434 for A3A, A3B, and corresponding catalytic mutant controls are shown in **Figure 3**.
1435
1436 **Figure 3 – Figure Supplement 6. APOBEC3 mutation signature analysis using an**
1437 **alternative NMF-based workflow.**
1438 (**A**) Mutation signature profiles extracted from Gan^R clone WGS using an NMF-based approach.
1439 The SBSs of each clone resolved into 2 signatures - an APOBEC3-like mutation signature shown
1440 in red and a background signature in blue. APOBEC3 mutation enrichment scores are shown
1441 above each bar, with significantly enriched values shown in red (Benjamini-Hochberg false
1442 discovery rate-corrected q-value < 0.05). The dashed line represents the average level of
1443 APOBEC3 signature mutations observed in the two eGFP control clones (*i.e.*, background signal).
1444 (**B**) Trinucleotide mutation profiles of signatures A and B derived using NMF with the former
1445 exhibiting an APOBEC3 SBS signature (C-to-T and C-to-G in TCA and TCT motifs).
1446
1447 **Figure 3 – Figure Supplement 7. Indel landscape of sequenced granddaughter clones.**
1448 Bar plots showing the percentage of each of the indicated indel types occurring in WGS from
1449 HAP1-TK-M9 granddaughter clones. Total numbers of single T deletions at homopolymers of 6

1450 or more are too numerous to plot on the same axis and are therefore listed here (A3A, 39.5%;
1451 A3B, 44.2%; A3A-E72A, 40.6%; A3B-E255A, 40.0%; A3H, 31.8%; and GFP, 27.4%). The cosine
1452 similarity of the indel landscape across all conditions is over 0.96 indicating no significant
1453 differences.
1454
1455 **Figure 4. Mesoscale properties of A3A and A3B *in vitro* and in the HAP1-TK-M9 system.**
1456 (**A-B**) Deamination kinetics of A3A and A3B using *SDHB* and *NUP93* DNA hairpin substrates in
1457 comparison to corresponding linear controls made by scrambling the 5' or 3' portion of the stem,
1458 respectively. See text for full description and **Figure 4 – Figure Supplement 1** for a genome-
1459 wide analysis and **Figure 4 – Figure Supplement 2** for a gel-based confirmation of DNA
1460 oligonucleotide integrity.
1461 (**C**) Rainfall plots of genome-wide intermutation distances showing APOBEC3 signature kataegic
1462 tracts (red arrows) in representative A3A- and A3B-expressing Gan^R clones (C>T mutations are
1463 red, C>G black, other SBS gray). Kataegis are defined here as ≥2 strand-coordinated APOBEC3
1464 signature SBS mutations within a 10 kbp window.
1465
1466 **Figure 4 – Figure Supplement 1. Hairpin mutation analysis.**
1467 A list of the total number of SBS mutations in each HAP1-TK-M9 clone and numbers and
1468 frequencies mutations predicted to occur in non-hairpin regions or hairpin loop regions of the
1469 genome (see **Methods** for additional information). Columns 1 and 2: The A3 or control condition
1470 and clone number as listed in Figure 2 – Figure Supplement 1. Column 3: The total number of
1471 SBS mutations in each clone by WGS (human genome: 3000 mbp). Columns 4 and 5: The total
1472 number and frequency of non-hairpin APOBEC3 signature TCW mutations per clone (estimated
1473 non-hairpin genomic DNA: 2991.375 mbp). Columns 6 and 7: Total number and frequency of
1474 APOBEC3 signature TCW mutations in 3-11 nucleotide loop regions of predicted chromosomal
1475 DNA hairpin structures (estimated ssDNA loop region genomic DNA: 8.625 mbp). For A3A data

1476 (red boxes) and A3B data (green boxes), the non-hairpin versus hairpin APOBEC3 TCW mutation
1477 frequencies are not significantly different (P = 0.25 and P = 0.19, respectively, by Welch's t-test).
1478 APOBEC3 TCW mutation frequencies are also not significantly different between the A3A and
1479 A3B data sets (red versus green boxes) for both non-hairpin regions (P = 0.087) or predicted
1480 ssDNA loop regions of hairpins (P = 0.149). However, in comparisons of A3A (red) and A3B (blue)
1481 APOBEC3 TCW mutation frequencies in ssDNA loop regions of hairpins and equivalent data sets
1482 from aggregate controls (blue boxes: the catalytic mutant of each protein and GFP), the A3A data
1483 set approaches statistical significance (P = 0.0655) and the A3B data are significantly different (P
1484 = 0.0367) by Welch's t-test.
1485
1486 **Figure 4 – Figure Supplement 2. Validation of *NUP93* and *SDHB* DNA hairpins and linear**
1487 **structures.**
1488 Native PAGE (left) and denaturing PAGE (right) analysis of the indicated oligonucleotide
1489 substrates. The hairpin substrates migrate faster under native conditions, and their mobility is
1490 similar to the linear derivatives under denaturing conditions. The only oligonucleotide not used in
1491 biochemical experiments in Figure 4 is the *NUP93*-noHP (no hairpin), which has half of the stem
1492 replaced by adenines (5'-6-carboxyfluorescein-GCAAGCTGTTCAAAAAAATGA) and is included
1493 here as an additional control.
1494
1495 **Figure 5. A composite origin of APOBEC3 signature mutations in breast cancer.**
1496 An unsupervised clustering analysis of similarity between the pentanucleotide SBS profiles from
1497 WGSs of the A3A, A3B, and control Gan^R clones described here versus those from primary breast
1498 tumor whole-genome sequencing data sets (ICGC, n=784). The APOBEC3 mutation signature is
1499 represented by both enrichment score and SBS2+13 (red), HRD signature as SBS3 (blue), and
1500 ageing signature as SBS1 (gray).
1501

1502 **Supplementary Table S1. Pentanucleotide contexts of SBS mutations in HAP1-TK-M9**
1503 **clones and ICGC breast tumors.**
1504 This supplementary Microsoft Excel file reports the number of SBS mutations occurring in each
1505 sample in each pentanucleotide context. Layer 1 lists SBS mutations in Hap1-TK-M9 clones
1506 described here, and Layer 2 lists SBS mutations in ICGC primary breast cancers. The ICGC data
1507 are organized from left-most column to right-most column from highest APOBEC3 signature
1508 enrichment (lowest Benjamini-Hochberg q-value) to lowest APOBEC3 signature enrichment
1509 (highest Benjamini-Hochberg q-value). Columns reporting data from tumors with significant
1510 APOBEC3 mutation signature enrichments are shaded gray.
1511

