Mutational impact of APOBEC3B and APOBEC3A in a human cell line

Running title: Mutational signatures of APOBEC3B and APOBEC3A

Matthew C. Jarvis[1,2,*], Michael A. Carpenter[1,2,3,4,5,*], Nuri A. Temiz[2,6,*], Margaret R. Brown[1,2], Kyle A. Richards[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: 64 character title; 196 word abstract; 5900 word main text (excluding methods, references, and legends); 5 Main Figures and 10 Supplementary Figures

Abstract

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

Introduction

Over the past decade, advances in DNA sequencing technologies and bioinformatics have helped to deconvolute a multitude of mutational processes that contribute to the genesis and evolution of cancer (see recent pan-cancer analysis by [1] and reviews by [2-4]). Through these approaches and complementary bench experiments, the APOBEC3 family of single-stranded (ss)DNA cytosine deaminases has emerged as one of the top three sources of single base substitution (SBS) mutation in cancer with particularly large contributions to tumors of the bladder, breast, cervix, lung, and head/neck. APOBEC3 signature mutations in cancer are defined as C-

to-T transitions and C-to-G transversions in 5'-TCW motifs (W = A or T; SBS2 and SBS13, respectively) [1, 5-10]. This definition is conservative because several APOBEC3 enzymes can also accommodate 5'-CG and 5'-methyl-CG ssDNA substrates [9-17], which can also lead to C-to-T transition mutations and overlap with the mutation signature attributable to spontaneous water-mediated deamination of cytosine and methyl-cytosine nucleotides. Spontaneous methyl-cytosine deamination is a clock-like mutation process that occurs predominantly in 5'-methyl-CG motifs and associates positively with a patient's biological age (ageing signature [18]), whereas APOBEC-catalyzed deamination is not evident in normal tissues, not associated with ageing, present in many primary tumors, and often enriched in metastases (APOBEC signature [1, 6, 8, 9, 18-23]).

The human APOBEC3 (A3) family is comprised of seven different enzymes with extensive homology and overlapping activities (reviewed by [24-26]), and it is unclear how much (or little) each contributes to the composite APOBEC3 mutation signature evident in tumor DNA sequences. The bulk of evidence favors two enzymes, APOBEC3A (A3A) and APOBEC3B (A3B), though to wildly different degrees depending on the study, and additional work has also implicated APOBEC3H (A3H). Evidence for A3A includes an intrinsic preference for 5'-TC substrates, high catalytic activity (highest of any human DNA deaminase), cell-wide localization, DNA damage responses, induction by the tumor virus HPV, and positive correlations between mRNA levels and tumor APOBEC3 mutation loads [11, 14, 16, 27-34]. Evidence for A3B includes all the same points except this enzyme is several-fold less active, localizes constitutively to the nuclear compartment, is induced by multiple DNA tumor viruses (HPV and polyomaviruses), and associates at both mRNA and protein levels with clinical outcomes [9, 11, 16, 19-21, 33-44]. Importantly, however, the APOBEC3 mutation signature can still accumulate in tumors that lack *A3B* due to a naturally occurring full gene deletion allele [45-47]. This observation with *A3B*-null tumors helped implicate a cell-wide, 5'-TC preferring variant of A3H (haplotype I) in causing an APOBEC mutation signature [46, 48, 49]. Another consideration is that A3A and A3B have been

reported to exhibit broader tetranucleotide preferences upon overexpression in yeast, 5'-YTCW vs RTCW, respectively [50-52].

Here, we report a human cellular system for mutation research and use it to perform unbiased comparisons of the mutagenic potential of A3A, A3B, and A3H. The human cell line HAP1 was engineered to express a single copy of the HSV-1 *thymidine kinase* (*TK*) gene, which enables the drug ganciclovir to be used to select rare *TK* mutants and quantify mutation frequencies. Moreover, the *TK* gene can be amplified readily from ganciclovir-resistant (Gan^R) clones by high-fidelity PCR and sequenced to provide initial assessments of mutation spectra (signatures) prior to undertaking additional experiments such as whole genome sequencing (WGS). Using this system, only expression of A3A and A3B cause significant increases in DNA damage signaling (γ-H2AX positivity), DNA breakage (alkaline COMETs), and Gan^R mutation frequencies. Sanger sequences from panels of individual *TK* mutant clones show a clear 5'-TC-biased mutation pattern including two hotspots, with no obvious stem-loop secondary structures. Whole genome sequences (WGS) further demonstrate that both A3A and A3B can generate the APOBEC3 mutational signatures SBS2 and SBS13. Moreover, although these A3A- and A3B-inflicted mutation signatures each cluster with the APOBEC3 signature extracted from different subsets of breast tumors, neither the A3A- nor the A3B-preferred motifs from this system is able to fully explain the observed composite APOBEC3 mutation signature in cancer suggesting that both enzymes may be acting in concert in most tumors.

Results

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

Model organisms such as *E. coli* and yeast are powerful systems for studying mutagens including DNA deaminases (*e.g.*, original studies with A3 enzymes [34, 53-57]). However, these model organisms do not recapitulate all of the DNA repair and regulatory mechanisms found in human cells. We therefore sought to combine strengths of both approaches by introducing a

single copy of a selectable reporter, the HSV-1 *thymidine kinase* (*TK*) gene, into the genome of the human cell line HAP1. TK confers exquisite sensitivity to the drug ganciclovir and, as for many antimicrobial agents, only *TK*-mutant, ganciclovir-resistant (Gan^R) cells are able to survive selection by this drug. The vast majority of Gan^R mutants can be characterized rapidly by conventional Sanger DNA sequencing because *TK* is a single open reading frame. Moreover, once informative Gan^R mutants are revealed by selection, secondary analyses including WGS can be used to uncover additional and potentially global features of a given mutation process.

The overall experimental workflow is shown in **Figure 1A**. To generate a "mother" clone of the commercially available HAP1 cell line, Sleeping Beauty (SB)-mediated transposition was used to introduce a single copy of a *TK-Neo* cassette into the genome [9]. Neo^R clones were selected with G418, expanded into healthy clonal populations (ca. 10^6 cells/ml), and screened for ganciclovir sensitivity (Gan^S). One mother clone, HAP1-TK-M9, was selected for further studies because it is Gan^S, it is mostly diploid (apart from a few pre-existing chromosome aberrations), it cultures, engineers, and clones well (below), and it shows a favorable *A3* expression profile (**Figure 1B**; **Figure 1 – Figure Supplement 1**). In particular, RT-qPCR showed that its *A3A* and *A3B* mRNA levels are lower than those of the original parent line and that *A3H* mRNA levels are very low and near the detection threshold. Accordingly, very low levels of ssDNA cytosine deaminase activity are detected in whole cell extracts (vector control lanes in **Figure 1C-E**). In addition to low *A3H* mRNA expression levels, genomic DNA sequencing showed that the only *A3H* allele is haplotype III (ΔAsn15), which is known to produce an unstable protein [58-60].

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

A panel of MLV-based human A3 expression constructs was assembled and used to produce viral supernatants and transduce the HAP1-TK-M9 mother clone (**Methods**). Uniform A3-expressing pools were selected with puromycin. A3A expression was demonstrated by immunoblotting and immunofluorescent (IF) microscopy using a newly developed rabbit anti-

human A3A monoclonal antibody, UMN13, that specifically binds this enzyme through a unique N-terminal epitope (**Figure 1C**, **1F**; **Figure 1 – Figure Supplement 2**). Expression of A3B was similarly shown by immunoblotting and IF microscopy using a rabbit anti-human A3B monoclonal antibody 5210-87-13 [61] (**Figure 1D**, **1F**). Expression of A3H haplotypes I and II was shown by immunoblotting and IF microscopy with a rabbit polyclonal antibody [58, 62] (**Figure 1E**, **1F**). Endogenous A3A, A3B, and A3H proteins were undetectable as expected based on low mRNA levels (above). In addition, whole cell extracts of A3A-, A3B-, and A3H-expressing HAP1-TK-M9 cells exhibited ssDNA cytosine deaminase activity, well above background levels from vector control or catalytic mutant pools constructed in parallel (**Figure 1C-E**). None of these experimental conditions caused overt decreases in cellular viability, proliferation rates, or clonogenicity.

We next asked whether these A3 enzymes trigger a DNA damage response and cause DNA breakage in HAP1-TK-M9 cells. The populations described above were stained for the DNA damage marker γ-H2AX and subjected to IF microscopy (representative low-resolution images in **Figure 1F**, quantification in **Figure 1G**). Only A3A and A3B triggered elevated levels of DNA damage as characterized by γ-H2AX staining. This was evident both as increased numbers of γ-H2AX foci and as elevated pan-nuclear γ-H2AX staining intensities, as reported in prior overexpression studies [9, 63, 64]. Catalytic mutant derivatives of A3A and A3B as well as active A3H-I and A3H-II showed vector control levels of γ-H2AX staining. A3A and A3B also exhibited DNA breaks as evidenced by larger tail moments in alkaline COMET assays (representative images in **Figure 1F**, quantification in **Figure 1H**). These results combined to demonstrate that both A3A and A3B are enzymatically active and capable of exerting DNA damage in HAP1-TK-M9 cells, consistent with prior DNA damage studies using different cell lines [9, 44, 65-68]. Despite showing cell-wide (including nuclear) localization and measurable DNA deaminase activity, A3H-I did not trigger significant increases in DNA damage or breakage. This result differs from an earlier report [48], which may have expressed A3H-I to higher levels with an inducible promoter in comparison to the stable levels established here with an MND promoter from an integrated

MLV-based construct (likely also single copy due to the low MOI used initially to establish each A3-expressing pool).

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

To directly test which A3 enzymes cause genomic mutation in the HAP1-TK-M9 system, 24 independent single-cell derived daughter clones were obtained for A3A, A3B, A3H-I, and A3H-II expressing conditions as well as catalytic mutant and vector controls. A classical fluctuation analysis was performed by growing each single cell clone for 1 month to $>10^7$ cells, subjecting each population to selection by ganciclovir, and allowing time for single Gan^R mutant cells to grow into countable colonies. Vector control conditions yielded a median Gan^R mutation frequency of 3 mutants per 5 million cells (mean = 3.6, SD = 2.8). In contrast, A3A- and A3B-expressing clones caused median Gan^R mutation frequencies to rise above 10 mutants per 5 million cells (A3A WT: median = 16, mean = 14, SEM = 2.2, p-value range = 1.7×10^{-4} to 3.1×10^{-5}; A3B WT: median = 11, mean = 12, SEM = 1.8, p-value range = 2.2×10^{-4} to 1.9×10^{-5} by Welch's t-test; **Figure 2A**). In comparison, expression of A3H-I or A3H-II or catalytic mutant derivatives of any of these DNA enzymes or empty vector controls failed to trigger increased Gan^R mutation frequencies (A3H-I WT: median = 3.0, mean = 3.6, SEM = 0.71, p-value range = 0.38 to 0.60; A3H-II WT: median = 2.0, mean = 2.8, SEM = 0.58, p-value range = 0.74 to 0.89 by Welch's t-test; **Figure 2A**).

We next asked what types of genetic alterations led to inactivation of the *TK* gene in Gan^R granddaughter clones derived from the different A3-expressing and control conditions. The *TK* gene was PCR-amplified from genomic DNA of Gan^R granddaughter clones and sequenced via Sanger sequencing. C-to-T and C-to-G mutations in an APOBEC-signature trinucleotide motif (APOBEC), all other single base substitution mutations (other SBS), and all insertion/deletion mutations (INDEL) were placed into groups for comparison (red, black, and blue tics in **Figure 2B**, respectively; individual sequence schematics in **Figure 2 – Figure Supplement 1**). *TK* sequences derived from A3A-expressing granddaughter clones harbored a greater number of

APOBEC mutations [12/20 clones contained at least 1 APOBEC mutation, 22 T(C>T/G)W mutations total, range of 0 to 3 SBS per sequence] relative to catalytic mutant control clones [3/19 clones contained 1 APOBEC mutation, 3 T(C>T/G)W mutations total, range of 0 to 1 SBS per sequence]. Similarly, *TK* sequences derived from A3B-expressing granddaughter clones also harbored a greater number of APOBEC mutations [11/20 clones contained at least 1 APOBEC mutation, 19 T(C>T/G)W mutations total, range of 0 to 3 SBS per sequence] relative to catalytic mutant control clones [2/18 clones contained 1 APOBEC mutation, 2 T(C>T/G)W mutations total, range of 0 to 1 SBS per sequence] (**Figure 2B; Figure 2 – Figure Supplement 1**). Additionally, *TK* cytosine bases 22 (Q8X) and 635 (R212K) emerged as mutational hotspots in both the A3A- and A3B-expressing clones (**Figure 2B**; considered further in **Discussion**). No differences were found in the number of other SBS or INDEL mutations between A3A- or A3B-expressing clones and controls.

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

209

210 **WGS of A3A and A3B clones reveals widespread SBS2 and SBS13 and suggests intrinsic**
211 **preferences**

212 To prepare for WGS of Gan^R granddaughter clones, RNA sequencing was done to confirm
213 expression of each exogenously expressed *A3* construct and compare mRNA levels relative to
214 established breast cell lines and primary breast tumors. All expression values were determined
215 relative to those of the conserved housekeeping gene *TBP* in order to be able to compare with
216 RNAseq data from different cell lines and tumors. Interestingly, the average *A3A* and *A3A-E72A*
217 mRNA levels in Gan^R granddaughter clones was over 5-fold higher than the average endogenous
218 *A3A* expression levels of APOBEC3-signature enriched breast cancer cell lines BT474 and MDA-
219 MD-453, breast cancer cell lines of the CCLE, or breast tumors of TCGA (**Figure 3 – Figure**
220 **Supplement 1A**). In other words, A3A levels were higher in this system due to difficulty identifying
221 a sufficiently weak promoter for constitutive expression. In comparison, *A3B* and *A3B-E255A*
222 mRNA levels in Gan^R granddaughter clones were similar to the averages reported for breast
223 cancer cell lines of the CCLE and breast tumors of the TCGA, and approximately 2-fold lower
224 than those of BT474 and MDA-MD-453 (**Figure 3 – Figure Supplement 1A**). *A3H* (haplotype I)
225 mRNA levels showed greater variance but only two clones were analyzed by RNAseq and WGS
226 due to negative results above with the *TK* mutation analysis.

227 The mRNA expression levels of the other four *A3* genes, as well as *AICDA* (*AID*),
228 *APOBEC1*, *APOBEC2*, and *APOBEC4*, were also quantified and compared with those of *A3A*,
229 *A3B*, and *A3H* (**Figure 3 – Figure Supplement 1B**). Endogenous *A3C* was expressed at similarly
230 high levels in all granddaughter clones, providing a robust internal control. Endogenous *A3F* and
231 *A3G* were expressed at lower but still detectable levels, and endogenous *AICDA*, *APOBEC1*,
232 *APOBEC2*, *APOBEC4*, and *A3D* were expressed at very low or undetectable levels. As expected,
233 levels of ectopically expressed *A3A*, *A3B*, and *A3H* mRNA exceeded those in the HAP1-TK-M9
234 parent clone as well as those in vector expressing granddaughter controls. In addition, protein

level expression of A3A, A3B, and A3H was confirmed in *TK* mutant granddaughter clones by immunoblotting and activity by ssDNA deamination assays (**Figure 3 – Figure Supplement 2**). Finally, the integration site of the *TK* reporter was determined using WGS reads that spanned both the 5' and 3'-ends of the integrated *TK* construct and the flanking human genome sequence. Data from multiple clones demonstrated a single integration site in chromosome 3, which is ~35 kbp from the nearest annotated coding or regulatory element.

To investigate mutational differences genome-wide, Illumina short-read WGS was done for randomly selected Gan^R granddaughter clones (specified by "WGS" in **Figure 2 – Figure Supplement 1**). Mutations unique to each granddaughter were identified by calling SBS variations versus the genomic DNA sequence of the HAP1-TK-M9 mother clone. This approach eliminated any somatic variation that accumulated in the Gan^S mother clone prior to transduction with each of the A3 or control expression constructs. Thus, all new SBS mutations were present in a significant proportion of reads from the Gan^R granddaughter clones and absent from the reads from the Gan^S mother clone and, as such, must have accumulated in the presence of an active A3 enzyme or a catalytic mutant control.

In A3A-expressing Gan^R granddaughter clones, the total number of unique SBSs ranged from 2057 to 5196 (n=6, median=3101, mean=3463, SD=1239). The total number of SBSs in A3B-expressing clones was approximately 2-fold lower, ranging from 1920 to 2622 (n=5, median=2346, mean=2334, SD=249). In comparison, the total number of SBSs in catalytic mutant control Gan^R granddaughter clones ranged from 1646 to 2182 (n=4; median=1913, mean=1913, SD=222). These results are summarized in dot plots in **Figure 3 – Figure Supplement 3**. Most importantly, analyses of the trinucleotide contexts of all unique SBS mutations revealed strong C-to-T and C-to-G mutation biases in 5'-TCA and 5'-TCT motifs in A3A-expressing clones and weaker, but still significant, mutation biases in the same motifs in A3B-expressing clones (**Figure 3A**; **Figure 3 – Figure Supplement 4**). In other words, only A3A- and A3B-expressing granddaughter clones exhibited APOBEC3 signature mutations.

This key result was confirmed using an independent metric, APOBEC3 signature enrichment scores [50, 69], which indicated that 6/6 A3A-expressing clones and 4/5 A3B-expressing clones have significant enrichments for APOBEC3 signature mutations, whereas clones expressing catalytically inactive A3A or A3B, as well as clones expressing A3H-I or vector control have none (**Figure 3 – Figure Supplement 5A**). An orthogonal bioinformatics approach, non-negative matrix factorization (NMF [70]), yielded similar results with "signature A" resembling SBS2 and SBS13 in A3A- and A3B-expressing clones (*i.e.*, APOBEC3 signature) and "signature B" occurring in all clones regardless of A3 functionality (**Figure 3 – Figure Supplement 5B**). In comparison, patterns of insertion/deletion (indel) mutations do not appear to be affected significantly by A3A or A3B (**Figure 3 – Figure Supplement 6**). Taken together, these experiments constitute the first cause-and-effect demonstration in a human cell line that both A3A and A3B can generate a clear APOBEC3 mutation signature (*i.e.*, both SBS2 and SBS13).