Figure 1

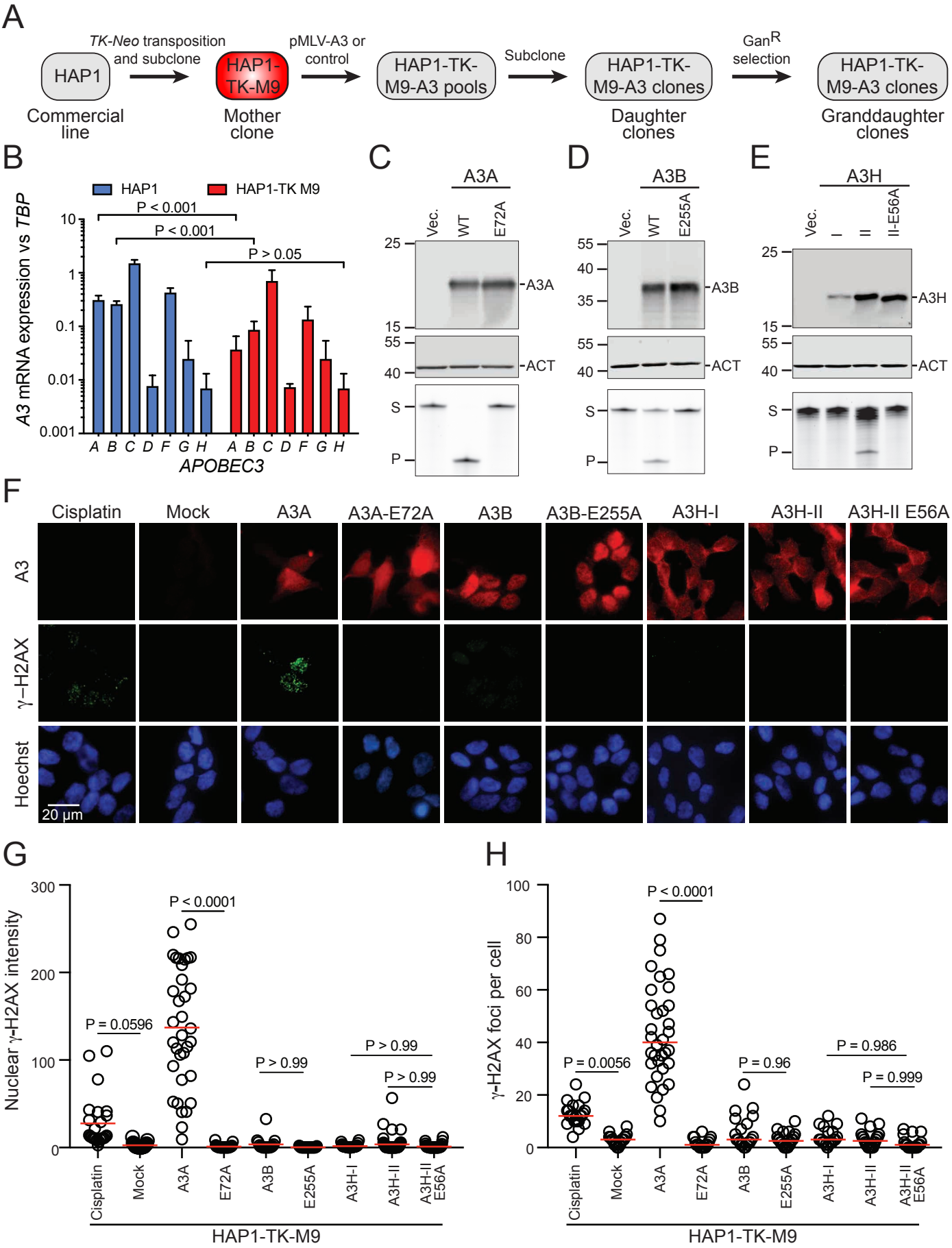

A
TK-Neo transposition and subclone
pMLV-A3 or control
Subclone
GanR selection
HAP1
HAP1-TK-M9
HAP1-TK-M9-A3 pools
HAP1-TK-M9-A3 clones
HAP1-TK-M9-A3 clones
Commercial line
Mother clone
Daughter clones
Granddaughter clones
B
HAP1
HAP1-TK M9
P < 0.001
P < 0.001
P > 0.05
A3 mRNA expression vs TBP
A B C D F G H A B C D F G H
APOBEC3
C
A3A
Vec.
WT
E72A
A3A
ACT
S
P
D
A3B
Vec.
WT
E255A
A3B
ACT
S
P
E
A3H
Vec.
I
II
II-E56A
A3H
ACT
S
P
F
Cisplatin
Mock
A3A
A3A-E72A
A3B
A3B-E255A
A3H-I
A3H-II
A3H-II E56A
A3
γ–H2AX
Hoechst
20 µm
G
Nuclear γ-H2AX intensity
P = 0.0596
P < 0.0001
P > 0.99
P > 0.99
P > 0.99
Cisplatin
Mock
A3A
E72A
A3B
E255A
A3H-I
A3H-II
A3H-II E56A
HAP1-TK-M9
H
γ-H2AX foci per cell
P = 0.0056
P < 0.0001
P = 0.96
P = 0.986
P = 0.999
Cisplatin
Mock
A3A
E72A
A3B
E255A
A3H-I
A3H-II
A3H-II E56A
HAP1-TK-M9

Figure 1 - Figure Supplement 1

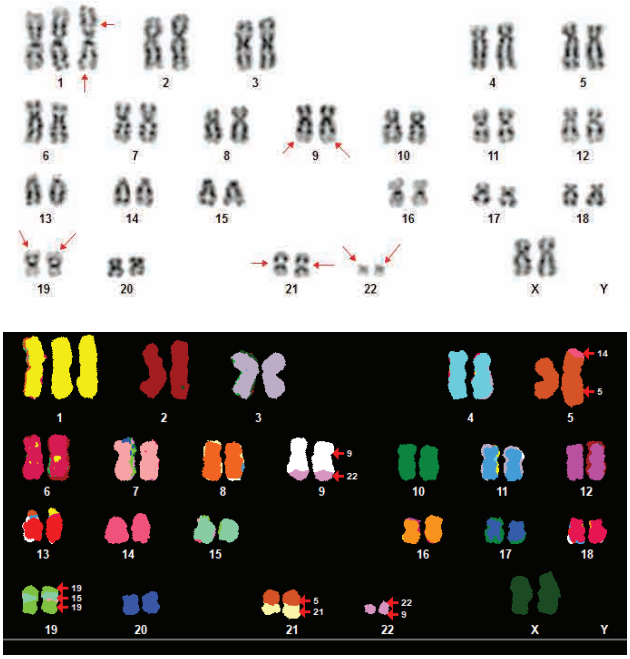


Figure 1 - Figure Supplement 2

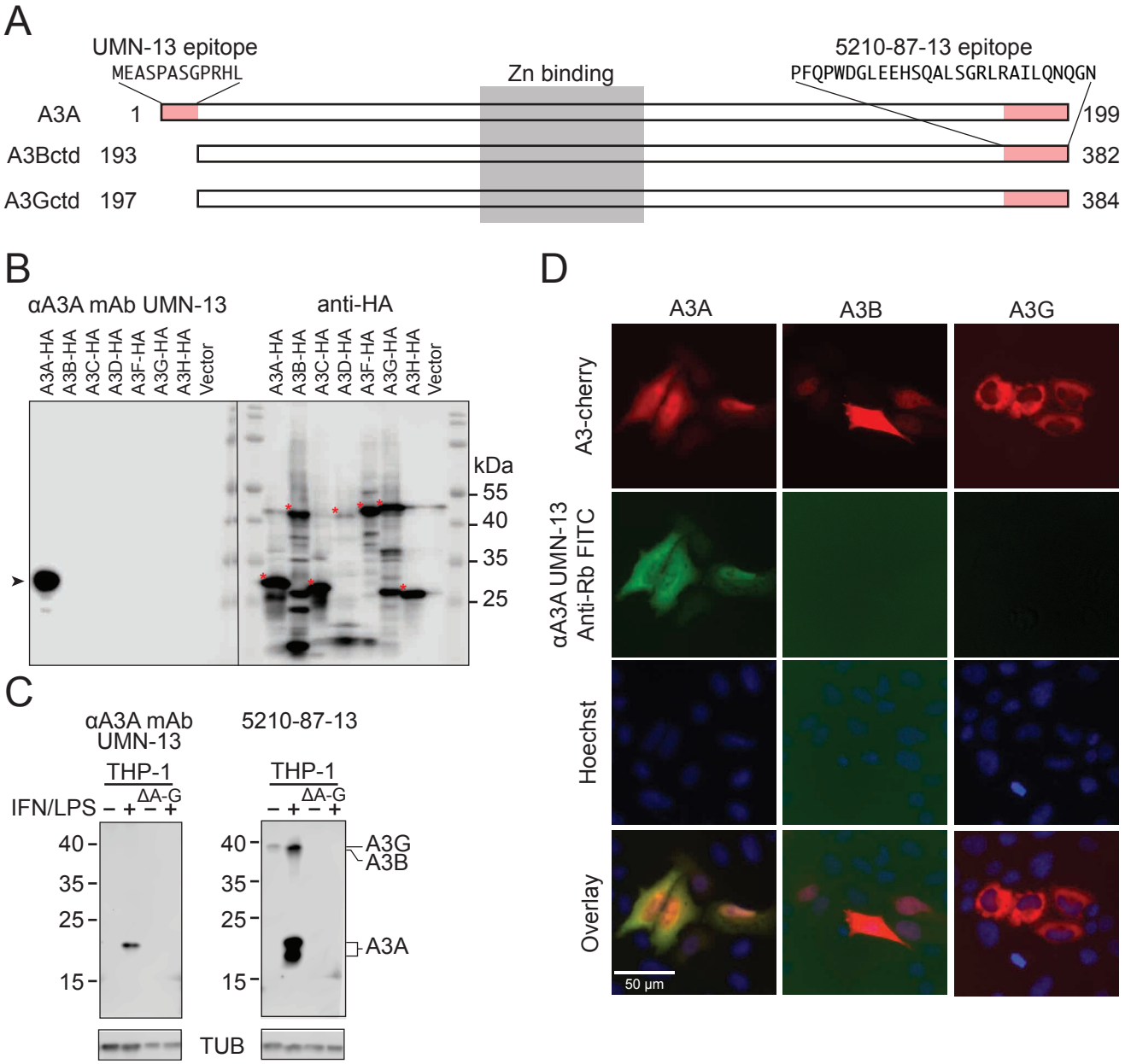


Figure 1 - Figure Supplement 3

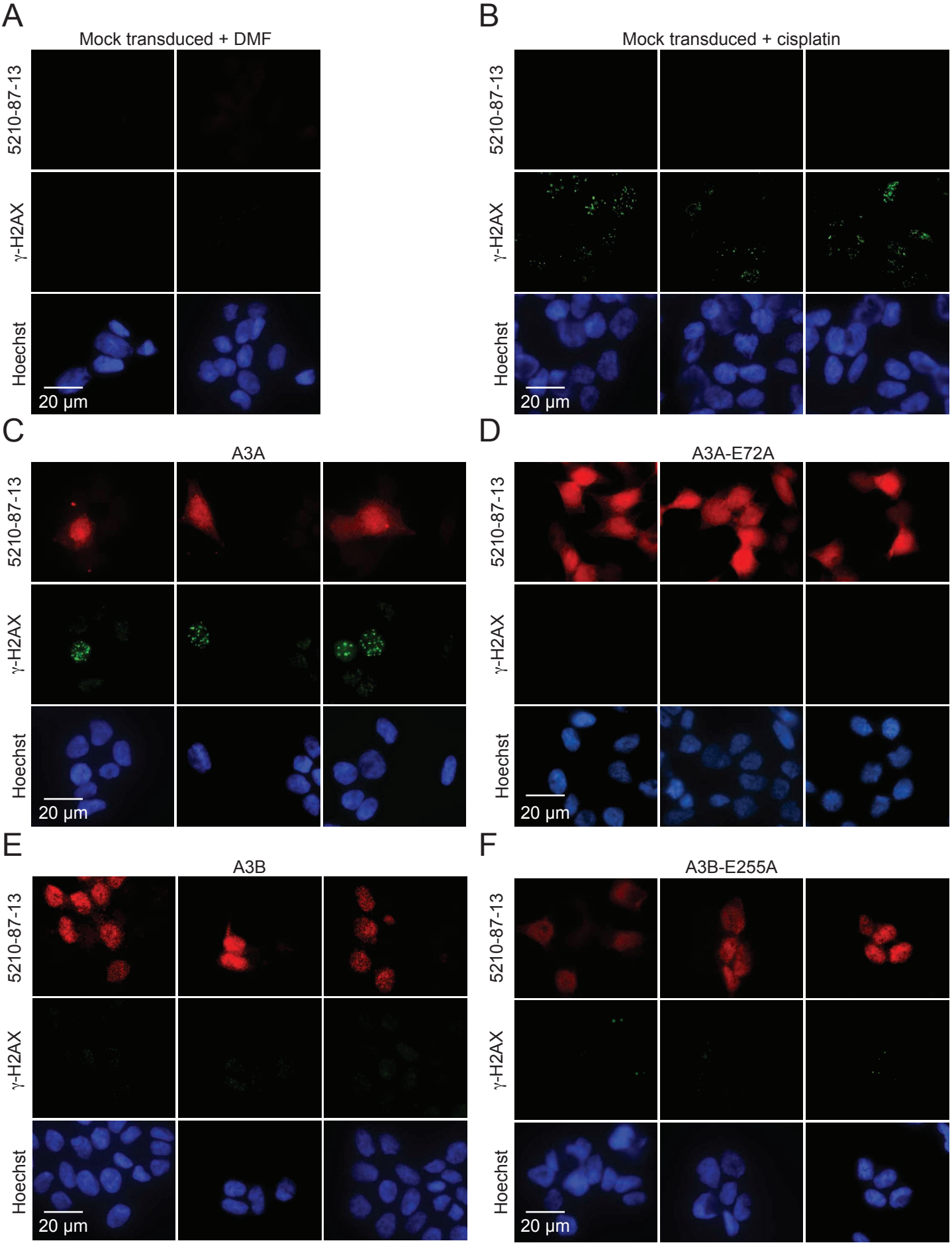


Figure 1 - Figure Supplement 4

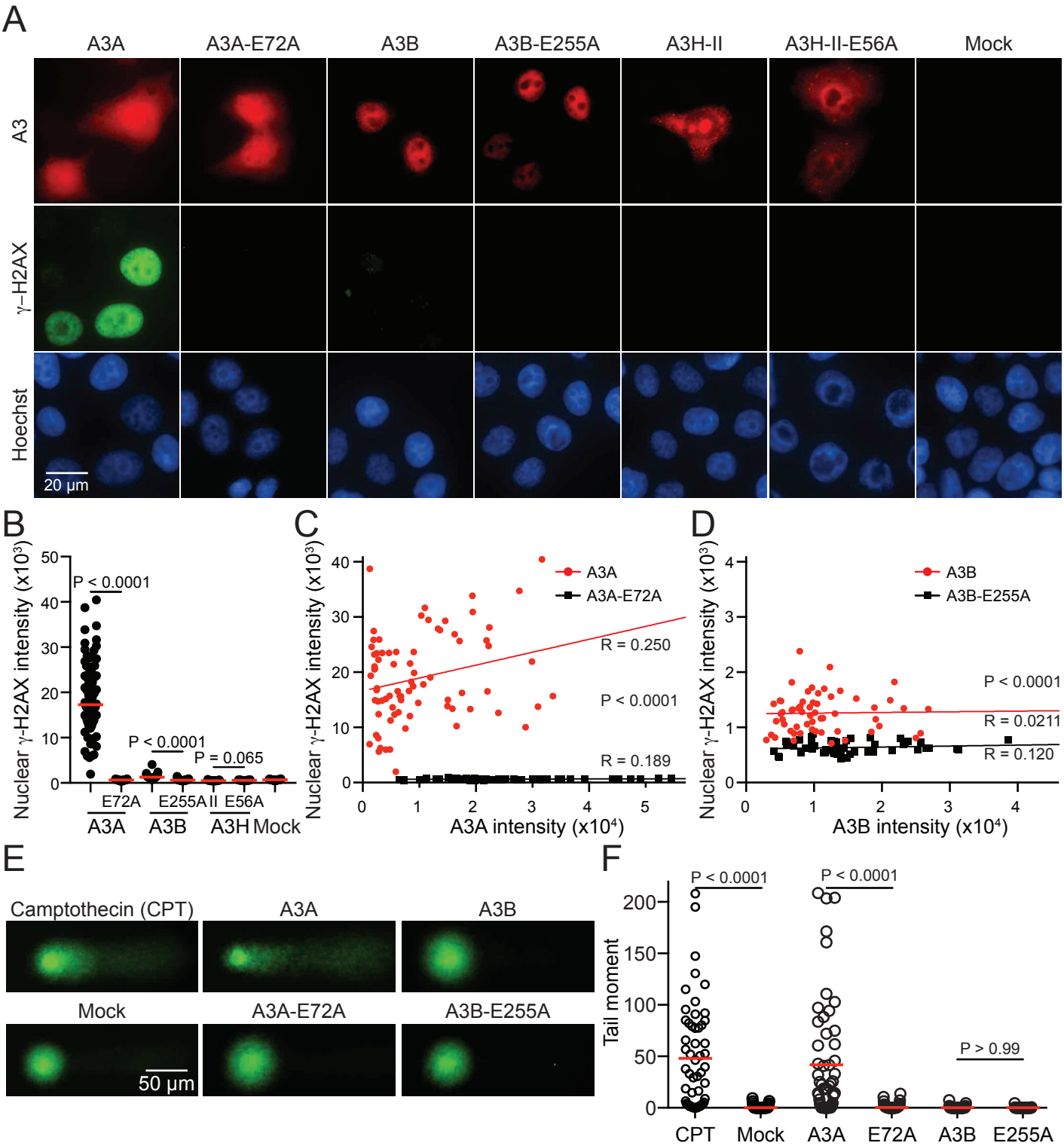