We next analyzed the broader pentanucleotide contexts of the 5'-TC-focused single base substitution mutations that accumulated in A3A- and A3B-expressing clones (n=6520 and n=1590, respectively) in comparison to those that accumulated in aggregate control clones (n=2934) as well as the overall distribution of 5'-TC in the human genome (n=339619283) (**Figure 3B**). First, a bias for +1 A over +1 T emerged in A3A-expressing clones (43.5% > 39.6%), whereas the opposite bias was evident in A3B-expressing clones (32.6% < 44.7%). For both enzymes, the percentage of +1 A and T (W) was similar (83.1% and 77.3%, respectively). Second, no significant bias was noted at the +2 position except a guanine is slightly over-represented in the pentanucleotide motifs derived from A3A-expressing conditions. Third and most importantly, an exceptionally strong bias for a pyrimidine nucleobase (C or T) occurred at the -2 position in A3A-expressing clones (68.6% YTC vs 31.4% RTC). These A3A results resemble the strong -2 pyrimidine bias reported for human A3A in murine hepatocellular carcinomas (70% YTCW vs 30% RTCW) [71] and for human A3A expression in yeast [50, 51]. Fourth, no bias was apparent for pyrimidines at the -2 position in A3B-expressing clones (50.4% YTC vs 49.6% RTC). This latter

result opposes prior data from yeast, where human A3B appeared to have a bias toward RTCW motifs (*i.e.*, RTCW > YTCW [50, 51]). Possible explanations for these contrasting results are included in **Discussion**).

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

X-ray structures have revealed a U-shaped bend in ssDNA substrates bound by A3A and A3B [16, 29], and other studies have indicated that similarly bent ssDNA loop regions of hairpins (*i.e.*, DNA cruciforms) may be preferred substrates for deamination by A3A but not A3B [52, 72, 73]. To ask whether this might extend to the HAP1-TK-M9 system described here, we analyzed our A3A and A3B *TK* PCR sequences and granddaughter clone WGS data for evidence of mutagenesis in the single-stranded loop regions of DNA hairpin structures. First, neither of the two A3A/B mutation hotspots in the *TK* gene reported above are part of obvious stem-loop structures. Second, none of the top-100 cruciform structures reported previously to harbor recurring APOBEC3 signature mutations in tumors [52] were mutated in our HAP1-TK-M9 WGS data sets. Third, in global comparisons of APOBEC3 signature mutations in predicted loop regions of stem-loop structures versus APOBEC3 signature mutations in non-cruciform structures, neither the A3A- nor the A3B-expressing conditions showed an increased proportion of TCW mutations in stem-loop regions relative to control conditions (Fisher's exact test, q = 1 for all comparisons; **Figure 4 – Figure Supplement 1**).

To assess relative rates of A3A and A3B-catalyzed deamination of hairpin versus non-hairpin substrates, we used ssDNA substrates for two previously reported A3A mutational hotspots in *SDHB* and *NUP93* [52, 74]. A3A and A3B were affinity-purified from human cells and incubated with these C-containing hairpin substrates over time. First, both A3A and A3B showed a strong preference for deaminating the *SDHB* hairpin substrate in comparison to a linear control with the same nucleobase content scrambled (~4- and ~8-fold preference, respectively; **Figure 4A**). In addition, A3A shows higher rates of deamination than A3B on both the hairpin and the

linear substrate in agreement with prior studies [52]. Thus, the relative deamination rates for *SDHB* substrates are: A3A/hairpin > A3B/hairpin > A3A/linear > A3B linear (120, 41, 31, and 5.1 nM/min, respectively).

However, a different picture emerged from analyses of deamination of *NUP93*-based substrates **Figure 4B**). Rates of A3A-catalyzed deamination were similarly high for the *NUP93* hairpin and linear control with the same nucleobase content scrambled. In contrast, A3B showed higher rates of deamination of the linear substrate and was only able to deaminate the hairpin substrate with low efficiencies and linear kinetics. Thus, the relative deamination rates for *NUP93* substrates are: A3A/hairpin = A3A/linear >> A3B linear > A3B/hairpin (93, 83, 7.7, and 1.3 nM/min, respectively). These results show that reaction rates can vary massively between substrates and suggest that caution should be used when attempting to extend individual structural and biochemical preferences to whole genomes. Nevertheless, these data conservatively show that *both* A3A and A3B can deaminate hairpin and linear substrates and, as proposed [52], it is possible to find sites such as the *NUP93* hairpin that are strongly (but not exclusively) preferred by a single A3 enzyme.

Another mesoscale feature of APOBEC3 mutagenesis in human cancer is clusters of strand-coordinated cytosine SBS mutations in TCA and TCT motifs most likely caused by processive deamination of exposed tracts of ssDNA (aka. *kataegis*) [5, 7, 70, 75]. No APOBEC3 signature *kataegic* events were observed in control conditions. However, several APOBEC3 signature *kataegic* events were evident in the genomic DNA of both A3A and A3B expressing granddaughter clones (**Figure 4C**). For instance, one A3A-attributable *kataegic* event was comprised of 7 T(C>T/G)A/T events within a 10 kbp window, and an A3B-attributable *kataegic* event included 8 T(C>T/G)A/T events within a 10 kbp window. Interestingly, however, the frequency of *kataegic* events did not differ significantly between A3A- and A3B-expressing granddaughter clones (A3A WT: median = 30, mean = 35 events, SD = 8.3; A3B WT: median = 12, mean = 15, SD = 6.9; p = 0.28 by Welch's t-test). These results are the first to indicate in a

339 human cell line that both A3A and A3B can cause kataegis.

340

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

342 Sequencing data from model systems such as HAP1-TK-M9 are powerful because the
343 resulting mutation signatures can help to establish cause-and-effect relationships for comparison
344 to more complex tumor WGS data sets to identify similarities and, potentially, to infer the precise
345 source of an observed mutation signature in a given tumor. We therefore performed an
346 unsupervised clustering analysis to compare the pentanucleotide cytosine mutation signatures
347 derived from sequencing A3A, A3B, and A3H expressing HAP1 clones and those from 794
348 primary breast tumors with WGS available through the ICGC data portal resource. This analysis
349 revealed three distinct tumor groups with respect to APOBEC3 signature mutations: 1) a group
350 that showed similarity to A3A-expressing HAP1-TK-M9 clones, 2) a group that showed similarity
351 to A3B-expressing HAP1-TK-M9 clones, and 3) a group that showed little to no significant
352 APOBEC3 mutation signature (**Figure 5**). As expected from a lack of substantial APOBEC3
353 signature mutations above in the *TK* reporter or in representative WGS, A3H-I expressing clones
354 and catalytic-inactive clones clustered with the non-APOBEC signature tumors.

355 Importantly, both the A3A-like and A3B-like groups were comprised of tumors that showed
356 high levels of APOBEC3 signature mutations and correspondingly high enrichment scores.
357 However, even in the A3A-like group, none of the breast tumors showed $\geq$70% APOBEC3
358 signature mutations in YTCW motifs (the overall bias observed here in the HAP1-TK-M9 system
359 and previously in murine tumors [71]). This key result strongly suggests that the observed
360 APOBEC3 mutation signatures in most breast tumors may be a composite resulting from both
361 A3A and A3B activity. In comparison, the homologous recombination repair deficiency signature
362 (SBS3) appeared underrepresented in both the A3A- and A3B-like tumor groups, and the ageing
363 signature (SBS1) occurred in all three groups regardless of A3 mutation status.

364

Discussion

Here we report the development and implementation of a novel system to investigate mutational processes in human cells (**Figure 1**). Like many bacterial and yeast model systems, the HAP1-TK-M9 system enables a uniform cytotoxic selection with ganciclovir such that only *TK* mutant cells survive. An analysis of clonally derived, A3A and A3B expressing *TK* mutants by Sanger sequencing of high-fidelity PCR amplicons demonstrates a strong shift in the mutational pattern from a variety of different base substitution mutations in control conditions to a strongly 5'TCW-biased pattern (**Figure 2**). Interestingly, the *TK* mutation spectra inflicted by A3A and A3B are virtually indistinguishable including two shared hotspots (Q8X, R212K) and no obvious mesoscale features. This may be due to the limited number of mutable cytosines in *TK* that confer resistance to ganciclovir (local base composition) and/or to selective pressure. Regardless of the precise explanation, an analogy can be drawn with the mutational spectrum of the *PIK3CA* gene in breast, head/neck, and others cancers, which has two prominent APOBEC3 mutation hotspots (E542K, E545K) and no obvious mesoscale structures [52, 76]. These observations combine to suggest that selective pressure has the potential to overshadow the intrinsic preferences of the different APOBEC3 enzymes and complicate assignment of direct cause-and-effect relationships.

Drawing direct connections between A3A and/or A3B and a given mutation, even a prominent hotspot, is additionally challenging due to the fact that both enzymes can deaminate DNA cytosines in linear substrates as well as single-stranded loop regions of stem-loop structures (*e.g.*, **Figure 4** and prior biochemical studies [52, 72, 73]). Thus, the *TK*-based system described here is capable of yielding informative, rapid, and inexpensive mutation data sets with positive results motivating genome-wide analyses where the vast majority of mutations are unselected and the larger mutation data sets will enable even stronger conclusions. A potential caveat is that an exclusive focus on Gan^R clones might overestimate the mutational impact of a given process. Earlier work also leveraged HAP1 cells to characterize a variety of different mutation sources but these studies did not incorporate a selectable reporter nor did they address the enzyme(s)

391 responsible for APOBEC3 signature mutations [77, 78].

392 The studies here are the first to demonstrate unambiguously that both A3A and A3B can
393 inflict an APOBEC3 mutation signature in human genomic DNA with, in both instances, ssDNA
394 deamination events immortalizing predominantly as C-to-T and C-to-G mutations in TCA and TCT
395 trinucleotide motifs (**Figure 2** and **Figure 3**). Over identical month-long timeframes, A3A causes
396 4.1-fold more APOBEC3 signature mutations in comparison to A3B (6520 vs 1590 mutations in 6
397 and 5 subclone WGSs, respectively). This difference can be explained in part by elevated A3A
398 expression levels in the HAP1-TK-M9 system (beyond levels in tumors or cell lines) and in part
399 by the higher intrinsic activity of this enzyme in comparison to A3B. Regardless, both A3A and
400 A3B yielded high levels of TC-focused APOBEC3 signature mutations, which enabled
401 comparisons between extended intrinsic preferences. Most importantly, A3A has a stronger
402 preference for a pyrimidine at the -2 position relative to the target cytosine (68.6% YTCW for A3A
403 vs 50.4% YTCW for A3B). This -2 pyrimidine bias mirrors recent WGS results from human A3A-
404 induced murine hepatocellular carcinomas (70% YTCW [71]), and it also resembles the bias
405 reported originally for human A3A expression in yeast [50, 51]. It also mirrors a strong -2
406 pyrimidine bias reported for A3A in the chicken B cell line DT40 published during preparation of
407 this manuscript [79]. However, importantly, A3B also shows a modest YTCW bias here in human
408 cells, which is different from the RTCW bias reported in yeast [50, 51]. This may be due to
409 differences in genetic and/or epigenetic factors including but not limited to base content,
410 chromatin state, DNA repair processes, and/or regulatory factors specific to human A3B
411 expression in human cells. Regardless of the precise explanation(s), the fact that both A3A and
412 A3B can exhibit YTCW mutational biases in human cells helps to inform interpretations of the
413 APOBEC3 enzyme responsible for the overall APOBEC3 mutation program in cancer. For
414 instance, the observed APOBEC3 mutation signature in tumors is likely to be a composite of A3A
415 and A3B because very few (if any) human cancers exhibit a 70% YTCW mutation biases indicative
416 of A3A exclusivity or a 52% YTCW bias indicative of A3B exclusivity. In support of this inference,

our unsupervised clustering analysis APOBEC3-signature motifs from HAP1-TK-M9 WGSs here and nearly 800 breast tumor WGSs identified both A3A- and A3B-like subgroups, with neither enzyme's preferred motif fully explaining the composite signature across breast cancer (**Figure 5**).

In addition to selective pressures and mesoscale features, additional factors are likely to influence the APOBEC3 signature component of an overall tumor mutational landscape including whether A3A and/or A3B is expressed, expression levels, duration of expression, intrinsic activity, and accessibility of chromosomal DNA (replication stress, R-loop levels, chromatin state, *etc*.). With regards to studies here with the HAP1-TK-M9 system, both A3A and A3B are expressed constitutively from the same promoter/construct for identical durations prior to ganciclovir selection, A3A is intrinsically more active than A3B, A3A is cell-wide and A3B predominantly nuclear, and yet these enzymes and cellular factors combine to yield remarkably similar *TK* mutation frequencies and only a 4-fold difference in overall genome-wide SBS mutation level. With respect to cancer, the *A3A* gene is expressed at lower levels than *A3B* in almost all cell lines and tumors, A3A is cell-wide or predominantly cytoplasmic where A3B is constitutively nuclear, and A3A has higher enzymatic activity that can vary from 2- to 100-fold above that of A3B depending on substrate (*e.g*., **Figure 4** and prior biochemical studies [11, 52]). It is therefore notable here that the overall genome-wide level of APOBEC3 signature mutation from A3A is only ~4-fold higher than that attributable to A3B. Endogenous *A3A* and *A3B* also have both distinct and overlapping transcription programs, and both genes can be induced by a variety of conditions including viral infection and inflammation [28, 33, 35-37, 39, 42, 80-86]. *In vivo*, *A3A* and *A3B* gene expression is also likely to be affected by the local tumor microenvironment, which can vary both between and within cancer types, as well as by a patient's global state of health. Taken together with unknown and likely lengthy multi-year durations of pre-cancer and early cancer development prior to clinical manifestation, deducing the exact fractions of mutations attributable to A3A and/or A3B may be fruitless. Rather, it may be more prudent to focus on developing

443 strategies to simultaneously diagnose and treat the contributions of both of these enzymes.

444 Independent whole genome sequencing experiments have provided additional information
445 on the APOBEC3 mutation process. Initial studies induced overexpression of A3B in HEK-293-
446 derived cell lines, documented the resulting DNA damage responses, and performed WGS to
447 assess genome-wide associations [65, 67]. However, an unambiguous APOBEC3 mutation
448 signature was difficult to extract from these whole genome sequences due to large numbers of
449 mutations attributable to compromised mismatch repair [65, 67]. A more recent study compared
450 *de novo* mutations occurring in APOBEC3 signature positive cell lines during multiple generations
451 of clonal outgrowth [87]. An intriguing finding from this work is that APOBEC3 signature mutations
452 may be able to occur in an episodic manner, accumulating in some generations and not others,
453 consistent with evidence discussed above that A3A and A3B expression can be induced by
454 multiple signal transduction pathways. Episodic mutagenesis, however, is unexpected in cell-
455 based systems in which stochastic mutagenesis should predominate given defined media and
456 well-controlled growth conditions. These studies were followed-up more recently by WGS
457 comparisons of subclones of the same cancer cell lines CRISPR-engineered to lack *A3A*, *A3B*,
458 or both genes [88]. The results of over 250 WGS combined to indicate that A3A may be the source
459 of a large fraction of observed APOBEC3 signature SBS mutations, A3B a smaller fraction, and
460 another as-yet-undefined APOBEC3 enzyme an additional minor fraction. These data are
461 complementary to the major results here, with both A3A and A3B proving capable of generating
462 genome-wide APOBEC3 signature mutations. Differences in the overall magnitude of A3A vs A3B
463 mutagenesis may be due to the factors described above including differential intrinsic activity,
464 protein expression levels, genomic DNA accessibility, cell culture conditions, and importantly
465 durations of mutagenesis. Both studies were necessarily done in model cellular systems, each
466 with obvious strengths, but neither is able to fully recapitulate the wide repertoire of factors that
467 impact the actual pre- and post-transformation environments *in vivo*, which are further likely to
468 differ between different tissue and tumor types.

A role for A3H, haplotypes I or II, in cancer is disfavored by our results here showing that these variants are incapable of eliciting DNA damage responses or increasing the *TK* mutation frequency. Two A3H-I expressing *TK* mutant clones were subjected to WGS and no APOBEC3 mutation signature was evident. In addition, no specific evidence for A3H emerged from sequencing clonally-derived cancer cell lines [88]. However, all of these cell-based studies have limitations as discussed above and have yet to fully eliminate A3H as a source of APOBEC3 signature mutations in cancer. For instance, A3H-I may take more time to inflict detectable levels of mutation, it may be subject to different transcriptional and post-transcriptional regulatory processes, and/or it may only be mutagenic in a subset of cancer types subject to different stresses and different selective pressures.

Ultimately, the studies here show that A3A and A3B are each individually capable of triggering overt DNA damage responses and inflicting a robust APOBEC3 mutation signature in human cells and, taken together with other work summarized above, support a model in which both of these enzymes contribute to the composite APOBEC3 mutation signature reported in many different tumor types. This conclusion is supported by clinical studies implicating A3A and/or A3B in a variety of different tumor phenotypes including drug resistance/susceptibility, metastasis, and immune responsiveness [9, 20, 21, 27, 32, 42, 47, 89-95]. Thus, efforts to diagnose and treat APOBEC3 signature-positive tumors should take both enzymes into account, not simply one or the other. Such longer-term goals are not trivial given the high degree of identity between A3A and the A3B catalytic domain (>90%), the related difficulty of developing specific and versatile antibodies for detecting each enzyme, and the fact that each can be regulated differentially by a wide variety of common factors including virus infection and inflammation. Thus, we are hopeful that the HAP1-TK-M9 system and A3A-specific rabbit monoclonal antibody described here will help to expedite the achievement of these goals.