Figure 2

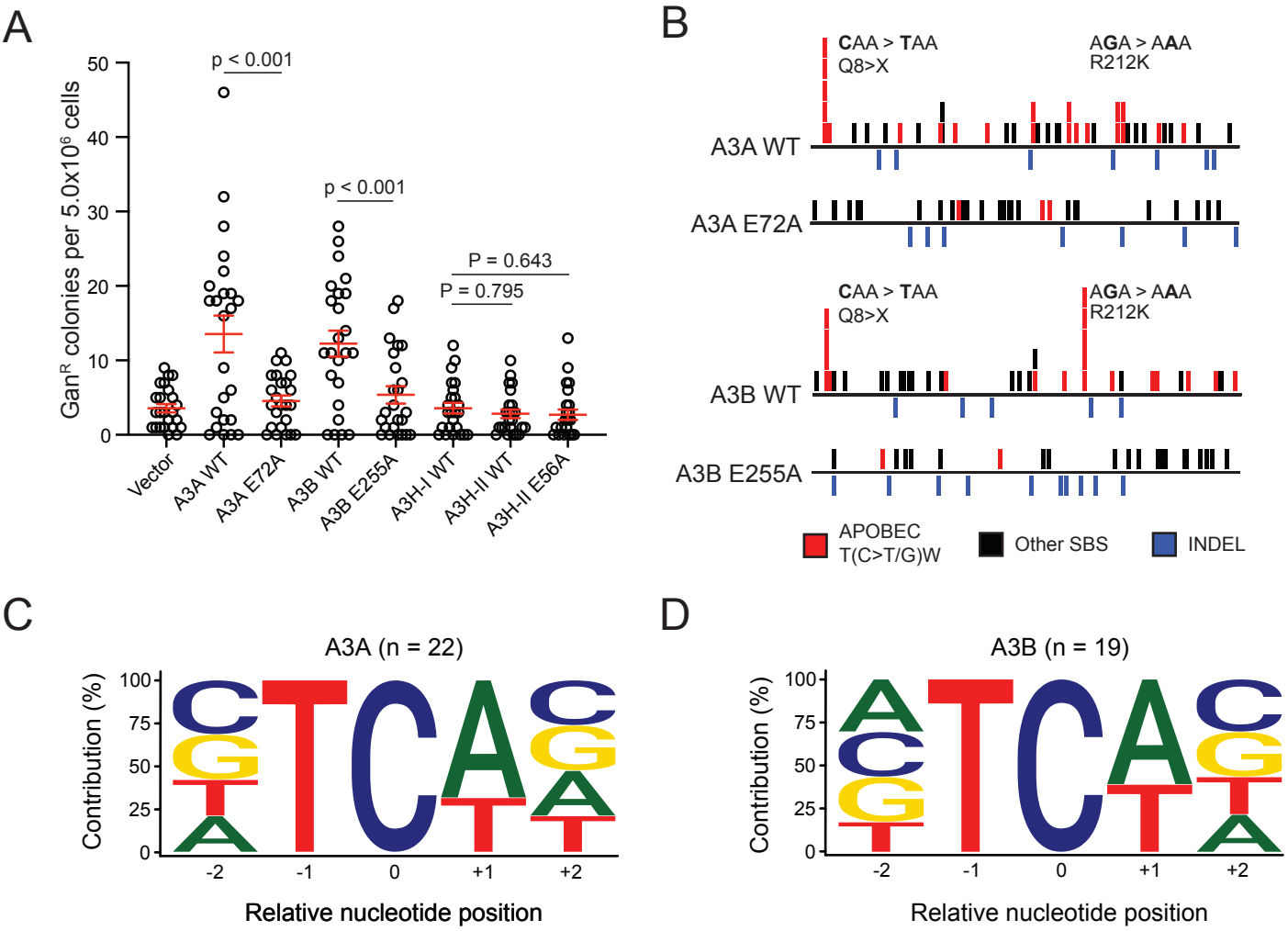

A
Gan^R colonies per 5.0x10^6 cells
p < 0.001
p < 0.001
P = 0.643
P = 0.795
Vector
A3A WT
A3A E72A
A3B WT
A3B E255A
A3H-I WT
A3H-II WT
A3H-II E56A
B
CAA > TAA
Q8>X
AGA > AAA
R212K
A3A WT
A3A E72A
A3B WT
A3B E255A
APOBEC
T(C>T/G)W
Other SBS