Materials and Methods

Key resources table

Reagent type (species) or resource	Designation	Source or reference	Identifiers	Additional information
Cell line (*homo sapiens,* male)	HAP1	Horizon	Cat#: C859	
Cell line (*homo sapiens,* male)	HAP1-TK-M9	Original clone, this study		Near diploid clone; expresses HSV-1 *TK* and is Gan^{S}
Cell line (*homo sapiens,* female)	293T	ATCC	Cat#: CRL-3216 RRID: CVCL_0063	
Cell line (*homo sapiens,* male)	THP1	ATCC	Cat#: TIB-202 RRID: CVCL_0006	
Antibody	Anti-A3A (rabbit Monoclonal)	This study	UMN13	WB (20 ng/mL)
Antibody	Anti-A3B (rabbit monoclonal)	[61]	Cat#:12397 RRID:AB_2721202	WB (1:1,000) IF (1:300)
Antibody	Anti-A3H (rabbit polyclonal)	Novus ARP10	Cat#: NBP1-91682 RRID: AB_2057523	WB (1:1,000) IF (1:300)

Antibody	Anti-γ-H2AX (S139) (rabbit polyclonal)	Millipore Sigma JW301	Cat#: H5912 RRID: AB_310406	IF (1:300)
Antibody	Anti-Tubulin (mouse monoclonal)	Sigma-Aldrich	Cat#: T5168 RRID: AB_477579	WB (1:20,000)
Recombinant DNA reagent	pQCXIP MLV vector	NovoPro	Cat#: V010396#	Parental transduction vector
Recombinant DNA reagent	pQCXIP-MND-A3Ai-IRES-Puro	This study	A3A with intron (pRH9978)	Request by contacting RSH
Recombinant DNA reagent	pQCXIP-MND-A3Ai-E72A-IRES-Puro	This study	A3A-E72A with intron (pRH9979)	Request by contacting RSH
Recombinant DNA reagent	pQCXIP-MND-A3Bi-IRES-Puro	This study	A3B with intron (pRH9980)	Request by contacting RSH
Recombinant DNA reagent	pQCXIP-MND-A3Bi-E255A-IRES-Puro	This study	A3B E255A with intron (pRH9981)	Request by contacting RSH
Recombinant DNA reagent	pQCXIP-MND-eGFP-IRES-Puro	This study	eGFP vector control (pRH9977)	Request by contacting RSH
Recombinant DNA reagent	pQCXIP-MND-A3H-Ii-IRES-Puro	This study	A3H hapI with intron (pRH9984)	Request by contacting RSH

Recombinant DNA reagent	pQCXIP-MND-A3H-IIi-IRES-Puro	This study	A3H hapII with intron (pRH9985)	Request by contacting RSH
Recombinant DNA reagent	pQCXIP-MND-A3H-IIi-E56A-IRES-Puro	This study	A3H hapII E56A with intron (pRH9986)	Request by contacting RSH
Software, algorithm	Burrows-Wheeler Aligner (BWA)		RRID: SCR_010910	
Software, algorithm	Fiji	Fiji	RRID: SCR_002285	
Software, algorithm	GraphPad Prism 6	GraphPad	RRID: SCR_002798	
Software, algorithm	ImageQuant	GE Healthcare	RRID: SCR_014246	
Software, algorithm	Image Studio	LI-COR Biosciences	RRID: SCR_015795	
Software, algorithm	MaxQuant version 1.5.2.8	MaxQuant	RRID: SCR_014485	
Software, algorithm	R for Statistical Computing		RRID: SCR_001905	
Software, algorithm	SpeedSeq		RRID: SCR_000469	

Software, algorithm	VarScan2		RRID: SCR_006849	
Other	BD LSRFortessa Flow Cytometer	BD Biosciences	RRID: SCR_019600	PI stain quantification
Other	LI-COR Odyssey FC	LI-COR	Cat#: 2800	WB imaging
Other	Typhoon FLA 7000	GE Healthcare	Cat#: 29-0044-13	Oligo cleavage assay imaging
Other	MiSeq Sequencing System	Illumina	RRID: SCR_016379	WGS library balancing validation
Other	NovaSeq 6000 Sequencing System	Illumina	RRID: SCR_016387	WGS

Cell lines and culture conditions

All cell lines were cultured at 37°C under 5% CO_2. HAP1 cells and derivatives were grown in IMDM (Invitrogen) supplemented with 10% fetal bovine serum (Sigma), penicillin (100 U/mL), and streptomycin (100 µg/mL). 293T cells were cultured in DMEM (Invitrogen) with 10% fetal bovine serum (Sigma), penicillin (100 U/mL), and streptomycin (100 µg/mL). THP1 cells were cultured in RPMI (ThermoFisher) supplemented with 10% fetal bovine serum (Sigma), penicillin (100 U/mL), and streptomycin (100 µg/mL). TransIT-LT1 (Mirus) was used for all transfections. All parent and clonal lineage cell lines tested negative for mycoplasma using a PCR-based assay

[96]. Puromycin (ThermoFisher) and neomycin (ThermoFisher) were used at 1 μg/mL and 50 μg/mL, respectively. Ganciclovir (ThermoFisher) was used at 5 μM to select *TK* mutant clones.

Creating the HAP1-TK-M9 system

The HAP1-TK-M9 system was generated by co-transfecting HAP1 parent cells with a plasmid expressing the Sleeping Beauty transposase and a separate plasmid with *TK-Neo* coding sequences flanked by SB recognition sites [9, 97]. Semi-confluent cells in 6 well plates were transfected, treated 24 hrs later with G418 (1 μg/mL), and subcloned by limiting dilution in 96 well plates to create single cell derivatives. Single cell clones were then expanded and characterized as described in the main text. The *A3H* genotype was determined by sequencing exon-specific PCR amplicons [98, 99] and further confirmed by WGS (below).

Standard molecular cloning procedures were used to create derivatives of MLV pQCXIP for expressing each A3 protein. First, pQCXIP was cut with MluI and PacI to excise the strong CMV promoter and replace it with a weaker MND promoter (a synthetic promoter containing regions of both the MLV LTR and the myeloproliferative sarcoma virus enhancer). Second, this new construct was cut with SfiI and BsiWI to insert intron-disrupted A3 coding sequences [71]. This was done for *A3A*, *A3B*, *A3H* haplotype-I, *A3H* haplotype-II, and appropriate catalytic mutant derivatives (E-to-A). An eGFP expressing construct was generated in parallel to use as a control in various experiments. All new constructs were confirmed by Sanger sequencing. Each MLV-based construct was co-transfected into 293T cells with appropriate packaging vectors and 48 hrs later the resulting viral supernatants were filtered (0.2 μm) and used to transduce semi-confluent HAP1-TK-M9 cells. After an additional 48 hrs incubation, transduced cells were selected with puromycin (1 μg/mL) and subcloned by limiting dilution to obtain A3 expressing daughter clones. These A3 expressing and control daughter clones were expanded for 1 month and characterized as described in the main text. No overt growth/proliferation defects were noted, and all granddaughter clones expanded at similar rates. Mutation frequencies were determined by

plating 5 x 10^6 cells in 96 well flat bottom plates, treating with 5 μM ganciclovir (ThermoFisher), and after 14 days incubation counting the number of *TK* mutant colonies that survived selection. Single Gan^R granddaughter clones were counted using a light microscope and expanded and characterized as described in the main text including immunoblotting, DNA deaminase activity assays, *TK* sequencing, and WGS.

Immunoblots (IB)

Cells were treated with trypsin EDTA and collected, washed in 1X PBS, and re-suspended in 100 μL of reducing sample buffer per one million cells [0.5 M Tris-HCl pH 6.8, 1% 2-mercaptoethanol, 10% sodium dodecyl sulfate (SDS), 50% glycerol]. Proteins were denatured by boiling samples for 20 min and resolved by SDS-polyacrylamide gel electrophoresis. Proteins were transferred to a PVDF-FL membrane (Millipore Sigma) and blocked in 5% milk in 1X PBS. Primary and secondary antibodies were incubated in blocking buffer, with the addition of 0.2% SDS for fluorescent antibodies. The primary antibodies used were anti-A3A/B (5210-87-13, IB: 1:1,000, IF: 1:300 [61]), anti-A3H (Novus, IB: 1:1,000, IF: 1:300 [58]), anti-Tubulin (Sigma Aldrich, 1:20,000), and anti-γ-H2AX (S139) (Millipore Sigma JBW301, 1:300). The secondary antibodies used were anti-rabbit HRP-linked (CST 7074, 1:2,000), IRDye 800CW goat anti-mouse (LI-COR Biosciences, 1:10,000), Alexa Fluor 680 goat anti-rabbit (Molecular Probes, Eugene, OR, USA, 1:10,000), and Alexa Fluor 680 goat anti-mouse (Molecular Probes, Eugene, OR, USA, 1:10,000). Membranes were imaged using a LI-COR Odyssey instrument or LI-COR Odyssey-Fc instrument for HRP visualization (LI-COR Biosciences).

DNA deaminase activity assays

Whole cell extract (WCE) assays: ssDNA deamination activities were measured using WCE prepared using 100 μL HED lysis buffer per 1 million cells (25 mM HEPES, 5 mM EDTA, 10% Glycerol, 1 mM DTT, and 1 protease inhibitor tablet). Samples were sonicated 3 times for 5

559 sec to ensure complete lysis. A3-containing lysates were incubated at 37°C for 2 hrs with purified
560 human UNG2 and a single-stranded DNA substrate containing a single TCA trinucleotide motif
561 and 3' FAM label (5'-ATTATTATTATTCAAATGGATTTATTTATTTATTTATTTATTT-FAM-3')
562 following established protocols [9, 63]. After this initial incubation, the reaction was treated with 1
563 mM NaOH for 10 min at 98°C. The reaction was run out on a 15% TBE acrylamide urea gel to
564 separate substrate oligo from cleaved product oligo and imaged on a Typhoon FLA 7000 with
565 ImageQuant TL 8.2.0 software (GE Healthcare).

566 Recombinant enzyme assays: A3A- and A3B-mycHis were prepared from transfected
567 293T as reported [14, 16, 100-102]. 25 nM of each protein was incubated for the indicated times
568 with 800 nM substrate in HEPES buffer (25 nM HEPES pH 7.4, 50 mM NaCl, 5 mM imidazole) at
569 37°C. Reactions were stopped by freezing in liquid nitrogen and then were heated to 95°C to
570 denature the enzymes. Reactions were then treated with 0.5 U/reaction Uracil DNA Glycosylase
571 (NEB, USA) for 10 min at 37°C. Sodium hydroxide was added to each reaction to 100 mM and
572 reactions were heated to 95°C for 5 minutes to cleave the DNA. Products were separated by 20%
573 TBE-Urea PAGE, imaged on a Typhoon FLA-7000 (GE Healthcare, USA), and quantified using
574 ImageQuant TL 8.2.0 software (GE Healthcare). Deamination of the target cytosine was
575 calculated by dividing the total reaction product (and any subsequent off-target cytosine
576 deamination events) by the total amount of starting substrate.

577

578 **PI staining and flow cytometry**

579 Propidium iodide staining was used to assess the ploidy of HAP1 clones relative to THP1
580 as a confirmed diploid control. Cells were trypsinized and suspended in 100 µL of 1X PBS per 1
581 million cells. Then, 500 µL ice-cold ethanol was added to cell suspensions and incubated at -20°C
582 for 1 hour to fix the cells. After fixation, cells were pelleted and washed in 1X PBS. Cells were
583 finally suspended in 500 µL of FxCycle PI stain (Invitrogen) and incubated for 30 minutes at RT
584 in the dark to stain the cells. Cells were spun down and resuspended in 300 µL of the PI stain

585 solution and placed in a 96 well round bottom plate for flow cytometry analysis using a BD
586 LSRFortessa flow cytometer (with high-throughput 96 well adapter system). A minimum of ten
587 thousand events were acquired for each cell line.
588
589 **RT-qPCR**
590 Total RNA was extracted using the High Pure RNA isolation kit (Roche). cDNA was
591 synthesized using SuperScript First-Strand RT (ThermoFisher). Quantification of mRNA was
592 done using validated primer sets for all human *A3* genes relative to the housekeeping gene *TBP*
593 [9, 81, 97, 98]. All RT-qPCR reaction were performed using SsoFast SYBR Green mastermix
594 (Bio-Rad) in 384-well plates on a LightCycler 480 (Roche) following the manufacturer's protocol.
595 Statistical analyses were done using GraphPad Prism 6 and R.
596
597 **Immunofluorescent microscopy (IF)**
598 IF was done as described [64, 103]. Cells were grown in 6 well tissue culture plates at low
599 density prior to fixation. Cells were fixed with 4% paraformaldehyde in PBS for 15 minutes at RT.
600 Cells were then permeabilized using PBS containing 0.2% Triton X-100 (Sigma Aldrich), then
601 rinsed with PBS. Cells were blocked using IF blocking solution and 0.1% Triton X-100 for 1 hr at
602 RT, then incubated with primary antibody overnight at 4°C. The primary antibodies used were
603 anti-A3A/B/G (5210-87-13, 1:300) [61], anti-A3A (UMN-13; this study), anti-A3H (Novus, 1:300)
604 [58], and anti-γ-H2AX (JBW301, Millipore Sigma, 1:300). Cells were then washed with PBS and
605 incubated with a fluoro-conjugated secondary antibody for 2 hrs at RT in the dark. The secondary
606 antibodies used were IRDye 800CW goat anti-mouse (LI-COR Biosciences, 1:10,000) and Alexa
607 Fluor 680 goat anti-rabbit (Molecular Probes, Eugene, OR, USA, 1:10,000). Primary and
608 secondary antibodies were diluted in blocking buffer. Hoescht 33342 (Mirus) was used at a final
609 concentration of 1 μg/mL to stain nuclei.
610

COMET assays

HAP1-TK-M9 clones expressing individual A3 enzymes were harvested at 70% confluency and resuspended in ice cold 1X phosphor-buffered saline (PBS, Ca^{2+} and Mg^{2+} free, $1.0x10^5$ cells/mL). The CometAssay kit and manufacturer's protocol were used for all alkaline comet assays (Trevigen). Cells were spread at low density on a glass slide and covered in low-melt agarose to keep the cells in position for lysis. After cell lysis, slides were washed 2 times with 1X UDG buffer (New England Biolabs), and then incubated with purified human UNG2 in 1X UDG buffer (4 ng/µL) for 1 hour at 37°C. After incubation, DNA unwinding and electrophoresis were done according to the manufacturer's alkaline protocol. Comets were analyzed using the OpenComet plugin for Image J [104] and statistical analyses were done in GraphPad Prism 6 and R.

***TK* sequencing**

Cells were harvested and genomic DNA was isolated using the Puregene DNA isolation protocol. The *TK* cassette was amplified from genomic DNA using 5'-ATCTTGGTGGCGTGAAACTC-3' and 5'-CTTCCGGTATTGTCTCCTTCC-3'. PCR products were cleaned-up using the GeneJet Gel Extraction Kit (Thermo Scientific) and Sanger sequenced with 4 different primers to cover the full open reading frame (5'-ATCTTGGTGGCGTGAAACTC-3', 5'-GGTCATGCTGCCCATAAGGTA-3', 5'-CCGTTCTGGCTCCTCATATC-3', and 5'-CTTCCGGTATTGTCTCCTTCC-3'.

Whole genome sequencing (WGS) and analyses

Genomic DNA was prepared from cell pellets (1 million cells) using Allprep DNA/RNA mini kit (Qiagen). Whole genome libraries were sequenced 150x2 bp on a NovaSeq 6000 (Illumina) to a target read depth of 30X coverage for all granddaughter clones as well as the parental HAP1-TK-M9 mother clone. Resulting sequences were aligned to the human genome (hg38) using

637 SpeedSeq [105], which relies on the Burrows-Wheeler Aligner, BWA (version 0.7.17). PCR
638 duplicates were removed using Picard (version 2.18.16). Reads were locally realigned around
639 InDels using GATK3 (version 3.6.0) tools RealignerTargetCreator to create intervals, followed by
640 IndelRealigner on the aligned bam files. Single base substitutions and small InDels were called
641 in each clone relative to the bam file generated from the HAP1-TK-M9 mother clone using Mutect2
642 from GATK3 (version 3.6.0). SBSs that passed the internal GATK3 filter with minimum 4 reads
643 supporting each variant, minimum 20 total reads at each variant site and a variant allele frequency
644 over 0.05 were used for downstream analysis. SBSs were analyzed in R (version 4.0.5) using the
645 MutationalPatterns [106] and deconstructSigs R packages (version 1.8.0 [107]). All visualizations
646 were generated using the ggplot2 package (version 3.3.5). The indel landscapes were generated
647 using the MutationalPatterns R package [106] following PCAWG definitions [1]. All
648 individual clone data from each condition were pooled for presentation.
649 COSMIC single base substitution mutation signatures (v3 – May 2019
650 https://cancer.sanger.ac.uk/cosmic/signatures/SBS/) were obtained from
651 https://www.synapse.org/#!Synapse:syn11738319. *De-novo* non-negative matrix factorization of
652 mutational signatures was performed with the "extract_signature" command from the
653 MutationalPatterns package, with a rank of 2 and 100 iterations. TCW mutation enrichment scores
654 were calculated as described [50, 69]. Sequence logos of -2 to +2 sequence surrounding C-to-T
655 mutations were created using the ggseqlogo (version 0.1) package.
656
657 **APOBEC expression and mutation signature analyses in TCGA and ICGC data sets**
658 TCGA primary breast tumors represented by both RNA-seq and whole exome sequencing
659 were downloaded from the Firehose GDAC resource through the Broad Institute pipeline
660 (http://gdac.broadinstitute.org/) for multiple tumor tissue types. APOBEC mutation signatures
661 were determined as described [5, 69] using the deconstructSigs R package [107]. APOBEC
662 mutation enrichment scores were calculated using the hg19 reference genome and published

663 methods [50]. Enrichment score significance was assessed using a Fisher exact test with
664 Benjamini-Hochberg false discovery rate (FDR) correction. All dataset analyses and
665 visualizations were conducted using R and the ggplot2 package (https://www.R-project.org/).
666
667 **_TK_ integration site determination**
668 To determine the integration site of the single copy *TK-Neo* construct, a TK reference
669 sequence was provided as an additional chromosome during alignment of the WGS reads to the
670 reference genome (hg38). Reads that mapped to this region were then categorized as discordant
671 and realigned using GRIDSS (v2.2) [108] to determine site of integration.
672
673 **Clustering analysis**
674 All primary breast tumor whole genome sequencing variant information from International
675 Cancer Genome Consortium (ICGC) was downloaded from the ICGC data portal
676 (https://dcc.icgc.org). SBSs used in these analyses included only C-to-T variants in TC
677 dinucleotide contexts (TCA, TCC, and TCT) and excluded all mutations in CG motifs due to
678 potential overlap with spontaneous water-mediated methyl-C deamination. SBSs meeting these
679 inclusion criteria from all clones expressing A3A, A3B, A3A-E72A, A3B-E255A, and A3H-I were
680 pooled per condition for this analysis. A matrix comprised of the number of mutations within a
681 pentanucleotide across all samples within a cancer type was generated, and counts were
682 normalized to frequency within each cancer type. The resulting matrix was then clustered using
683 the hclust function in R with the classical Euclidean distance as the distance method for clustering,
684 which was then plotted as dendrograms. Mutation signatures were calculated using
685 deconstructSigs as described above.
686
687 **Data and code availability**

All alignment files (FASTQ and BAM format) are available through the Sequence Read Archive under the BioProject accession number PRJNA832427.