INDEL
C
A3A (n = 22)
Contribution (%)
Relative nucleotide position
D
A3B (n = 19)
Contribution (%)
Relative nucleotide position

Figure 2 - Figure Supplement 1

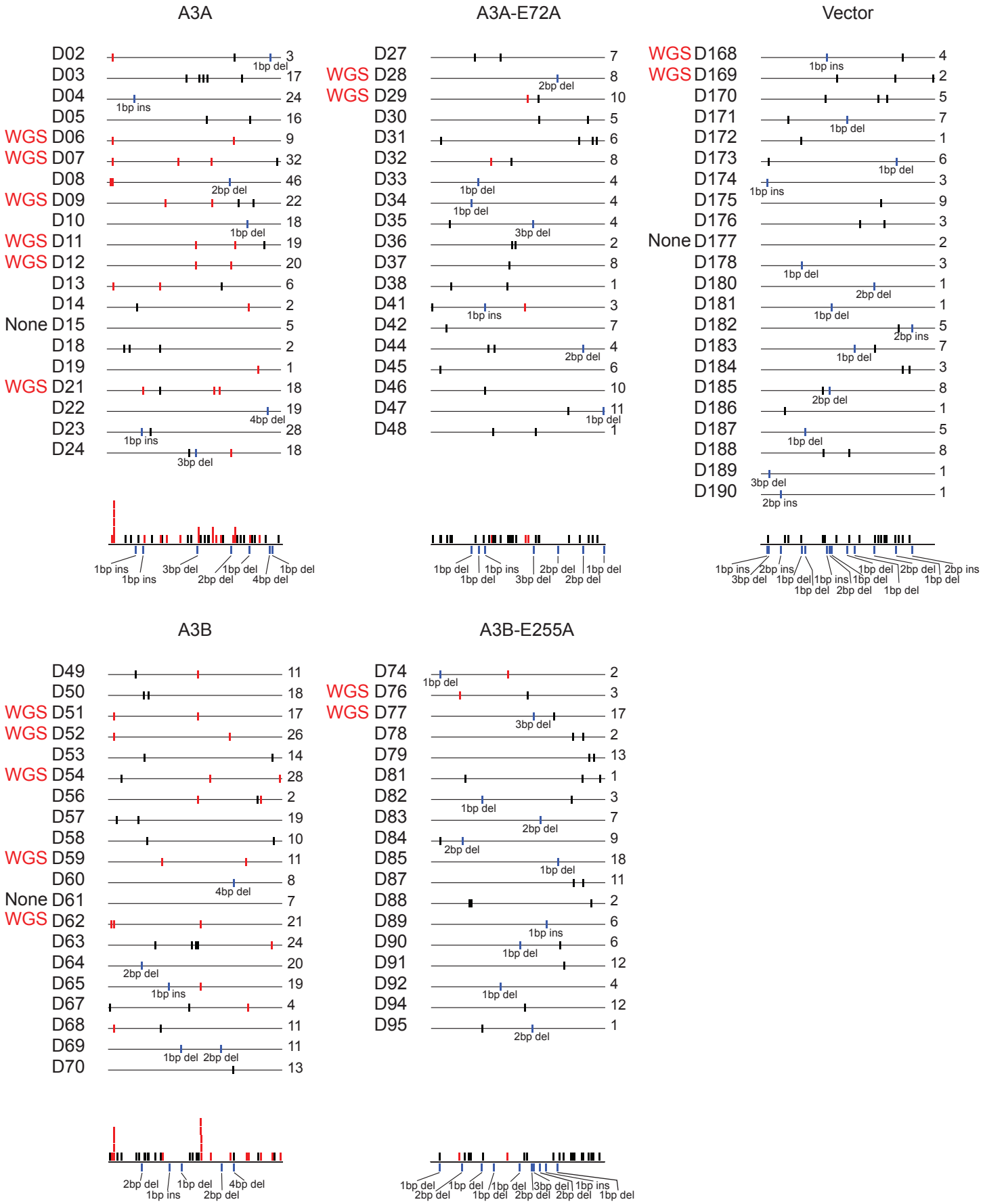


Figure 3

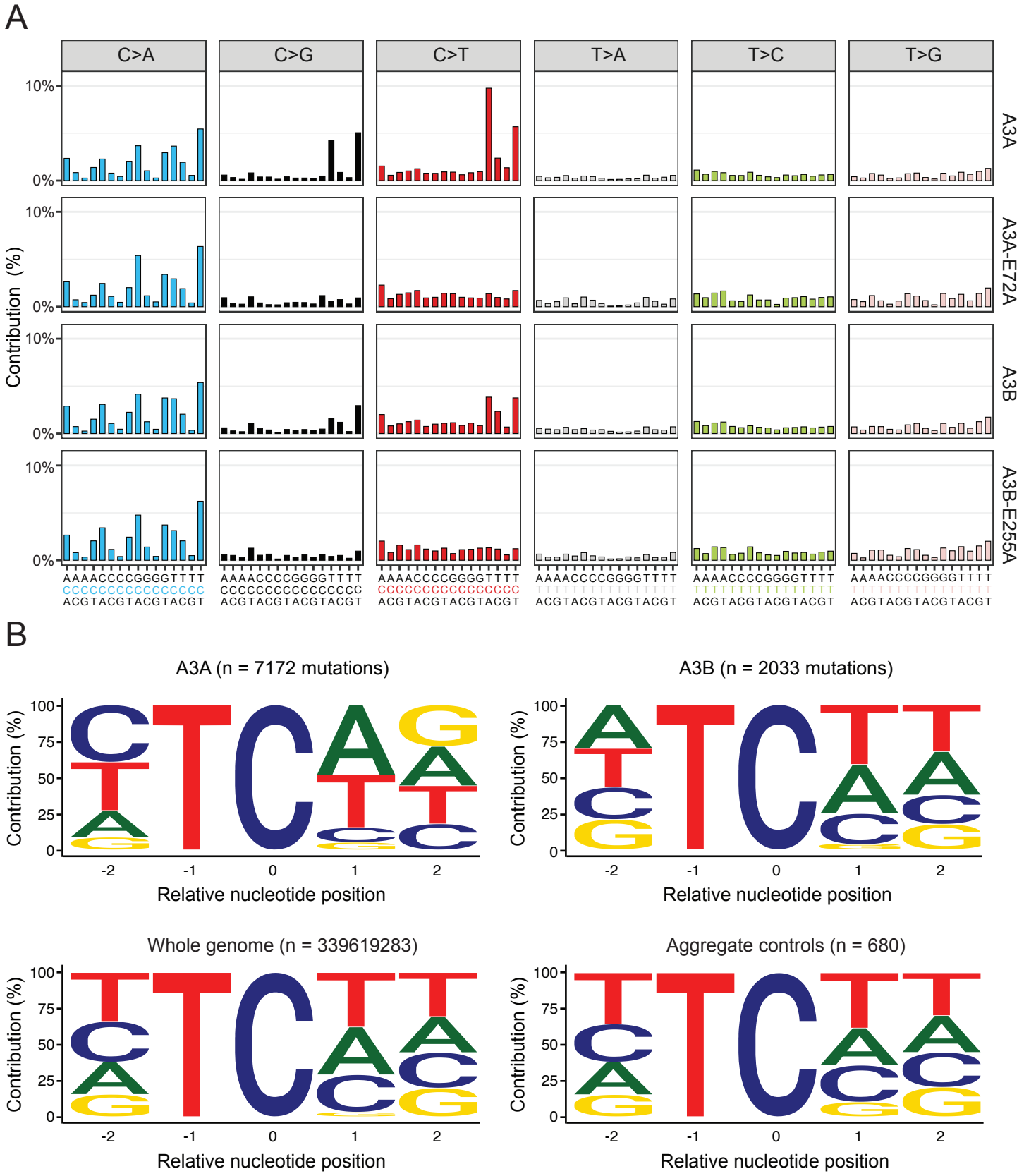
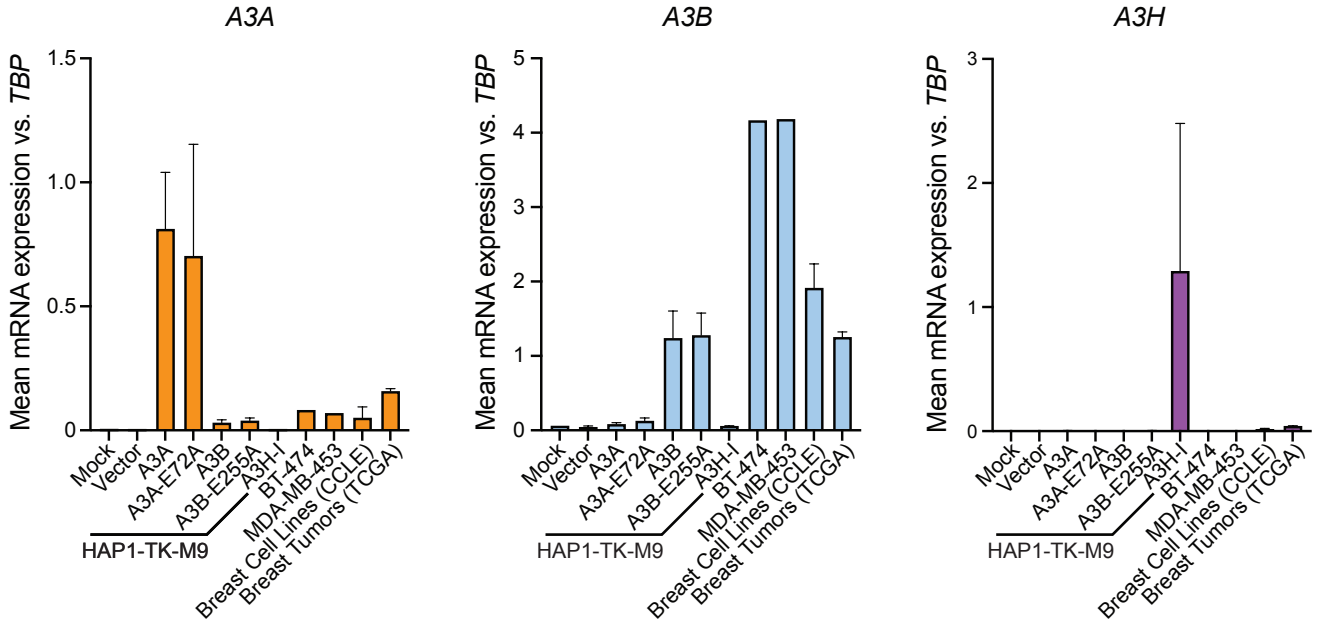