Author contributions

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

Additional contributions

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

Additional information

Funding

These studies were supported by NCI P01-CA234228, NCI P50-CA247749, and a CPRIT Established Investigator Recruitment Award. Salary support for MCJ was provided in part by NCI T32-CA009138 and subsequently NCI F31-CA243306. RSH is an Investigator of the Howard

Hughes Medical Institute and the Ewing Halsell President's Council Distinguished Chair at University of Texas Health San Antonio. The authors have no competing interests to declare.

Compliance with ethical standards

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

References

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

738 Epub 2013/08/16. doi: 10.1038/nature12477 [pii]. PubMed PMID: 23945592.

739 6. Roberts SA, Lawrence MS, Klimczak LJ, Grimm SA, Fargo D, Stojanov P, et al. An APOBEC
740 cytidine deaminase mutagenesis pattern is widespread in human cancers. Nat Genet.
741 2013;45(9):970-6. doi: 10.1038/ng.2702. PubMed PMID: 23852170; PubMed Central
742 PMCID: PMC3789062.

743 7. Nik-Zainal S, Alexandrov LB, Wedge DC, Van Loo P, Greenman CD, Raine K, et al.
744 Mutational processes molding the genomes of 21 breast cancers. Cell. 2012;149(5):979-
745 93. Epub 2012/05/23. doi: S0092-8674(12)00528-4 [pii] 10.1016/j.cell.2012.04.024.
746 PubMed PMID: 22608084.

747 8. Nik-Zainal S, Davies H, Staaf J, Ramakrishna M, Glodzik D, Zou X, et al. Landscape of
748 somatic mutations in 560 breast cancer whole-genome sequences. Nature.
749 2016;534(7605):47-54. doi: 10.1038/nature17676. PubMed PMID: 27135926; PubMed
750 Central PMCID: PMC4910866.

751 9. Burns MB, Lackey L, Carpenter MA, Rathore A, Land AM, Leonard B, et al. APOBEC3B is
752 an enzymatic source of mutation in breast cancer. Nature. 2013;494(7437):366-70. Epub
753 2013/02/08. doi: 10.1038/nature11881 [pii]. PubMed PMID: 23389445.

754 10. Leonard B, Hart SN, Burns MB, Carpenter MA, Temiz NA, Rathore A, et al. APOBEC3B
755 upregulation and genomic mutation patterns in serous ovarian carcinoma. Cancer Res.
756 2013;73(24):7222-31. Epub 2013/10/25. doi: 10.1158/0008-5472.CAN-13-1753. PubMed
757 PMID: 24154874; PubMed Central PMCID: PMC3867573.

758 11. Ito F, Fu Y, Kao SA, Yang H, Chen XS. Family-wide comparative analysis of cytidine and
759 methylcytidine deamination by eleven human APOBEC proteins. J Mol Biol.
760 2017;429(12):1787-99. doi: 10.1016/j.jmb.2017.04.021. PubMed PMID: 28479091;
761 PubMed Central PMCID: PMC5530319.

12. Ito F, Yang H, Xiao X, Li SX, Wolfe A, Zirkle B, et al. Understanding the structure, multimerization, subcellular localization and mC selectivity of a genomic mutator and anti-HIV factor APOBEC3H. Sci Rep. 2018;8(1):3763. doi: 10.1038/s41598-018-21955-0. PubMed PMID: 29491387; PubMed Central PMCID: PMC5830531.

13. Hultquist JF, Lengyel JA, Refsland EW, LaRue RS, Lackey L, Brown WL, et al. Human and rhesus APOBEC3D, APOBEC3F, APOBEC3G, and APOBEC3H demonstrate a conserved capacity to restrict Vif-deficient HIV-1. J Virol. 2011;85(21):11220-34. Epub 2011/08/13. doi: JVI.05238-11 [pii] 10.1128/JVI.05238-11. PubMed PMID: 21835787; PubMed Central PMCID: PMC3194973.

14. Carpenter MA, Li M, Rathore A, Lackey L, Law EK, Land AM, et al. Methylcytosine and normal cytosine deamination by the foreign DNA restriction enzyme APOBEC3A. J Biol Chem. 2012;287(41):34801-8. Epub 2012/08/17. doi: 10.1074/jbc.M112.385161 [pii]. PubMed PMID: 22896697; PubMed Central PMCID: PMC3464582.

15. Suspène R, Aynaud MM, Vartanian JP, Wain-Hobson S. Efficient deamination of 5-methylcytidine and 5-substituted cytidine residues in DNA by human APOBEC3A cytidine deaminase. PloS one. 2013;8(6):e63461. Epub 2013/07/11. doi: 10.1371/journal.pone.0063461 PONE-D-13-02127 [pii]. PubMed PMID: 23840298; PubMed Central PMCID: PMC3688686.

16. Shi K, Carpenter MA, Banerjee S, Shaban NM, Kurahashi K, Salamango DJ, et al. Structural basis for targeted DNA cytosine deamination and mutagenesis by APOBEC3A and APOBEC3B. Nat Struct Mol Biol. 2017;24(2):131-9. doi: 10.1038/nsmb.3344. PubMed PMID: 27991903; PubMed Central PMCID: PMC5296220.

17. Schutsky EK, Nabel CS, Davis AKF, DeNizio JE, Kohli RM. APOBEC3A efficiently deaminates methylated, but not TET-oxidized, cytosine bases in DNA. Nucleic Acids Res.

2017;45(13):7655-65. Epub 2017/05/05. doi: 10.1093/nar/gkx345. PubMed PMID: 28472485; PubMed Central PMCID: PMCPMC5570014.

18. Alexandrov LB, Jones PH, Wedge DC, Sale JE, Campbell PJ, Nik-Zainal S, et al. Clock-like mutational processes in human somatic cells. Nat Genet. 2015;47(12):1402-7. Epub 2015/11/10. doi: 10.1038/ng.3441. PubMed PMID: 26551669; PubMed Central PMCID: PMCPMC4783858.

19. Burns MB, Temiz NA, Harris RS. Evidence for APOBEC3B mutagenesis in multiple human cancers. Nat Genet. 2013;45(9):977-83. Epub 2013/07/16. doi: 10.1038/ng.2701 [pii]. PubMed PMID: 23852168.

20. Sieuwerts AM, Schrijver WA, Dalm SU, de Weerd V, Moelans CB, Ter Hoeve N, et al. Progressive APOBEC3B mRNA expression in distant breast cancer metastases. PloS one. 2017;12(1):e0171343. doi: 10.1371/journal.pone.0171343. PubMed PMID: 28141868; PubMed Central PMCID: PMC5283735 sharing data and materials.

21. Sieuwerts AM, Willis S, Burns MB, Look MP, Meijer-Van Gelder ME, Schlicker A, et al. Elevated APOBEC3B correlates with poor outcomes for estrogen-receptor-positive breast cancers. Hormones & cancer. 2014;5(6):405-13. doi: 10.1007/s12672-014-0196-8. PubMed PMID: 25123150; PubMed Central PMCID: PMC4228172.

22. Angus L, Smid M, Wilting SM, van Riet J, Van Hoeck A, Nguyen L, et al. The genomic landscape of metastatic breast cancer highlights changes in mutation and signature frequencies. Nat Genet. 2019;51(10):1450-8. Epub 2019/10/02. doi: 10.1038/s41588-019-0507-7. PubMed PMID: 31570896; PubMed Central PMCID: PMCPMC6858873.

23. Moore L, Cagan A, Coorens THH, Neville MDC, Sanghvi R, Sanders MA, et al. The mutational landscape of human somatic and germline cells. Nature. 2021;597(7876):381-6. Epub 2021/08/27. doi: 10.1038/s41586-021-03822-7. PubMed PMID: 34433962.

24. Harris RS, Dudley JP. APOBECs and virus restriction. Virology. 2015;479-480C:131-45. doi: 10.1016/j.virol.2015.03.012. PubMed PMID: 25818029; PubMed Central PMCID: PMC4424171.

25. Simon V, Bloch N, Landau NR. Intrinsic host restrictions to HIV-1 and mechanisms of viral escape. Nat Immunol. 2015;16(6):546-53. doi: 10.1038/ni.3156. PubMed PMID: 25988886.

26. Silvas TV, Schiffer CA. APOBEC3s: DNA-editing human cytidine deaminases. Protein Sci. 2019;28(9):1552-66. Epub 2019/06/27. doi: 10.1002/pro.3670. PubMed PMID: 31241202; PubMed Central PMCID: PMCPMC6699113.

27. Cortez LM, Brown AL, Dennis MA, Collins CD, Brown AJ, Mitchell D, et al. APOBEC3A is a prominent cytidine deaminase in breast cancer. PLoS Genet. 2019;15(12):e1008545. Epub 2019/12/17. doi: 10.1371/journal.pgen.1008545. PubMed PMID: 31841499.

28. Westrich JA, Warren CJ, Klausner MJ, Guo K, Liu CW, Santiago ML, et al. Human papillomavirus 16 E7 stabilizes APOBEC3A protein by inhibiting Cullin 2-dependent protein degradation. J Virol. 2018;92(7). Epub 2018/01/26. doi: 10.1128/JVI.01318-17. PubMed PMID: 29367246; PubMed Central PMCID: PMCPMC5972886.

29. Kouno T, Silvas TV, Hilbert BJ, Shandilya SMD, Bohn MF, Kelch BA, et al. Crystal structure of APOBEC3A bound to single-stranded DNA reveals structural basis for cytidine deamination and specificity. Nat Commun. 2017;8:15024. Epub 2017/04/30. doi: 10.1038/ncomms15024. PubMed PMID: 28452355; PubMed Central PMCID: PMCPMC5414352.

30. Green AM, Landry S, Budagyan K, Avgousti DC, Shalhout S, Bhagwat AS, et al. APOBEC3A damages the cellular genome during DNA replication. Cell Cycle. 2016;15(7):998-1008. doi: 10.1080/15384101.2016.1152426. PubMed PMID: 26918916.

834 31. Chen H, Lilley CE, Yu Q, Lee DV, Chou J, Narvaiza I, et al. APOBEC3A is a potent inhibitor
835 of adeno-associated virus and retrotransposons. Curr Biol. 2006;16(5):480-5. PubMed
836 PMID: 16527742.
837 32. Faden DL, Thomas S, Cantalupo PG, Agrawal N, Myers J, DeRisi J. Multi-modality analysis
838 supports APOBEC as a major source of mutations in head and neck squamous cell
839 carcinoma. Oral oncology. 2017;74:8-14. doi: 10.1016/j.oraloncology.2017.09.002.
840 PubMed PMID: 29103756.
841 33. Warren CJ, Xu T, Guo K, Griffin LM, Westrich JA, Lee D, et al. APOBEC3A functions as a
842 restriction factor of human papillomavirus. J Virol. 2015;89(1):688-702. doi:
843 10.1128/JVI.02383-14. PubMed PMID: 25355878; PubMed Central PMCID:
844 PMC4301161.
845 34. Taylor BJ, Nik-Zainal S, Wu YL, Stebbings LA, Raine K, Campbell PJ, et al. DNA deaminases
846 induce break-associated mutation showers with implication of APOBEC3B and 3A in
847 breast cancer kataegis. Elife. 2013;2:e00534. Epub 2013/04/20. doi: 10.7554/eLife.00534
848 00534 [pii]. PubMed PMID: 23599896; PubMed Central PMCID: PMC3628087.
849 35. Verhalen B, Starrett GJ, Harris RS, Jiang M. Functional upregulation of the DNA cytosine
850 deaminase APOBEC3B by polyomaviruses. J Virol. 2016;90(14):6379-86. doi:
851 10.1128/JVI.00771-16. PubMed PMID: 27147740.
852 36. Vieira VC, Leonard B, White EA, Starrett GJ, Temiz NA, Lorenz LD, et al. Human
853 papillomavirus E6 triggers upregulation of the antiviral and cancer genomic DNA
854 deaminase APOBEC3B. MBio. 2014;5(6). doi: 10.1128/mBio.02234-14. PubMed PMID:
855 25538195; PubMed Central PMCID: PMC4278539.
856 37. Starrett GJ, Serebrenik AA, Roelofs PA, McCann JL, Verhalen B, Jarvis MC, et al.
857 Polyomavirus T antigen induces APOBEC3B expression using an LXCXE-dependent and

858 TP53-independent mechanism. MBio. 2019;10(1). doi: 10.1128/mBio.02690-18. PubMed
859 PMID: 30723127.
860 38. Walker BA, Wardell CP, Murison A, Boyle EM, Begum DB, Dahir NM, et al. APOBEC
861 family mutational signatures are associated with poor prognosis translocations in multiple
862 myeloma. Nat Commun. 2015;6:6997. doi: 10.1038/ncomms7997. PubMed PMID:
863 25904160; PubMed Central PMCID: PMC4568299.
864 39. Roelofs PA, Goh CY, Chua BH, Jarvis MC, Stewart TA, McCann JL, et al. Characterization
865 of the mechanism by which the RB/E2F pathway controls expression of the cancer genomic
866 DNA deaminase APOBEC3B. Elife. 2020;9. Epub 2020/09/29. doi: 10.7554/eLife.61287.
867 PubMed PMID: 32985974; PubMed Central PMCID: PMCPMC7553775.
868 40. Salamango DJ, McCann JL, Demir O, Brown WL, Amaro RE, Harris RS. APOBEC3B nuclear
869 localization requires two distinct N-terminal domain surfaces. J Mol Biol.
870 2018;430(17):2695-708. doi: 10.1016/j.jmb.2018.04.044. PubMed PMID: 29787764.
871 41. de Bruin EC, McGranahan N, Mitter R, Salm M, Wedge DC, Yates L, et al. Spatial and
872 temporal diversity in genomic instability processes defines lung cancer evolution. Science.
873 2014;346(6206):251-6. doi: 10.1126/science.1253462. PubMed PMID: 25301630.
874 42. Argyris PP, Wilkinson PE, Jarvis MC, Magliocca KR, Patel MR, Vogel RI, et al. Endogenous
875 APOBEC3B overexpression characterizes HPV-positive and HPV-negative oral epithelial
876 dysplasias and head and neck cancers. Modern pathology : an official journal of the United
877 States and Canadian Academy of Pathology, Inc. 2020. Epub 2020/07/08. doi:
878 10.1038/s41379-020-0617-x. PubMed PMID: 32632179.
879 43. Yamazaki H, Shirakawa K, Matsumoto T, Hirabayashi S, Murakawa Y, Kobayashi M, et al.
880 Endogenous APOBEC3B overexpression constitutively generates DNA substitutions and
881 deletions in myeloma cells. Sci Rep. 2019;9(1):7122. Epub 2019/05/11. doi:

882 10.1038/s41598-019-43575-y. PubMed PMID: 31073151; PubMed Central PMCID:
883 PMCPMC6509214.
884 44. Yamazaki H, Shirakawa K, Matsumoto T, Kazuma Y, Matsui H, Horisawa Y, et al.
885 APOBEC3B reporter myeloma cell lines identify DNA damage response pathways leading
886 to APOBEC3B expression. PloS one. 2020;15(1):e0223463. Epub 2020/01/09. doi:
887 10.1371/journal.pone.0223463. PubMed PMID: 31914134; PubMed Central PMCID:
888 PMCPMC6948746.
889 45. Nik-Zainal S, Wedge DC, Alexandrov LB, Petljak M, Butler AP, Bolli N, et al. Association of
890 a germline copy number polymorphism of APOBEC3A and APOBEC3B with burden of
891 putative APOBEC-dependent mutations in breast cancer. Nat Genet. 2014;46(5):487-91.
892 doi: 10.1038/ng.2955. PubMed PMID: 24728294; PubMed Central PMCID:
893 PMC4137149.
894 46. Starrett GJ, Luengas EM, McCann JL, Ebrahimi D, Temiz NA, Love RP, et al. The DNA
895 cytosine deaminase APOBEC3H haplotype I likely contributes to breast and lung cancer
896 mutagenesis. Nat Commun. 2016;7:12918. doi: 10.1038/ncomms12918. PubMed PMID:
897 27650891.
898 47. Chen TW, Lee CC, Liu H, Wu CS, Pickering CR, Huang PJ, et al. APOBEC3A is an oral
899 cancer prognostic biomarker in Taiwanese carriers of an APOBEC deletion polymorphism.
900 Nat Commun. 2017;8(1):465. doi: 10.1038/s41467-017-00493-9. PubMed PMID:
901 28878238; PubMed Central PMCID: PMC5587710.
902 48. Hix MA, Wong L, Flath B, Chelico L, Cisneros GA. Single-nucleotide polymorphism of the
903 DNA cytosine deaminase APOBEC3H haplotype I leads to enzyme destabilization and
904 correlates with lung cancer. NAR Cancer. 2020;2(3):zcaa023. Epub 2020/09/29. doi:
905 10.1093/narcan/zcaa023. PubMed PMID: 32984821; PubMed Central PMCID:

PMCPMC7503452.