A
C>A
C>G
C>T
T>A
T>C
T>G
A3A
A3A-E72A
A3B
A3B-E255A
Contribution (%)
10%
0%
AAAACCCCGGGGTTTT
CCCCCCCCCCCCCCCC
ACGTACGTACGTACGT
B
A3A (n = 7172 mutations)
A3B (n = 2033 mutations)
Whole genome (n = 339619283)
Aggregate controls (n = 680)
Contribution (%)
100
75
50
25
0
-2
-1
0
1
2
Relative nucleotide position

Figure 3 - Figure Supplement 1

A



B

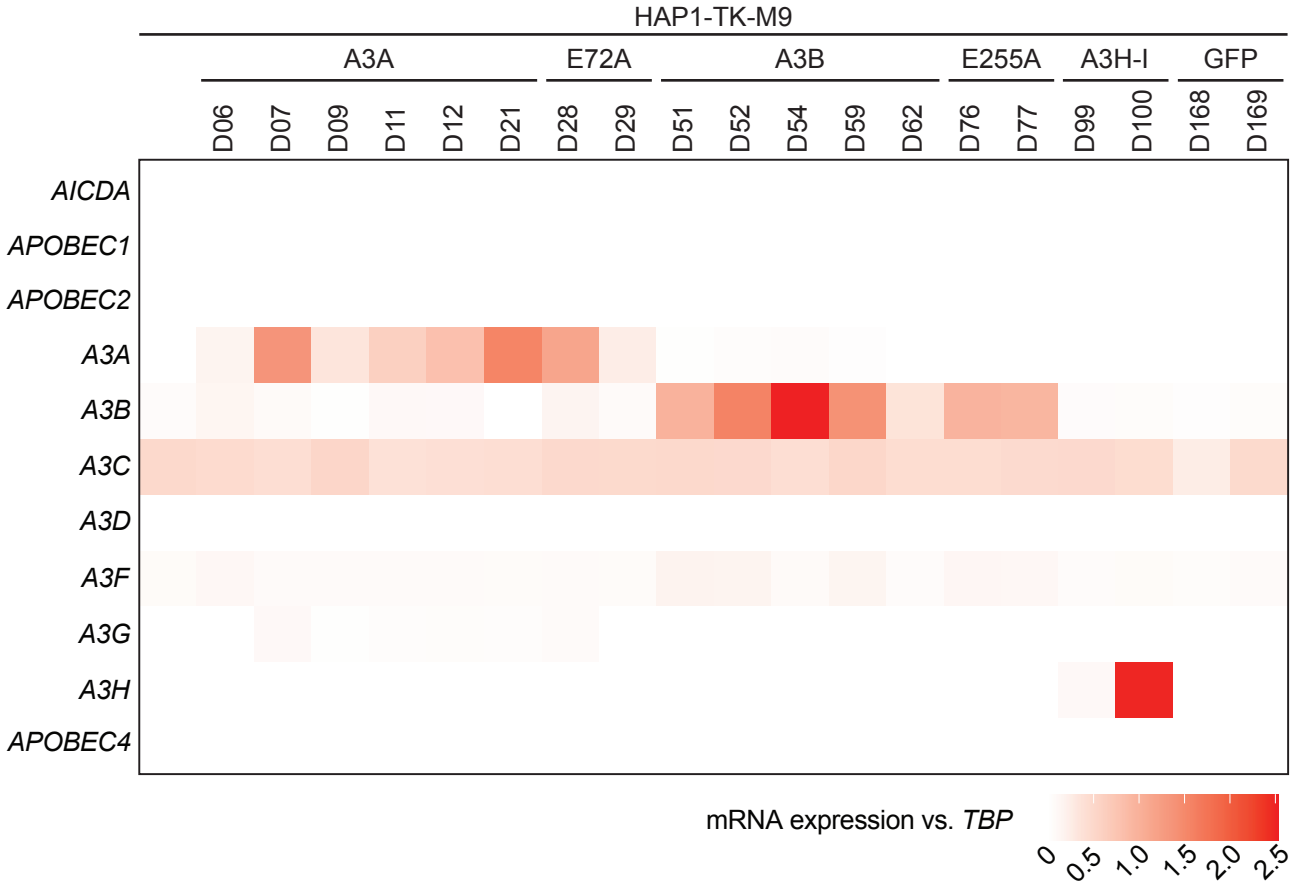


Figure 3 - Figure Supplement 2

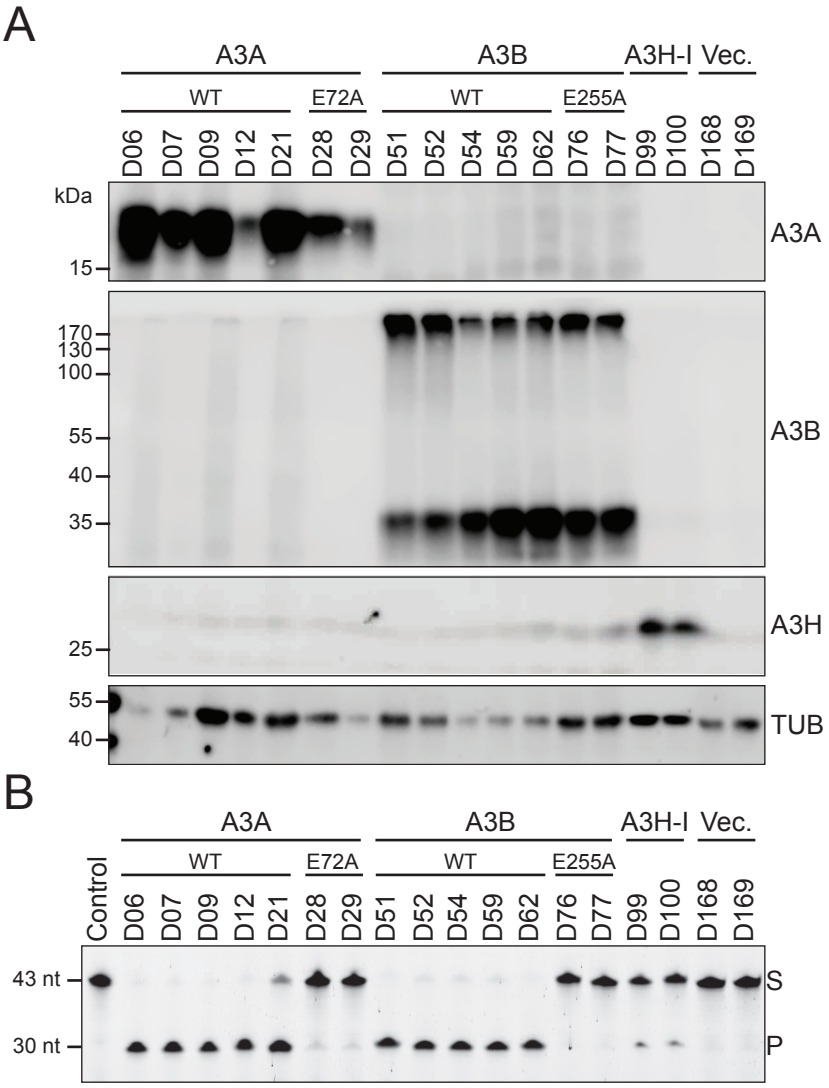


Figure 3 - Figure Supplement 3

Sample	Clone	Chromosome	Integration position	Gene	Location
A3A	D09	18	2573779 – 2574079	*NDC80*	Intron
A3A-E72A	D28	17	76436512 – 76437151	*UBE2O*	Intron
A3B	D51	14	102141202 – 102141502	*WDR20*	Intron
A3B-E255A	D76	20	1117560 – 1117860	*PSMF1*	Intron
A3H-I	D99	1	149776687 – 149776987	None	Intergenic