49. Feng Y, Love RP, Ara A, Baig TT, Adolph MB, Chelico L. Natural polymorphisms and oligomerization of human APOBEC3H contribute to single-stranded DNA scanning ability. J Biol Chem. 2015;290(45):27188-203. doi: 10.1074/jbc.M115.666065. PubMed PMID: 26396192; PubMed Central PMCID: PMC4646388.

50. Chan K, Roberts SA, Klimczak LJ, Sterling JF, Saini N, Malc EP, et al. An APOBEC3A hypermutation signature is distinguishable from the signature of background mutagenesis by APOBEC3B in human cancers. Nat Genet. 2015;47(9):1067-72. doi: 10.1038/ng.3378. PubMed PMID: 26258849.

51. Hoopes JI, Cortez LM, Mertz TM, Malc EP, Mieczkowski PA, Roberts SA. APOBEC3A and APOBEC3B preferentially deaminate the lagging strand template during DNA replication. Cell Rep. 2016;14(6):1273-82. doi: 10.1016/j.celrep.2016.01.021. PubMed PMID: 26832400; PubMed Central PMCID: PMC4758883.

52. Buisson R, Langenbucher A, Bowen D, Kwan EE, Benes CH, Zou L, et al. Passenger hotspot mutations in cancer driven by APOBEC3A and mesoscale genomic features. Science. 2019;364(6447). Epub 2019/06/30. doi: 10.1126/science.aaw2872. PubMed PMID: 31249028; PubMed Central PMCID: PMCPMC6731024.

53. Petersen-Mahrt SK, Harris RS, Neuberger MS. AID mutates *E. coli* suggesting a DNA deamination mechanism for antibody diversification. Nature. 2002;418(6893):99-103. PubMed PMID: 12097915.

54. Harris RS, Petersen-Mahrt SK, Neuberger MS. RNA editing enzyme APOBEC1 and some of its homologs can act as DNA mutators. Mol Cell. 2002;10(5):1247-53. PubMed PMID: 12453430.

55. Mayorov VI, Rogozin IB, Adkison LR, Frahm C, Kunkel TA, Pavlov YI. Expression of human

AID in yeast induces mutations in context similar to the context of somatic hypermutation at G-C pairs in immunoglobulin genes. BMC Immunol. 2005;6(1):10. PubMed PMID: 15949042.

56. Schumacher AJ, Nissley DV, Harris RS. APOBEC3G hypermutates genomic DNA and inhibits Ty1 retrotransposition in yeast. Proc Natl Acad Sci U S A. 2005;102(28):9854-9. PubMed PMID: 16000409.

57. Chan K, Sterling JF, Roberts SA, Bhagwat AS, Resnick MA, Gordenin DA. Base damage within single-strand DNA underlies in vivo hypermutability induced by a ubiquitous environmental agent. PLoS Genet. 2012;8(12):e1003149. doi: 10.1371/journal.pgen.1003149. PubMed PMID: 23271983; PubMed Central PMCID: PMC3521656.

58. Refsland EW, Hultquist JF, Luengas EM, Ikeda T, Shaban NM, Law EK, et al. Natural polymorphisms in human APOBEC3H and HIV-1 Vif combine in primary T lymphocytes to affect viral G-to-A mutation levels and infectivity. PLoS Genet. 2014;10(11):e1004761. doi: 10.1371/journal.pgen.1004761. PubMed PMID: 25411794; PubMed Central PMCID: PMC4238949.

59. Ebrahimi D, Richards CM, Carpenter MA, Wang J, Ikeda T, Becker JT, et al. Genetic and mechanistic basis for APOBEC3H alternative splicing, retrovirus restriction, and counteraction by HIV-1 protease. Nat Commun. 2018;9(1):4137. doi: 10.1038/s41467-018-06594-3. PubMed PMID: 30297863; PubMed Central PMCID: PMC6175962.

60. Chesarino NM, Emerman M. Polymorphisms in human APOBEC3H differentially regulate ubiquitination and antiviral activity. Viruses. 2020;12(4). Epub 2020/04/03. doi: 10.3390/v12040378. PubMed PMID: 32235597; PubMed Central PMCID: PMCPMC7232234.

954 61. Brown WL, Law EK, Argyris PP, Carpenter MA, Levin-Klein R, Ranum AN, et al. A rabbit
955 monoclonal antibody against the antiviral and cancer genomic DNA mutating enzyme
956 APOBEC3B. Antibodies (Basel). 2019;8(3). Epub 2019/09/24. doi:
957 10.3390/antib8030047. PubMed PMID: 31544853.
958 62. Shaban NM, Shi K, Lauer KV, Carpenter MA, Richards CM, Salamango D, et al. The antiviral
959 and cancer genomic DNA deaminase APOBEC3H is regulated by an RNA-mediated
960 dimerization mechanism. Mol Cell. 2018;69(1):75-86 e9. doi:
961 10.1016/j.molcel.2017.12.010. PubMed PMID: 29290613.
962 63. Lackey L, Demorest ZL, Land AM, Hultquist JF, Brown WL, Harris RS. APOBEC3B and
963 AID have similar nuclear import mechanisms. J Mol Biol. 2012;419(5):301-14. Epub
964 2012/03/27. doi: 10.1016/j.jmb.2012.03.011 S0022-2836(12)00266-5 [pii]. PubMed
965 PMID: 22446380; PubMed Central PMCID: PMC3368237.
966 64. Lackey L, Law EK, Brown WL, Harris RS. Subcellular localization of the APOBEC3 proteins
967 during mitosis and implications for genomic DNA deamination. Cell Cycle.
968 2013;12(5):762-72. Epub 2013/02/08. doi: 10.4161/cc.23713 [pii]. PubMed PMID:
969 23388464; PubMed Central PMCID: PMC3610724.
970 65. Akre MK, Starrett GJ, Quist JS, Temiz NA, Carpenter MA, Tutt AN, et al. Mutation processes
971 in 293-based clones overexpressing the DNA cytosine deaminase APOBEC3B. PloS one.
972 2016;11(5):e0155391. doi: 10.1371/journal.pone.0155391. PubMed PMID: 27163364;
973 PubMed Central PMCID: PMC4862684.
974 66. Landry S, Narvaiza I, Linfesty DC, Weitzman MD. APOBEC3A can activate the DNA damage
975 response and cause cell-cycle arrest. EMBO Rep. 2011;12(5):444-50. Epub 2011/04/05.
976 doi: embor201146 [pii] 10.1038/embor.2011.46. PubMed PMID: 21460793.
977 67. Nikkila J, Kumar R, Campbell J, Brandsma I, Pemberton HN, Wallberg F, et al. Elevated

APOBEC3B expression drives a kataegic-like mutation signature and replication stress-related therapeutic vulnerabilities in p53-defective cells. British journal of cancer. 2017;117(1):113-23. doi: 10.1038/bjc.2017.133. PubMed PMID: 28535155; PubMed Central PMCID: PMC5520199.

68. Mussil B, Suspene R, Aynaud MM, Gauvrit A, Vartanian JP, Wain-Hobson S. Human APOBEC3A isoforms translocate to the nucleus and induce DNA double strand breaks leading to cell stress and death. PloS one. 2013;8(8):e73641. doi: 10.1371/journal.pone.0073641. PubMed PMID: 23977391; PubMed Central PMCID: PMC3748023.

69. Jarvis MC, Ebrahimi D, Temiz NA, Harris RS. Mutation signatures including APOBEC in cancer cell lines. JNCI Cancer Spectr. 2018;2(1). doi: 10.1093/jncics/pky002. PubMed PMID: 29888758; PubMed Central PMCID: PMC5993214.

70. Alexandrov LB, Nik-Zainal S, Wedge DC, Campbell PJ, Stratton MR. Deciphering signatures of mutational processes operative in human cancer. Cell Rep. 2013;3(1):246-59. Epub 2013/01/16. doi: 10.1016/j.celrep.2012.12.008. PubMed PMID: 23318258; PubMed Central PMCID: PMCPMC3588146.

71. Law EK, Levin-Klein R, Jarvis MC, Kim H, Argyris PP, Carpenter MA, et al. APOBEC3A catalyzes mutation and drives carcinogenesis in vivo. J Exp Med. 2020;217(12). Epub 2020/09/02. doi: 10.1084/jem.20200261. PubMed PMID: 32870257.

72. Jalili P, Bowen D, Langenbucher A, Park S, Aguirre K, Corcoran RB, et al. Quantification of ongoing APOBEC3A activity in tumor cells by monitoring RNA editing at hotspots. Nat Commun. 2020;11(1):2971. Epub 2020/06/14. doi: 10.1038/s41467-020-16802-8. PubMed PMID: 32532990; PubMed Central PMCID: PMCPMC7293259.

73. Langenbucher A, Bowen D, Sakhtemani R, Bournique E, Wise JF, Zou L, et al. An extended

APOBEC3A mutation signature in cancer. Nat Commun. 2021;12(1):1602. Epub 2021/03/13. doi: 10.1038/s41467-021-21891-0. PubMed PMID: 33707442; PubMed Central PMCID: PMCPMC7952602.

74. Sharma S, Patnaik SK, Taggart RT, Kannisto ED, Enriquez SM, Gollnick P, et al. APOBEC3A cytidine deaminase induces RNA editing in monocytes and macrophages. Nat Commun. 2015;6:6881. doi: 10.1038/ncomms7881. PubMed PMID: 25898173; PubMed Central PMCID: PMC4411297.

75. Roberts SA, Sterling J, Thompson C, Harris S, Mav D, Shah R, et al. Clustered mutations in yeast and in human cancers can arise from damaged long single-strand DNA regions. Mol Cell. 2012;46(4):424-35. doi: 10.1016/j.molcel.2012.03.030. PubMed PMID: 22607975; PubMed Central PMCID: PMC3361558.

76. Henderson S, Chakravarthy A, Su X, Boshoff C, Fenton TR. APOBEC-mediated cytosine deamination links PIK3CA helical domain mutations to human papillomavirus-driven tumor development. Cell Rep. 2014;7(6):1833-41. doi: 10.1016/j.celrep.2014.05.012. PubMed PMID: 24910434.

77. Zou X, Koh GCC, Nanda AS, Degasperi A, Urgo K, Roumeliotis TI, et al. A systematic CRISPR screen defines mutational mechanisms underpinning signatures caused by replication errors and endogenous DNA damage. Nat Cancer. 2021;2(6):643-57. Epub 2021/06/25. doi: 10.1038/s43018-021-00200-0. PubMed PMID: 34164627; PubMed Central PMCID: PMCPMC7611045.

78. Zou X, Owusu M, Harris R, Jackson SP, Loizou JI, Nik-Zainal S. Validating the concept of mutational signatures with isogenic cell models. Nat Commun. 2018;9(1):1744. Epub 2018/05/03. doi: 10.1038/s41467-018-04052-8. PubMed PMID: 29717121; PubMed Central PMCID: PMCPMC5931590.

79. DeWeerd RA, Nemeth E, Poti A, Petryk N, Chen CL, Hyrien O, et al. Prospectively defined patterns of APOBEC3A mutagenesis are prevalent in human cancers. Cell Rep. 2022;38(12):110555. Epub 2022/03/24. doi: 10.1016/j.celrep.2022.110555. PubMed PMID: 35320711.

80. Warren CJ, Westrich JA, Doorslaer KV, Pyeon D. Roles of APOBEC3A and APOBEC3B in human papillomavirus infection and disease progression. Viruses. 2017;9(8). doi: 10.3390/v9080233. PubMed PMID: 28825669; PubMed Central PMCID: PMC5580490.

81. Leonard B, McCann JL, Starrett GJ, Kosyakovsky L, Luengas EM, Molan AM, et al. The PKC/NF-kappaB signaling pathway induces APOBEC3B expression in multiple human cancers. Cancer Res. 2015;75(21):4538-47. doi: 10.1158/0008-5472.CAN-15-2171-T. PubMed PMID: 26420215; PubMed Central PMCID: PMC4631676.

82. Periyasamy M, Patel H, Lai CF, Nguyen VT, Nevedomskaya E, Harrod A, et al. APOBEC3B-mediated cytidine deamination is required for estrogen receptor action in breast cancer. Cell Rep. 2015;13(1):108-21. doi: 10.1016/j.celrep.2015.08.066. PubMed PMID: 26411678; PubMed Central PMCID: PMC4597099.

83. Periyasamy M, Singh AK, Gemma C, Kranjec C, Farzan R, Leach DA, et al. p53 controls expression of the DNA deaminase APOBEC3B to limit its potential mutagenic activity in cancer cells. Nucleic Acids Res. 2017;45(19):11056-69. doi: 10.1093/nar/gkx721. PubMed PMID: 28977491; PubMed Central PMCID: PMC5737468.

84. Cho RJ, Alexandrov LB, den Breems NY, Atanasova VS, Farshchian M, Purdom E, et al. APOBEC mutation drives early-onset squamous cell carcinomas in recessive dystrophic epidermolysis bullosa. Sci Transl Med. 2018;10(455). doi: 10.1126/scitranslmed.aas9668. PubMed PMID: 30135250.

85. Oh S, Bournique E, Bowen D, Jalili P, Sanchez A, Ward I, et al. Genotoxic stress and viral

infection induce transient expression of APOBEC3A and pro-inflammatory genes through two distinct pathways. Nat Commun. 2021;12(1):4917. Epub 2021/08/15. doi: 10.1038/s41467-021-25203-4. PubMed PMID: 34389714; PubMed Central PMCID: PMCPMC8363607.

86. Maruyama W, Shirakawa K, Matsui H, Matsumoto T, Yamazaki H, Sarca AD, et al. Classical NF-kappaB pathway is responsible for APOBEC3B expression in cancer cells. Biochem Biophys Res Commun. 2016;478(3):1466-71. doi: 10.1016/j.bbrc.2016.08.148. PubMed PMID: 27577680.

87. Petljak M, Alexandrov LB, Brammeld JS, Price S, Wedge DC, Grossmann S, et al. Characterizing mutational signatures in human cancer cell lines reveals episodic APOBEC mutagenesis. Cell. 2019;176(6):1282-94 e20. Epub 2019/03/09. doi: 10.1016/j.cell.2019.02.012. PubMed PMID: 30849372; PubMed Central PMCID: PMCPMC6424819.

88. Petljak M, Dananberg A, Chu K, Bergstrom EN, Striepen J, von Morgen P, et al. Mechanisms of APOBEC3 mutagenesis in human cancer cells. Nature. 2022;607(7920):799-807. Epub 20220720. doi: 10.1038/s41586-022-04972-y. PubMed PMID: 35859169; PubMed Central PMCID: PMCPMC9329121.

89. Law EK, Sieuwerts AM, LaPara K, Leonard B, Starrett GJ, Molan AM, et al. The DNA cytosine deaminase APOBEC3B promotes tamoxifen resistance in ER-positive breast cancer. Sci Adv. 2016;2(10):e1601737. doi: 10.1126/sciadv.1601737. PubMed PMID: 27730215.

90. Chen Z, Wen W, Bao J, Kuhs KL, Cai Q, Long J, et al. Integrative genomic analyses of APOBEC-mutational signature, expression and germline deletion of APOBEC3 genes, and immunogenicity in multiple cancer types. BMC medical genomics. 2019;12(1):131. Epub

2019/09/20. doi: 10.1186/s12920-019-0579-3. PubMed PMID: 31533728; PubMed Central PMCID: PMCPMC6751822.

91. Glaser AP, Fantini D, Wang Y, Yu Y, Rimar KJ, Podojil JR, et al. APOBEC-mediated mutagenesis in urothelial carcinoma is associated with improved survival, mutations in DNA damage response genes, and immune response. Oncotarget. 2018;9(4):4537-48. Epub 2018/02/13. doi: 10.18632/oncotarget.23344. PubMed PMID: 29435122; PubMed Central PMCID: PMCPMC5796993.

92. Middlebrooks CD, Banday AR, Matsuda K, Udquim KI, Onabajo OO, Paquin A, et al. Association of germline variants in the APOBEC3 region with cancer risk and enrichment with APOBEC-signature mutations in tumors. Nat Genet. 2016;48(11):1330-8. Epub 2016/10/28. doi: 10.1038/ng.3670. PubMed PMID: 27643540; PubMed Central PMCID: PMCPMC6583788.

93. Cannataro VL, Gaffney SG, Sasaki T, Issaeva N, Grewal NKS, Grandis JR, et al. APOBEC-induced mutations and their cancer effect size in head and neck squamous cell carcinoma. Oncogene. 2019;38(18):3475-87. Epub 2019/01/17. doi: 10.1038/s41388-018-0657-6. PubMed PMID: 30647454; PubMed Central PMCID: PMCPMC6499643.

94. Driscoll CB, Schuelke MR, Kottke T, Thompson JM, Wongthida P, Tonne JM, et al. APOBEC3B-mediated corruption of the tumor cell immunopeptidome induces heteroclitic neoepitopes for cancer immunotherapy. Nat Commun. 2020;11(1):790. Epub 2020/02/09. doi: 10.1038/s41467-020-14568-7. PubMed PMID: 32034147; PubMed Central PMCID: PMCPMC7005822.

95. Tsuboi M, Yamane A, Horiguchi J, Yokobori T, Kawabata-Iwakawa R, Yoshiyama S, et al. APOBEC3B high expression status is associated with aggressive phenotype in Japanese breast cancers. Breast cancer. 2016;23(5):780-8. Epub 2015/10/20. doi: 10.1007/s12282-

1098 015-0641-8. PubMed PMID: 26476745.

1099 96. Uphoff CC, Drexler HG. Detecting mycoplasma contamination in cell cultures by polymerase
1100 chain reaction. Methods Mol Biol. 2011;731:93-103. Epub 2011/04/26. doi: 10.1007/978-
1101 1-61779-080-5_8. PubMed PMID: 21516400.

1102 97. Stenglein MD, Burns MB, Li M, Lengyel J, Harris RS. APOBEC3 proteins mediate the
1103 clearance of foreign DNA from human cells. Nat Struct Mol Biol. 2010;17(2):222-9.
1104 PubMed PMID: 20062055.

1105 98. Refsland EW, Stenglein MD, Shindo K, Albin JS, Brown WL, Harris RS. Quantitative
1106 profiling of the full *APOBEC3* mRNA repertoire in lymphocytes and tissues: implications
1107 for HIV-1 restriction. Nucleic Acids Res. 2010;38:4274-84. PubMed PMID: 20308164.

1108 99. Wang J, Becker JT, Shi K, Lauer KV, Salamango DJ, Aihara H, et al. The role of RNA in
1109 HIV-1 Vif-mediated degradation of APOBEC3H. J Mol Biol. 2019;431(24):5019-31.
1110 Epub 2019/10/20. doi: 10.1016/j.jmb.2019.09.014. PubMed PMID: 31628948; PubMed
1111 Central PMCID: PMCPMC6948013.

1112 100. Shi K, Carpenter MA, Kurahashi K, Harris RS, Aihara H. Crystal structure of the DNA
1113 deaminase APOBEC3B catalytic domain. J Biol Chem. 2015;290(47):28120-30. doi:
1114 10.1074/jbc.M115.679951. PubMed PMID: 26416889; PubMed Central PMCID:
1115 PMC4653671.

1116 101. Shaban NM, Yan R, Shi K, Moraes SN, Cheng AZ, Carpenter MA, et al. Cryo-EM
1117 structure of the EBV ribonucleotide reductase BORF2 and mechanism of APOBEC3B
1118 inhibition. Science Advances. 2022;In press.

1119 102. McCann JL, Cristini A, Law EK, Lee SY, Tellier M, Carpenter MA, et al. R-loop
1120 homeostasis and cancer mutagenesis promoted by the DNA cytosine deaminase
1121 APOBEC3B. BioRxiv. 2021.

1122 103. McNamara RP, McCann JL, Gudipaty SA, D'Orso I. Transcription factors mediate the
1123 enzymatic disassembly of promoter-bound 7SK snRNP to locally recruit P-TEFb for
1124 transcription elongation. Cell Rep. 2013;5(5):1256-68. Epub 2013/12/10. doi:
1125 10.1016/j.celrep.2013.11.003. PubMed PMID: 24316072; PubMed Central PMCID:
1126 PMCPMC3882317.
1127 104. Gyori BM, Venkatachalam G, Thiagarajan PS, Hsu D, Clement MV. OpenComet: an
1128 automated tool for comet assay image analysis. Redox Biol. 2014;2:457-65. Epub
1129 2014/03/14. doi: 10.1016/j.redox.2013.12.020. PubMed PMID: 24624335; PubMed
1130 Central PMCID: PMCPMC3949099.
1131 105. Chiang C, Layer RM, Faust GG, Lindberg MR, Rose DB, Garrison EP, et al. SpeedSeq:
1132 ultra-fast personal genome analysis and interpretation. Nat Methods. 2015;12(10):966-8.
1133 Epub 2015/08/11. doi: 10.1038/nmeth.3505. PubMed PMID: 26258291; PubMed Central
1134 PMCID: PMCPMC4589466.
1135 106. Blokzijl F, Janssen R, van Boxtel R, Cuppen E. MutationalPatterns: comprehensive
1136 genome-wide analysis of mutational processes. Genome medicine. 2018;10(1):33. Epub
1137 2018/04/27. doi: 10.1186/s13073-018-0539-0. PubMed PMID: 29695279; PubMed
1138 Central PMCID: PMCPMC5922316.
1139 107. Rosenthal R, McGranahan N, Herrero J, Taylor BS, Swanton C. DeconstructSigs:
1140 delineating mutational processes in single tumors distinguishes DNA repair deficiencies
1141 and patterns of carcinoma evolution. Genome Biol. 2016;17:31. doi: 10.1186/s13059-016-
1142 0893-4. PubMed PMID: 26899170; PubMed Central PMCID: PMC4762164.
1143 108. Cameron DL, Schroder J, Penington JS, Do H, Molania R, Dobrovic A, et al. GRIDSS:
1144 sensitive and specific genomic rearrangement detection using positional de Bruijn graph
1145 assembly. Genome Res. 2017;27(12):2050-60. Epub 2017/11/04. doi:

1146 10.1101/gr.222109.117. PubMed PMID: 29097403; PubMed Central PMCID:

1147 PMCPMC5741059.

1148

Main and Supplementary Figure Legends

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

(**A**) Schematic of the construction of the HAP1-TK-M9 system and overall experimental workflow (see **Figure 1 – Figure Supplement 1** for representative HAP1-TK-M9 karyotype images).

(**B**) *A3* mRNA levels in parental HAP1 cells in comparison to the HAP1-TK-M9 daughter clone by RT-qPCR (*** = $p < 0.001$, NS = non-significant, Welch's t-test).

(**C-E**) Immunoblots of A3A, A3B, and A3H following transduction of HAP1-TK-M9 cells. Tubulin (Tub) is a loading control (see **Figure 1 – Figure Supplement 2** for anti-A3A mAb validation). The lower images show ssDNA deaminase activity of extracts from the same cells (S, substrate; P, product).

(**F**) Immunofluorescent microscopy images of HAP1-TK-M9 cells expressing the indicated A3 enzymes (see methods for primary antibodies). A3 staining is green, γ-H2AX staining is red, and nuclei are blue from Hoechst. A3A is cell wide, A3B is predominantly nuclear, and A3H-I and -II are cell wide with nucleolar concentrations (scale bar = 25 μm). The lower row shows representative COMET images visualized with SYBR Gold (1:10,000) in the GFP channel.

(**G-H**) Quantification of γ-H2AX staining and COMET moments (each data point represents an independent cell; p-values determined with Welch's t-test).

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

Representative G-band and spectral karyotype (SKY) images of HAP1-TK-M9 M-phase chromosomes showing a near diploid DNA content and previously reported aberrations including the reciprocal chromosome 9:22 translocation (Philadelphia chromosome) characteristic of CML tumor cells. The Y-chromosome is missing, as reported previously for the KBM7 parent line of HAP1.

Figure 1 – Supplementary Figure 2. Validation of a custom rabbit-anti human APOBEC3A mAb UMN-13.

(**A**) Schematic of human A3A, A3B, and A3G indicating the unique N-terminal epitope used here to generate an A3A-specific mAb UMN-13. The schematic also shows the C-terminal epitope used previously to generate the versatile 5210-87-13 mAb that recognizes these three enzymes.

(**B**) Comparative immunoblots of whole cell extracts from 293T cells expressing each of the 7 human A3 family members and probed with either our custom rabbit anti-human A3A mAb UMN-13 (left) or a commercial anti-HA mAb as an expression control (right). The positions of the full-length proteins are indicated by red asterisks.

(**C**) Comparative immunoblots of whole cell extracts from the monocytic cell line THP-1 and a clonal derivative lacking A3A-through-A3G (ΔA-G) plus/minus treatment with LPS/IFN-α to induce expression of multiple A3s including A3A and A3G. The UMN-13 mAb blot on the left shows a single band representing full-length A3A (starting at Met1), and the 5210-87-13 mAb blot on the right shows A3G (strong top band), A3B (weak band just below A3G), and both A3A translation products (strong band for full-length A3A starting at Met1 and a faster-migrating band for the shorter isoform starting at Met13).

(**D**) IF microscopy images of 293T cells expressing A3A-mCherry, A3B-mCherry, or A3G-mCherry. Only the A3A construct is detected by the UMN-13 mAb as indicated by green signal in the same cells and cellular compartments as the red signal.

Figure 2. Characterization of *TK* mutations in ganciclovir-resistant clones.

(**A**) A dot plot of Gan^R colonies generated under the indicated A3 expression conditions. Each data point represents the number of Gan^R mutants in a single clonal culture (mean +/- SD shown with significance assessed using Welch's t-test).

(**B**) Schematics representing all *TK* mutations observed under the indicated A3 expression conditions (APOBEC3 signature T(C>T/G)W mutations in red, other SBSs in black, and InDels in

blue; see **Figure 2 – Figure Supplement 1** for schematics of individual *TK* mutants for these and vector control conditions). Q8X and R212K mutation hotspots are labeled.

(**C-D**) Pentanucleotide logos depicting -2 and +2 sequence preferences flanking all T(C>T/G)W mutations observed under the indicated conditions.

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

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

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

(**A**) Trinucleotide profiles of pooled SBSs across all clones sequenced for each listed experimental condition (A3A WT, n=6; A3A E72A, n=2; A3B WT, n=5; A3B E255A, n=2. APOBEC3 signature T(C>T/G)W mutations are highlighted by red dashed-line boxes. See **Figure 3 – Figure Supplement 1 and 2** for mRNA and protein level expression confirmation, respectively, and **Figure 3 – Figure Supplement 4** for SBS profiles from each WGS).

(**B**) Pentanucleotide logos depicting -2, +1 and +2 sequence preferences flanking all C-to-T and C-to-G mutated TC motifs in WGS from HAP1-TK-M9 cells expressing A3A or A3B in comparison to aggregate controls (catalytic mutants, A3H-I, and GFP only conditions, which do not show evidence for APOBEC3 signature mutations; **Figure 3 – Figure Supplement 4**). The genome-wide distribution of nucleobases flanking TC is also shown for comparison.

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

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

(**B**) A heatmap depicting mean expression levels of all 7 human *APOBEC3* family members, in addition to *APOBEC1*, *APOBEC2*, *APOBEC3*, and *AICDA*, relative to those of the housekeeping gene *TBP* (RNA-seq values are FKPM; n≥2 for each condition to provide matching data sets for the Gan^R clones subjected to WGS). Endogenous *A3C* provides a robust internal control.

Figure 3 – Supplementary Figure 2. A3 protein expression in Gan^R clones.

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

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

Figure 3 – Supplementary Figure 3. Numbers of single base substitution mutations in individual ganciclovir-resistant clones by whole genome sequencing.

(**A-B**) Dot plots showing total numbers of cytosine mutations in NCN and TCW motifs, respectively, in WGS from individual Gan^R granddaughter clones (P-values using Welch's T-test).

Figure 3 – Supplementary Figure 4. Single base substitution mutation profiles of individual ganciclovir-resistant clones by whole genome sequencing.

Trinucleotide profiles of the single base substitution mutations found in WGSs from the indicated Gan^R clones (clone names and total SBS numbers are indicated to the right in each profile). Composite profiles for A3A, A3B, and corresponding catalytic mutants are shown in **Figure 3**.

Figure 3 – Supplementary Figure 5. APOBEC3 mutation signature extracted from WGSs using complementary methods.

(**A**) Mutation signature profiles extracted from Gan^{R} clone WGS using an NMF-based approach. For each clone, all SBSs resolved into 2 signatures - an APOBEC-like signature shown in red and a background signature in blue. APOBEC3 mutation enrichment scores are shown above each bar, with significantly enriched values shown in red (BH FDR corrected q-value < 0.05). The dashed line represents the average level of APOBEC3 signature mutations observed in the two eGFP control clones (*i.e.*, background signal).

(**B**) Trinucleotide profiles of the two signatures derived from NMF in panel A clearly showing APOBEC3 signature T(C>T/G)W mutations.

Figure 3 – Supplementary Figure 6. Indel landscape of sequenced granddaughter clones.

Histograms showing indel types and proportions in WGS from the indicated HAP1 granddaughter clones.

Figure 4. Mesoscale properties of A3A and A3B *in vitro* and in the HAP1-TK-M9 system.

(**A-B**) DNA deamination kinetics of A3A and A3B using *SDHB* and *NUP93* hairpin substrates in comparison to corresponding linear controls made by scrambling the 5' or 3' portion of the stem, respectively. See text for full description and **Figure 4 – Figure Supplement 1** for a genome-wide analysis.

(**C**) Rainfall plots of genome-wide intermutation distances showing APOBEC3 signature kataegic tracts (red arrows) in representative A3A- and A3B-expressing Gan^{R} clones (C>T mutations are red, C>G black, other SBS gray).

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

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

base substitution mutations found in predicted hairpin substrates (red) versus non-hairpin contexts (grey). The total number of APOBEC3 signature T(C>T/G)W or other SBS mutations is shown in the center of each pie chart. Hairpin stems and loops were defined liberally as any feature between 3 and 100 nucleotides, and a hairpin is considered mutated if it has a SBS mutation in any position of the loop region.

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

An unsupervised clustering analysis of similarity between the pentanucleotide SBS profiles from WGSs of the A3A, A3B, and control Gan^{R} clones described here versus primary breast tumor whole-genome sequencing data (ICGC, n=794). The APOBEC3 signature is represented by both enrichment score and SBS2+13 (red), HRD signature as SBS3 (blue), and ageing signature as SBS1 (gray).