Figure 3 - Figure Supplement 4

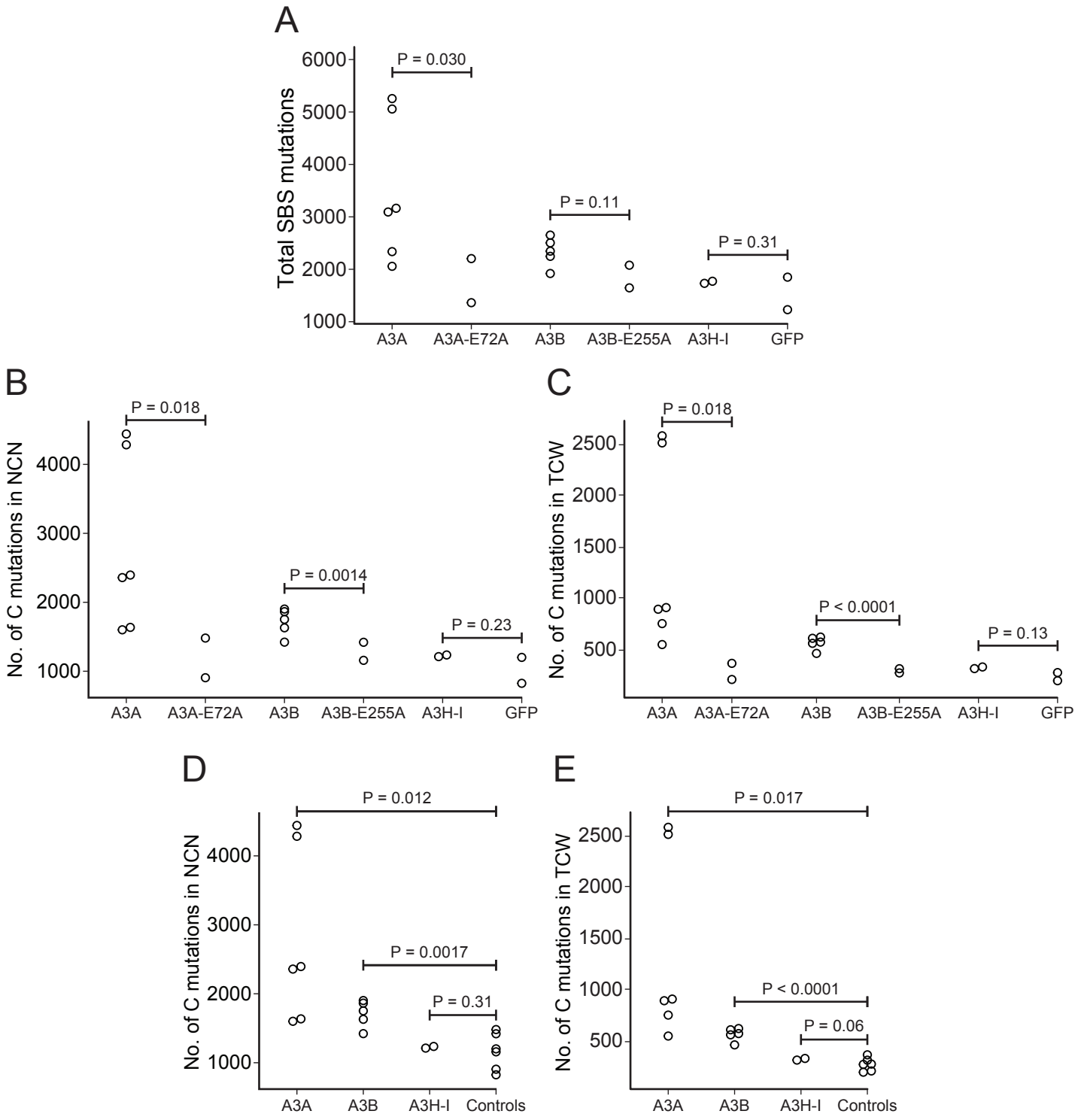


Figure 3 - Figure Supplement 5

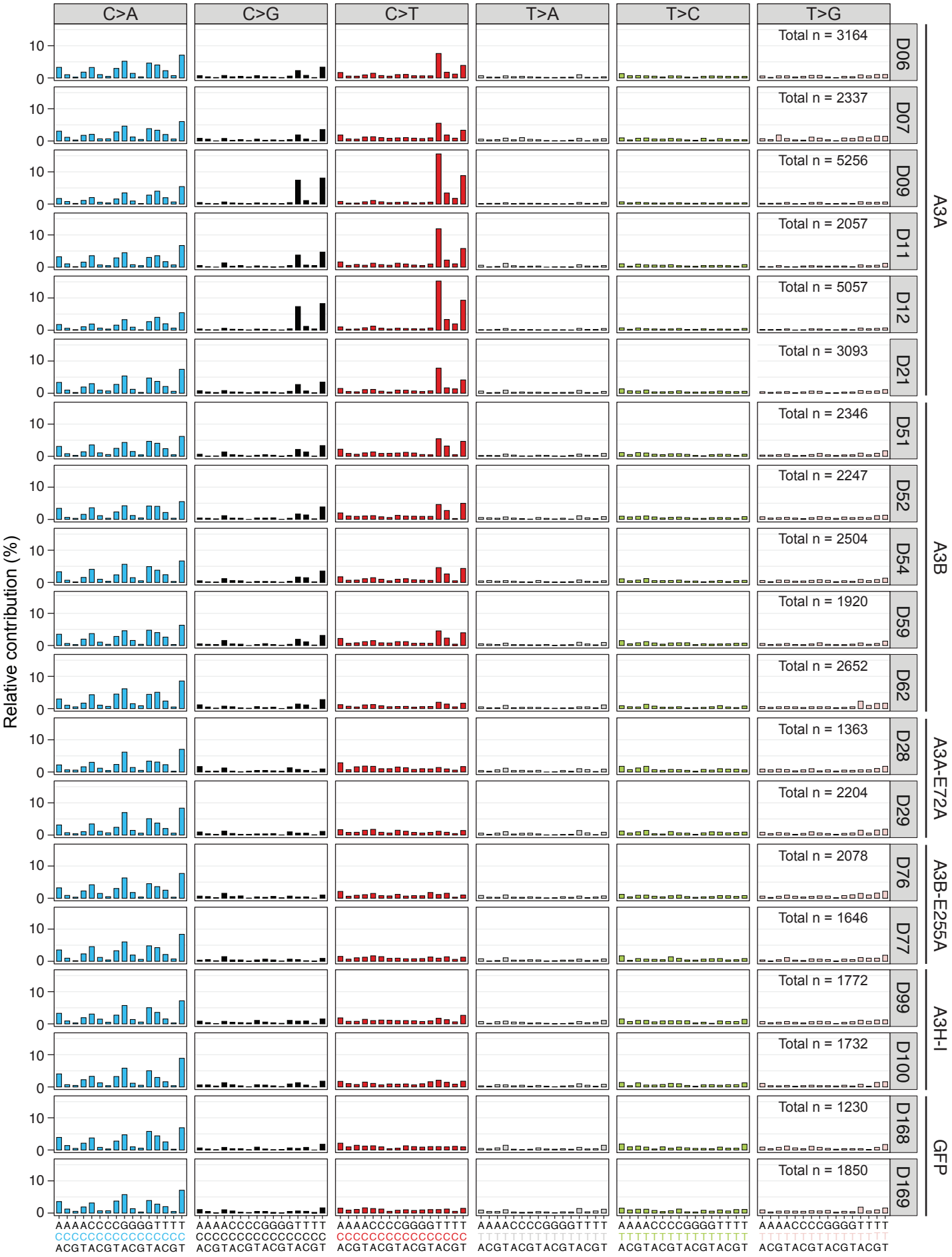


Figure 3 - Figure Supplement 6

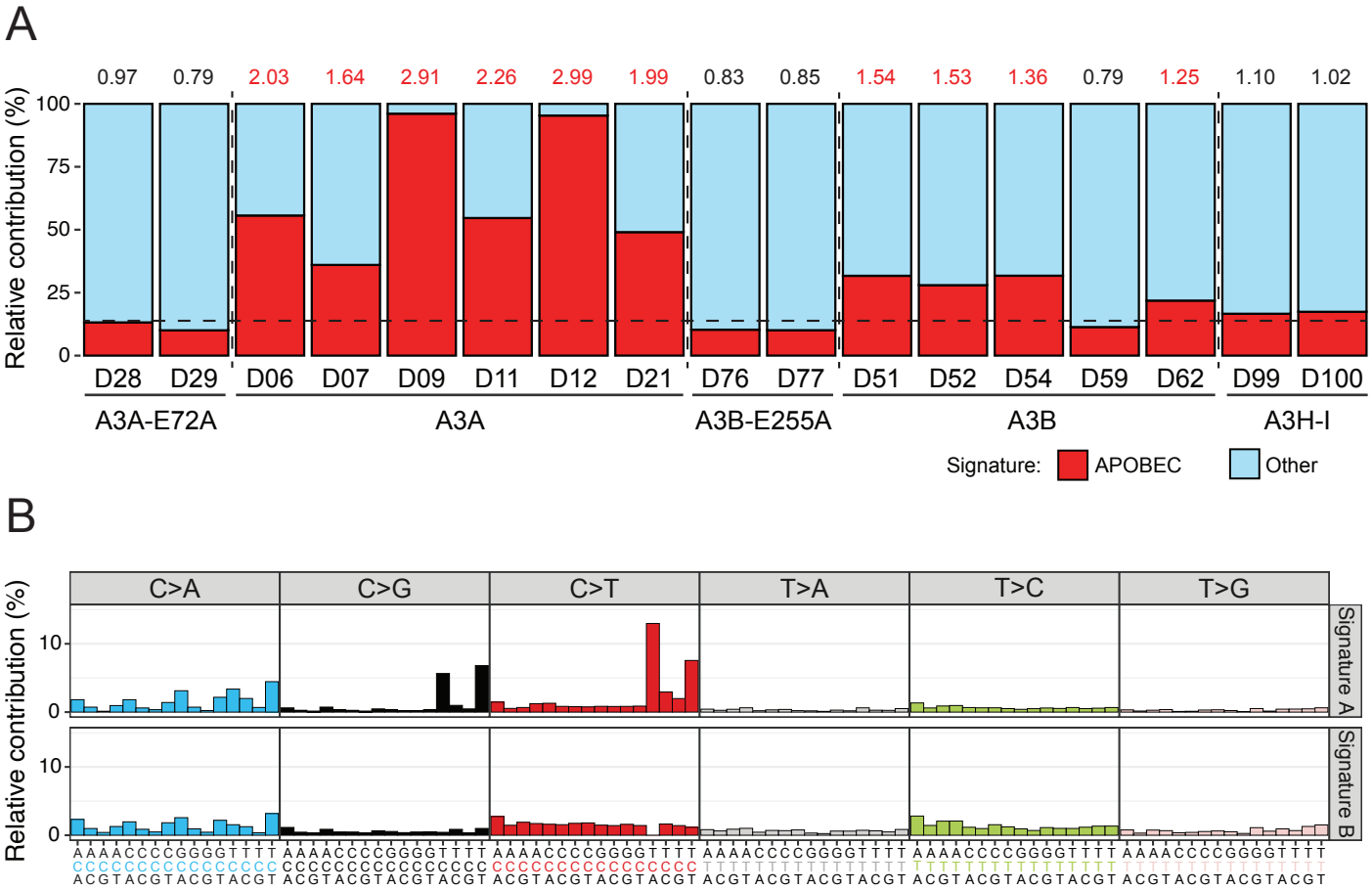


Figure 3 - Figure Supplement 7

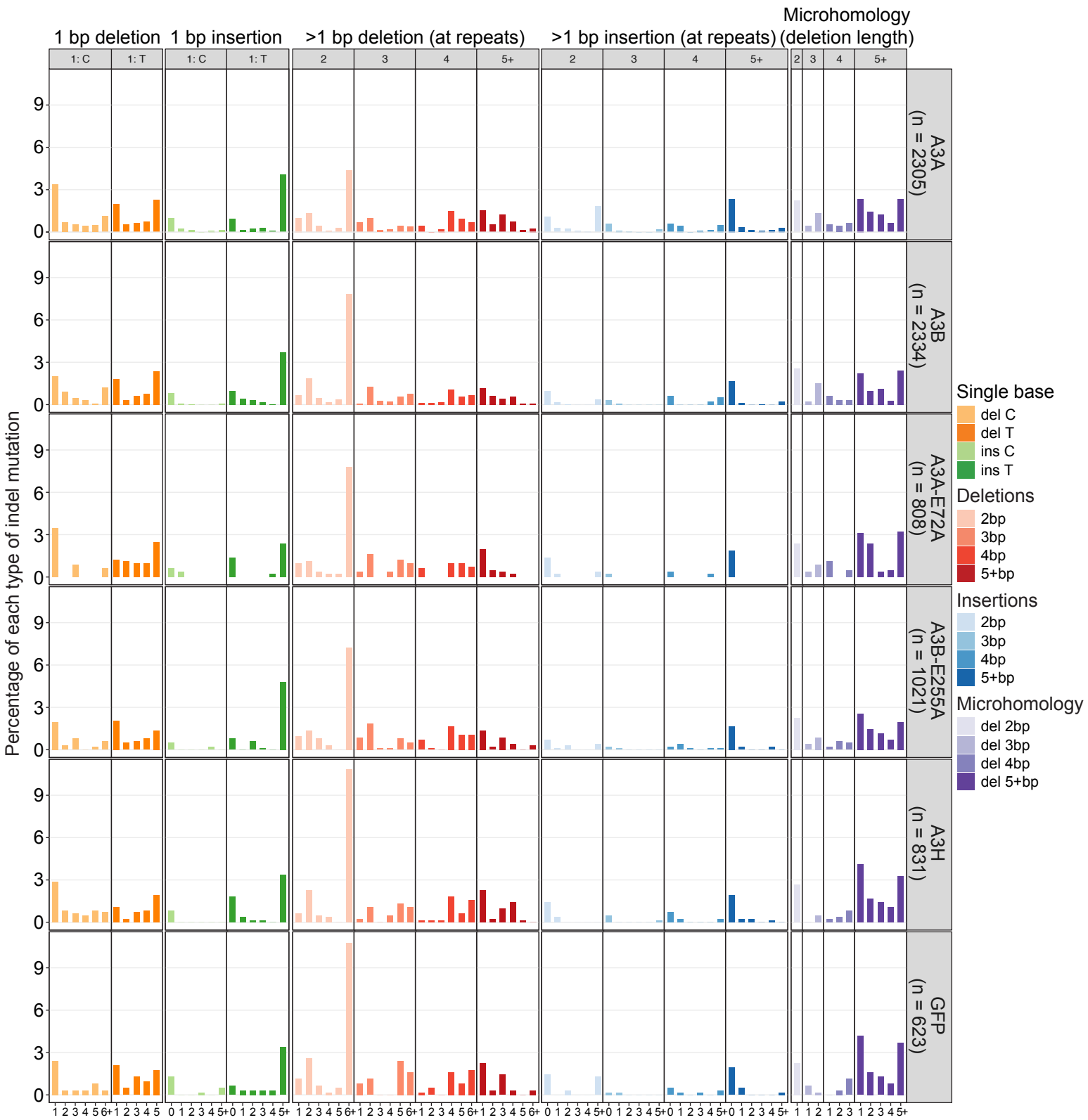


Figure 4

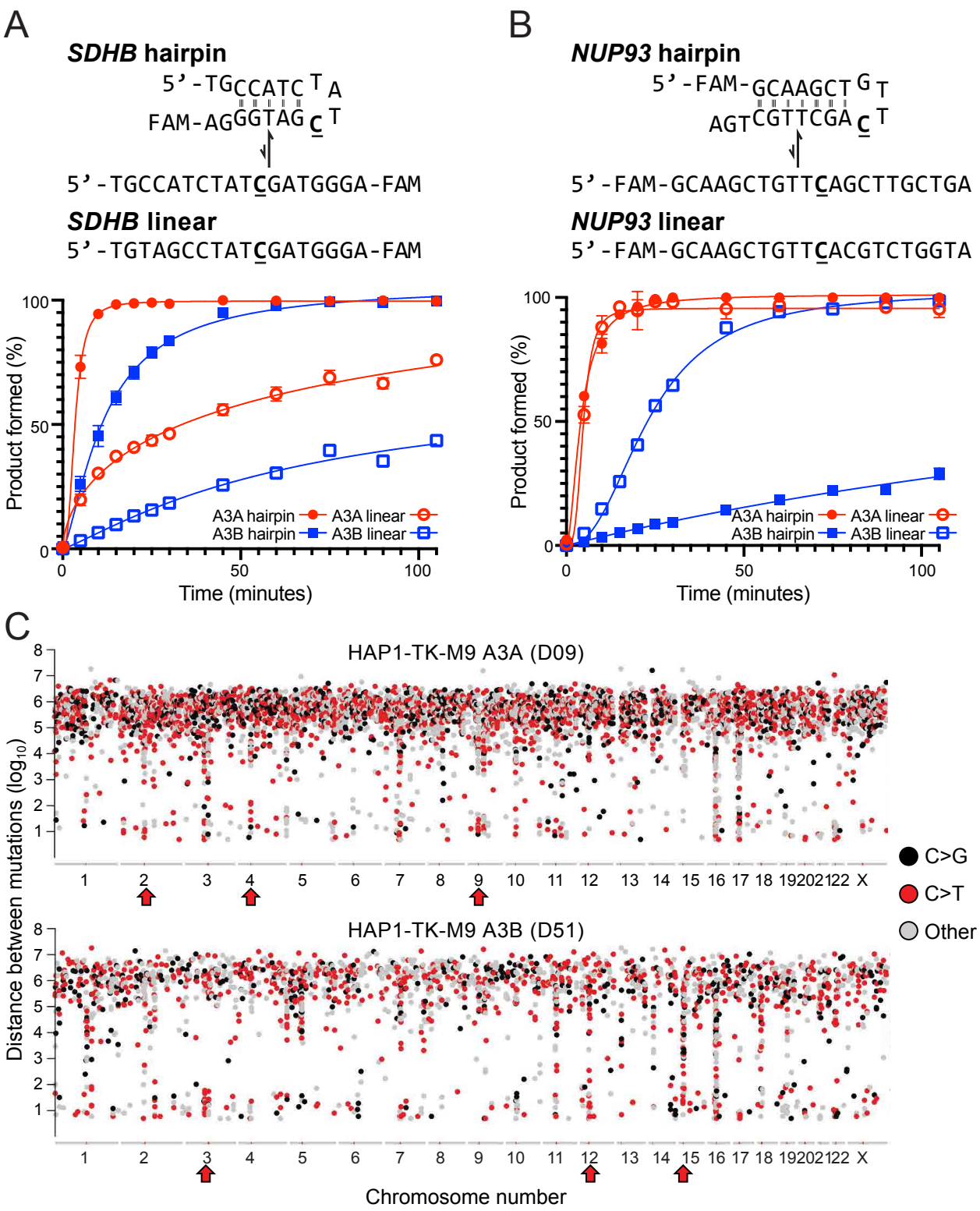

A
SDHB hairpin
5'-TGCCATCTA
FAM-AGGGTAG T
5'-TGCCATCTATCGATGGGA-FAM
SDHB linear
5'-TGTAGCCTATCGATGGGA-FAM
B
NUP93 hairpin
5'-FAM-GCAAGCTG T
AGTCGTTCGA T
5'-FAM-GCAAGCTGTTCAGCTTGCTGA
NUP93 linear
5'-FAM-GCAAGCTGTTCACGTCTGGTA
Product formed (%)
Time (minutes)
A3A hairpin
A3A linear
A3B hairpin
A3B linear
C
HAP1-TK-M9 A3A (D09)
HAP1-TK-M9 A3B (D51)
Distance between mutations (log10)
Chromosome number
C>G
C>T
Other

Figure 4 - Figure Supplement 1