Figure 1

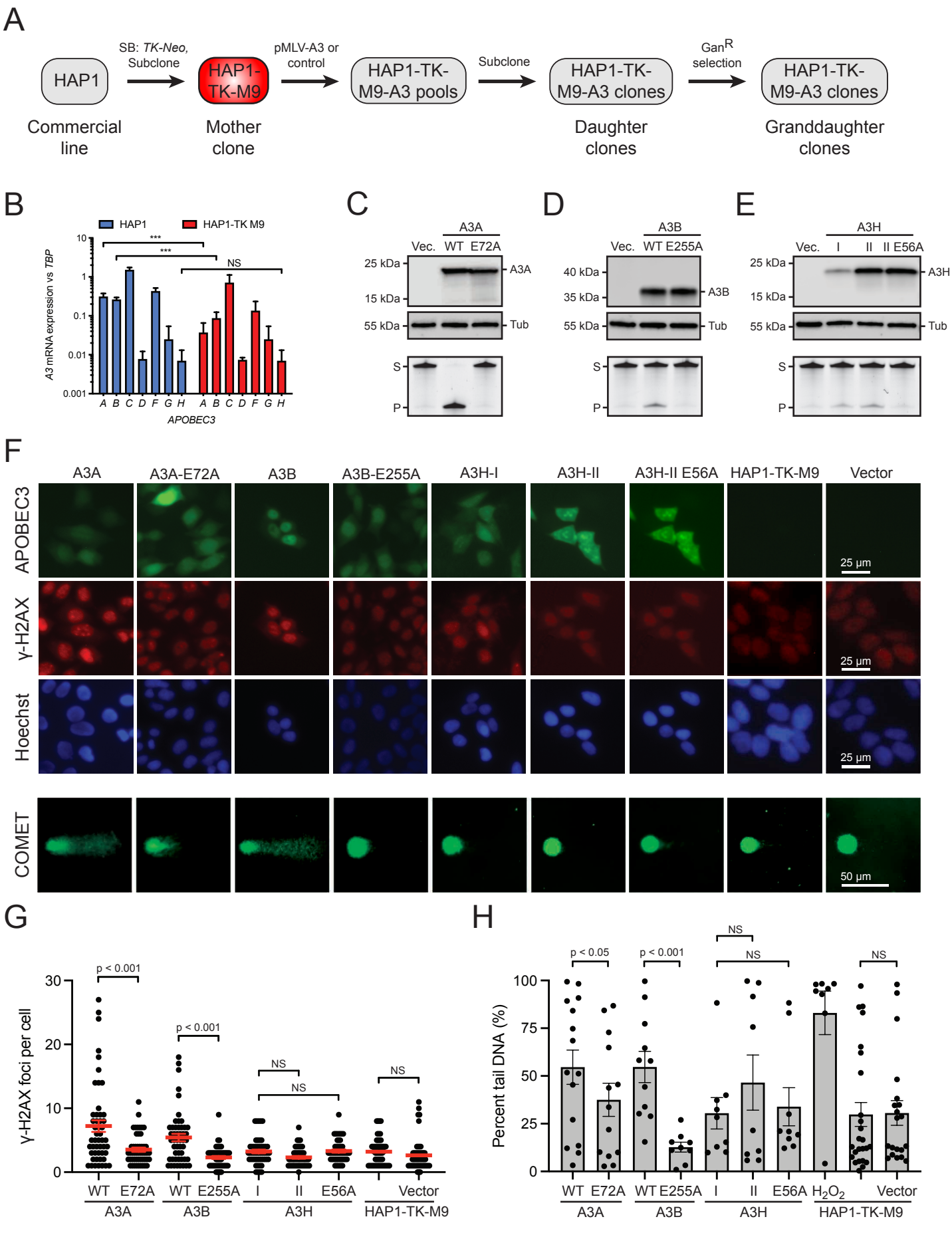


Figure 1 - Supp 1

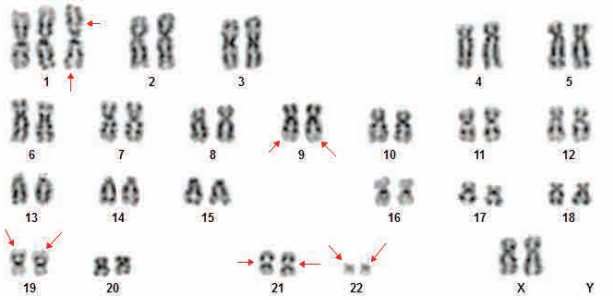


Figure 1 - Supp 2

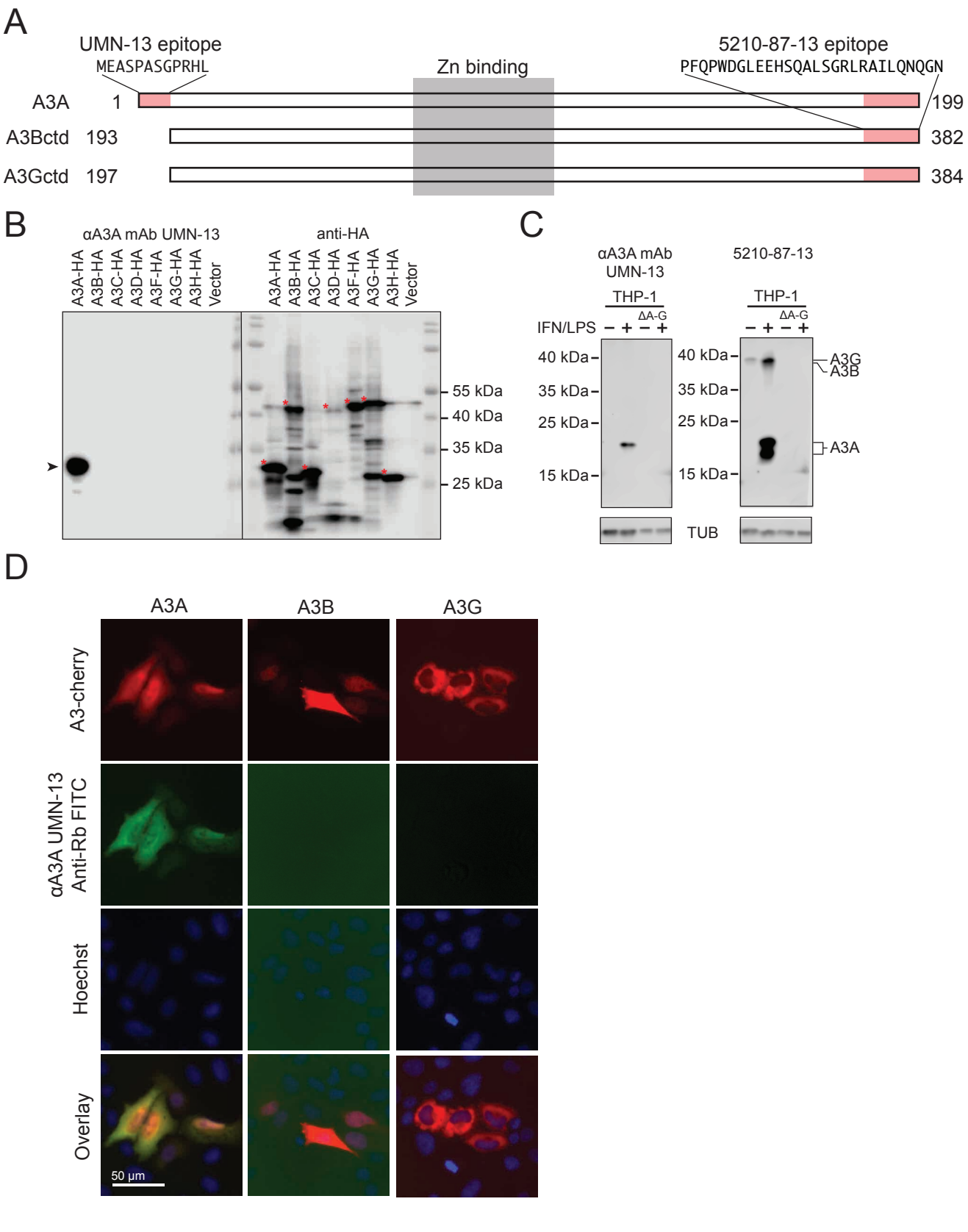


Figure 2

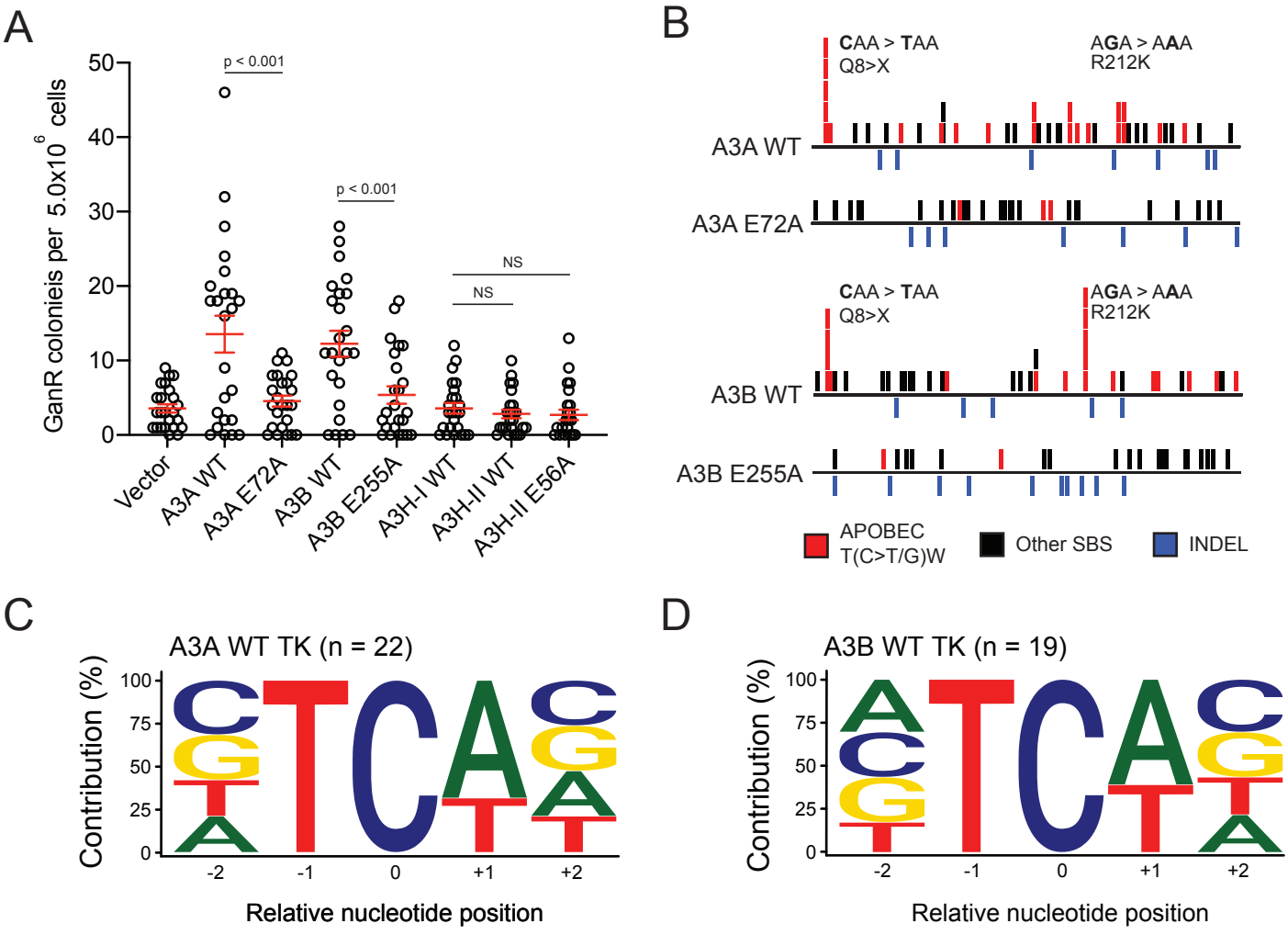


Figure 2 - Supp 1

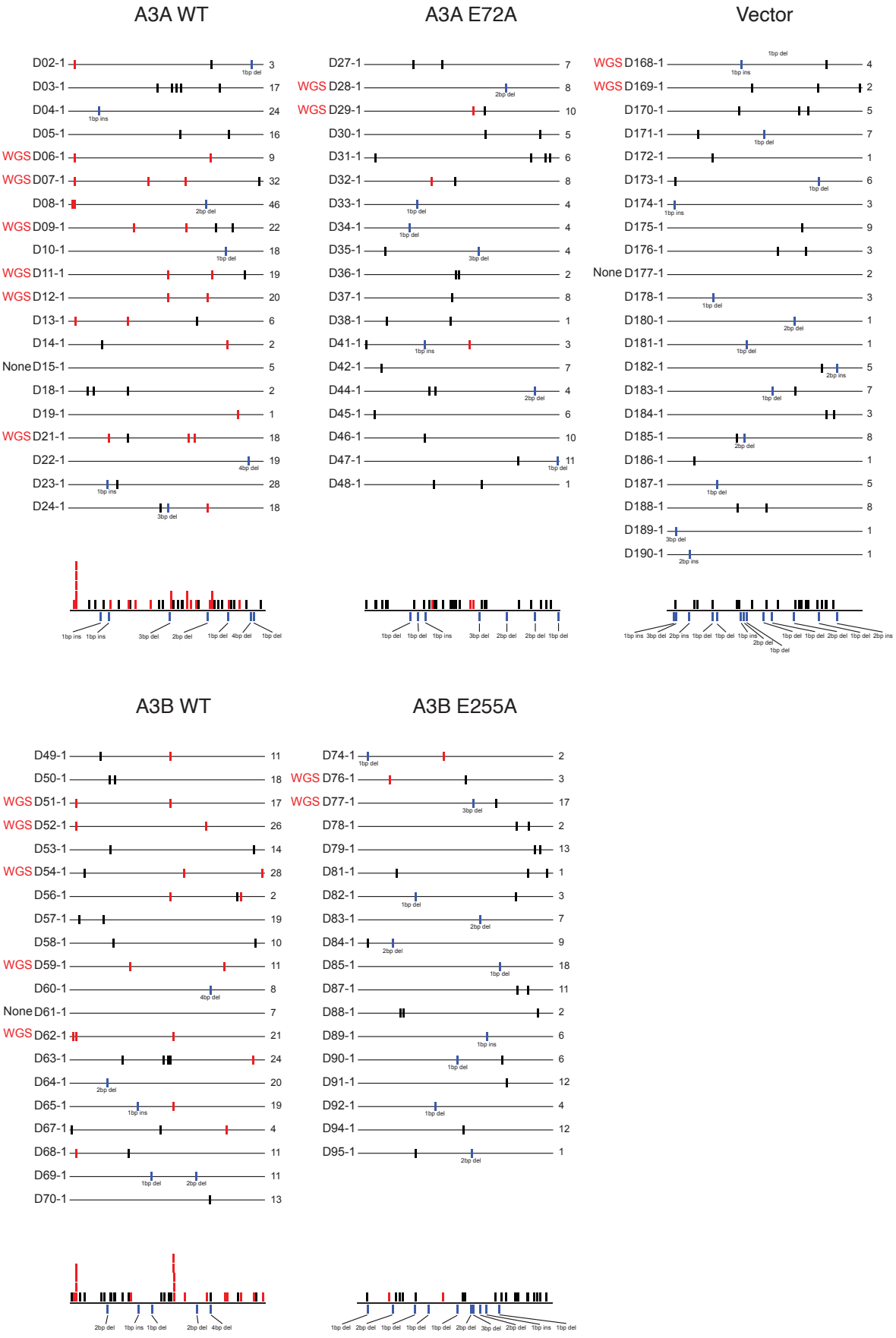
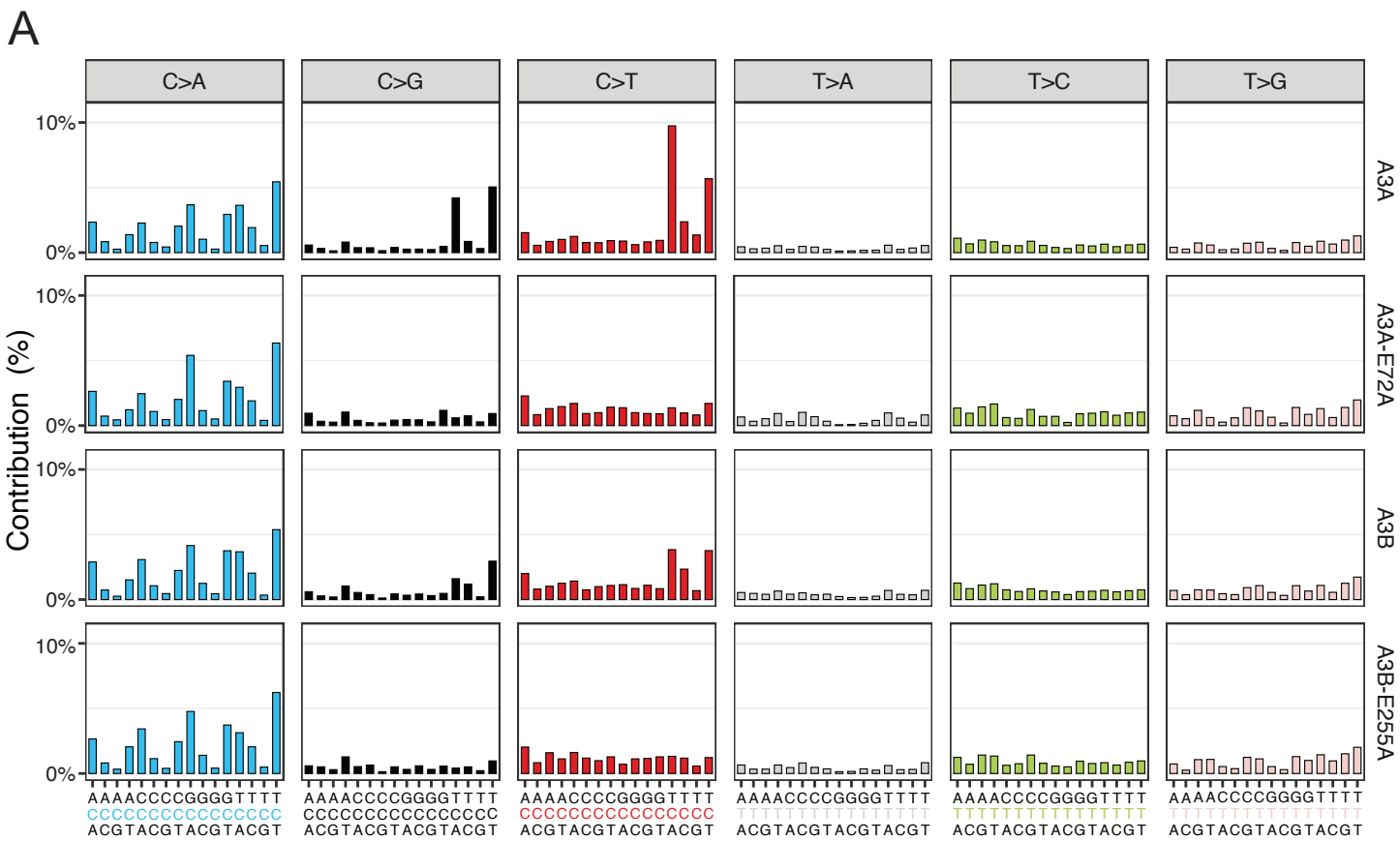