Condition	Clone	Number of SBS mutations	Number of non-hairpin TCW mutations	Non-hairpin TCW mutations per mbp	Number of hairpin loop TCW mutations	Hairpin loop TCW mutations per mbp
A3A-E72A	D28	1363	66	0.022	0	0
A3A-E72A	D29	2204	95	0.032	0	0
A3A	D06	3164	533	0.178	3	0.348
A3A	D07	2337	325	0.109	1	0.116
A3A	D09	5256	2066	0.691	15	1.739
A3A	D11	2057	536	0.179	1	0.116
A3A	D12	5057	2014	0.673	16	1.855
A3A	D21	3093	556	0.186	4	0.464
A3B-E255A	D76	2078	74	0.025	0	0
A3B-E255A	D77	1646	60	0.020	0	0
A3B	D51	2346	361	0.121	4	0.464
A3B	D52	2247	340	0.114	1	0.116
A3B	D54	2504	353	0.118	2	0.232
A3B	D59	1920	259	0.087	1	0.116
A3B	D62	2652	206	0.069	1	0.116
A3H-I	D99	1772	120	0.040	0	0
A3H-I	D100	1732	121	0.040	0	0
GFP	D168	1230	53	0.018	0	0
GFP	D169	1850	89	0.030	0	0

Figure 4 - Figure Supplement 2

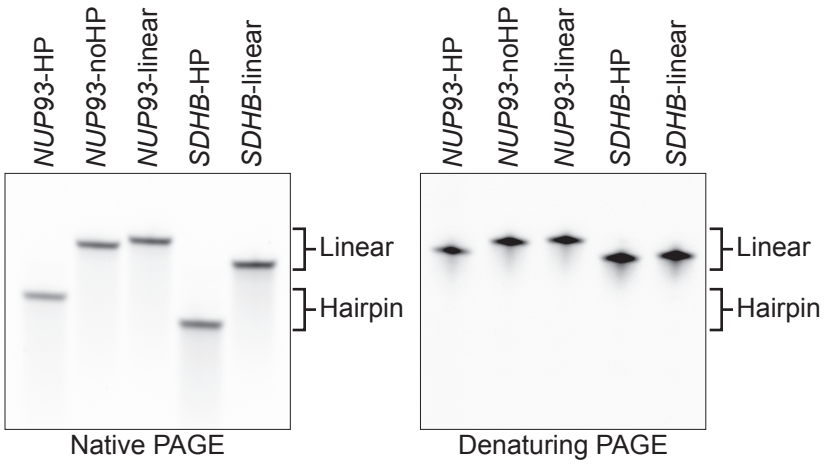
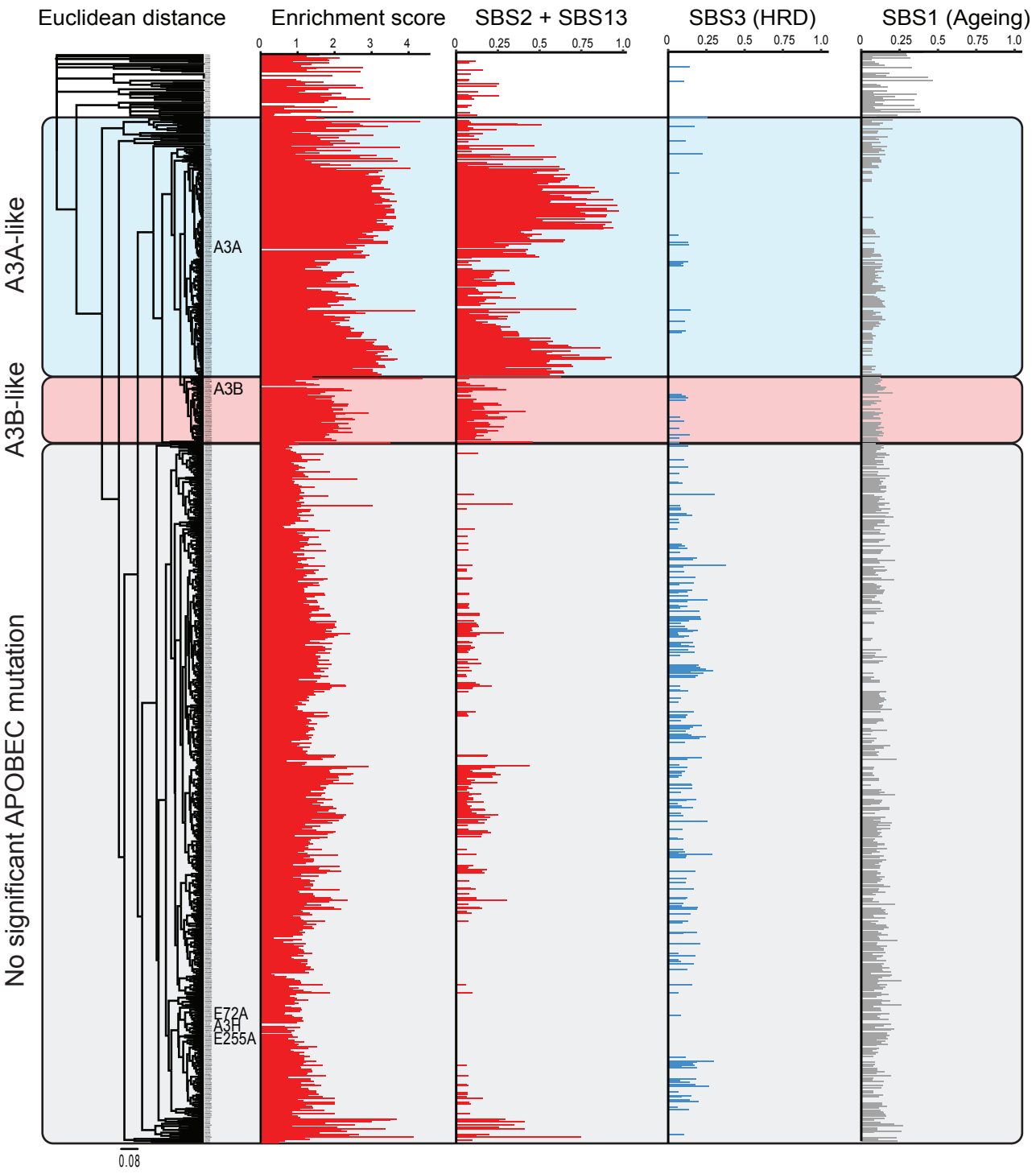


Figure 5


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