Figure 3



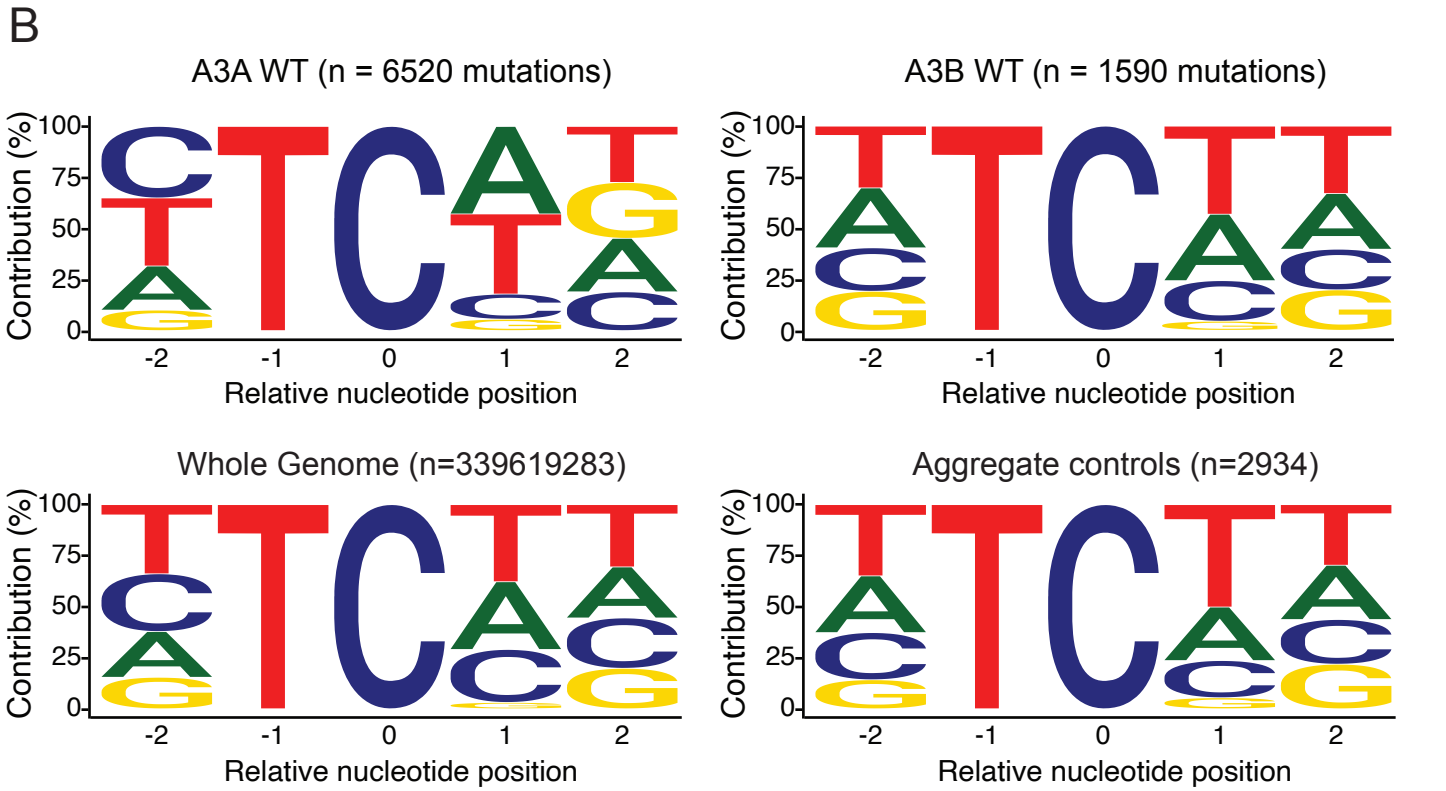
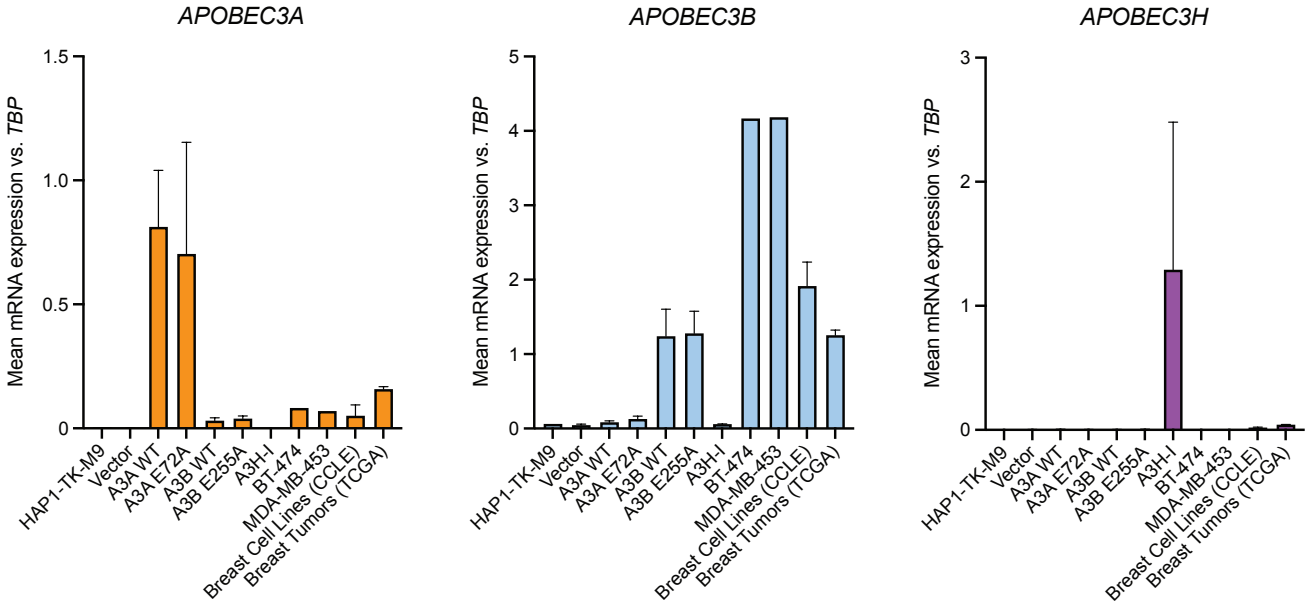


Figure 3 - Supp 1

A



B

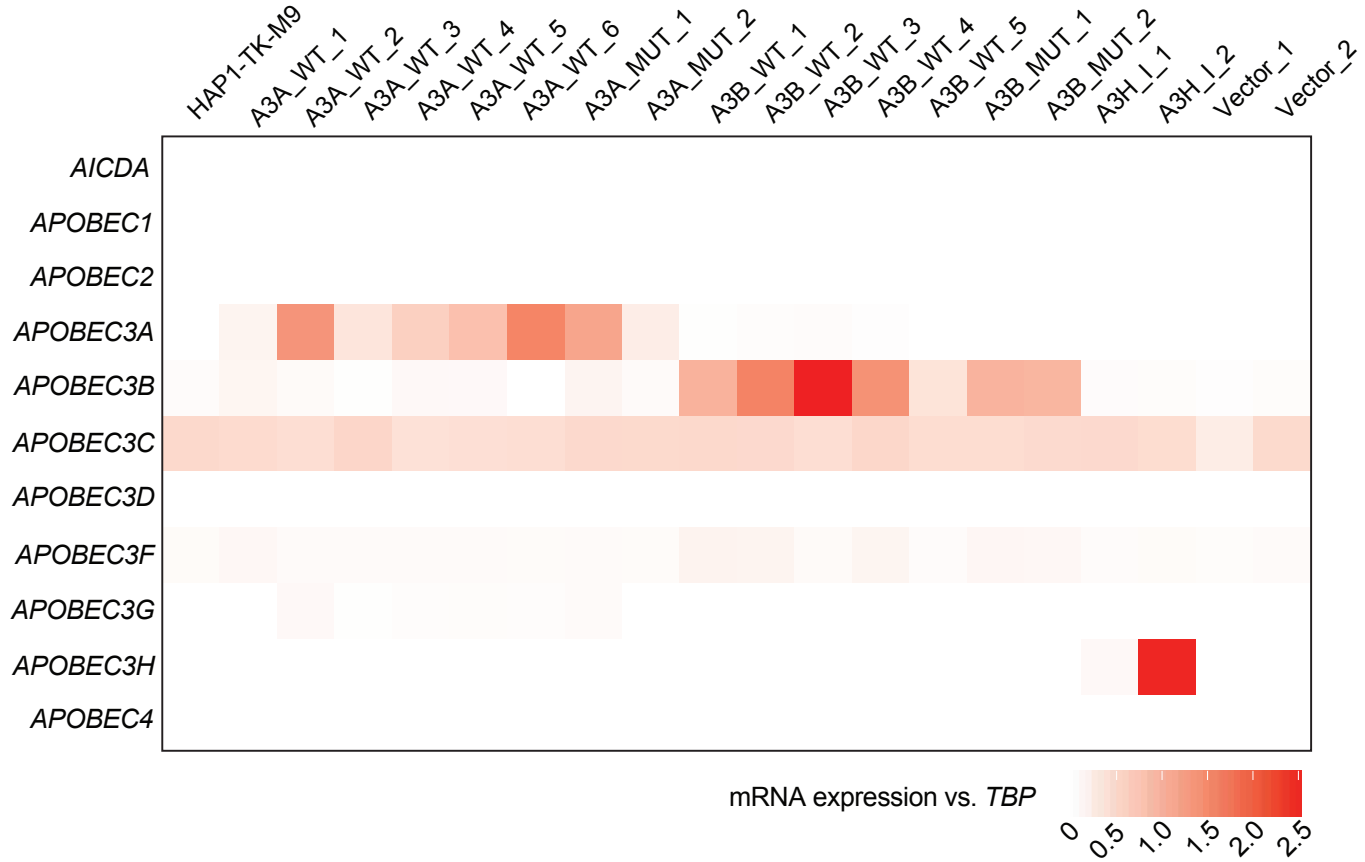


Figure 3 - Supp 2

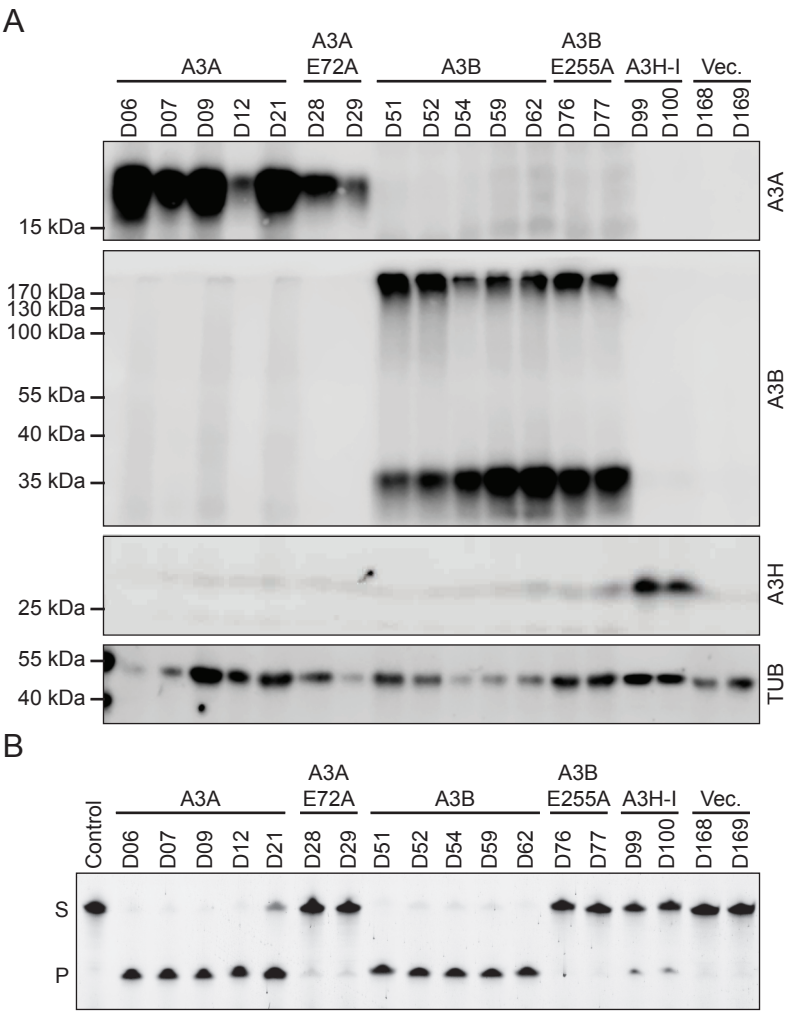
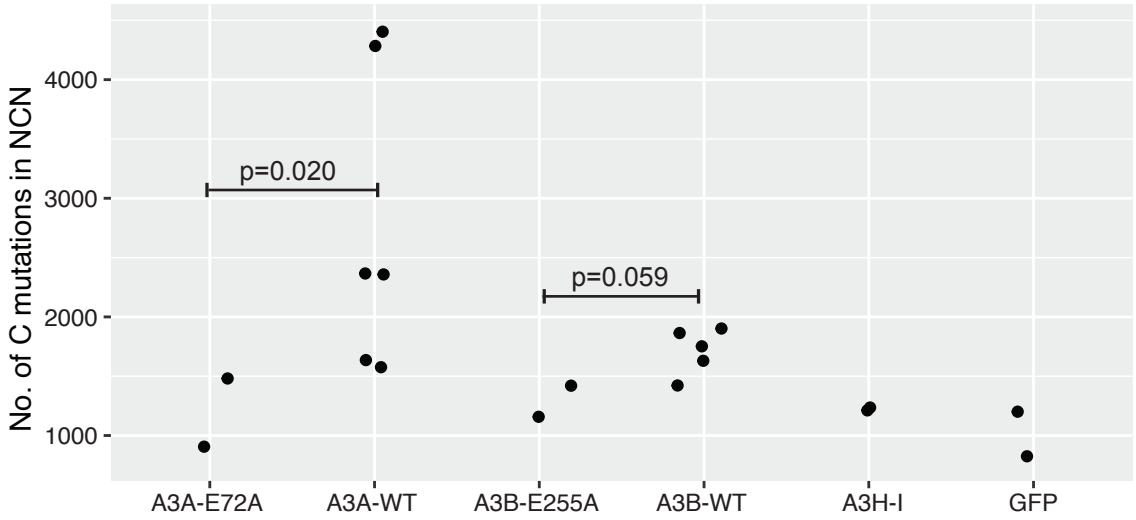


Figure 3 Supp 3

A



B

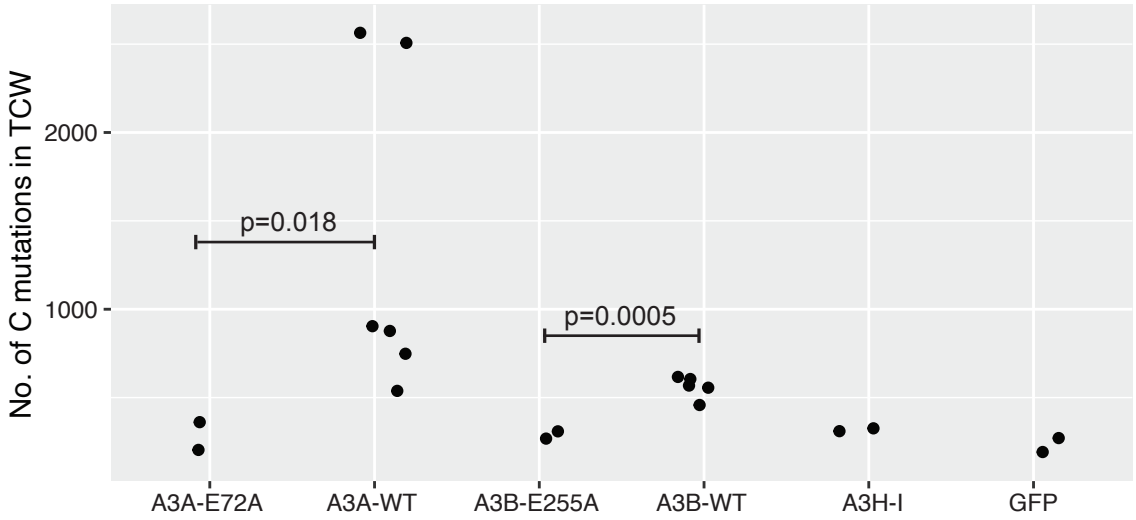


Figure 3 - Supp 4

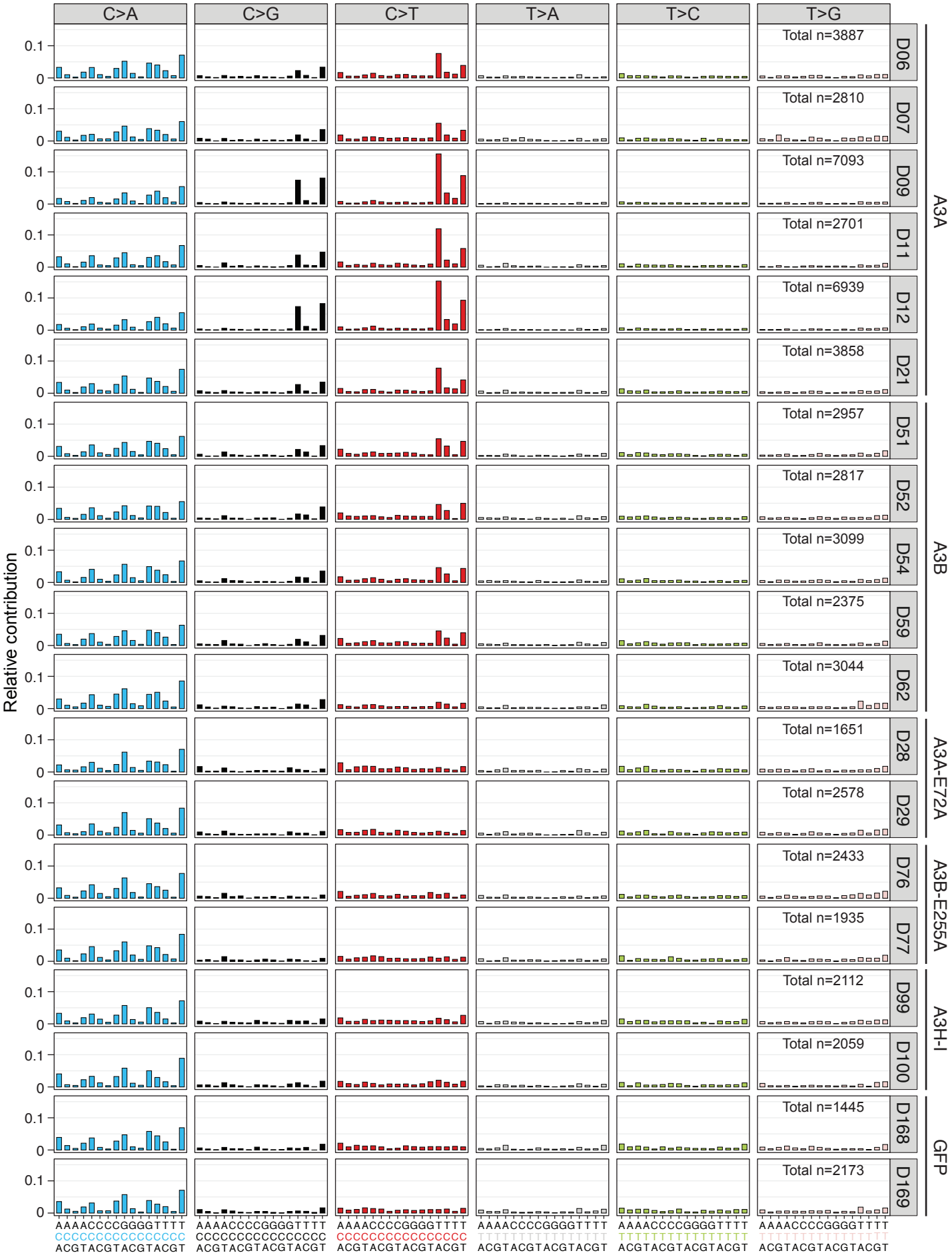


Figure 3 - Supp 5

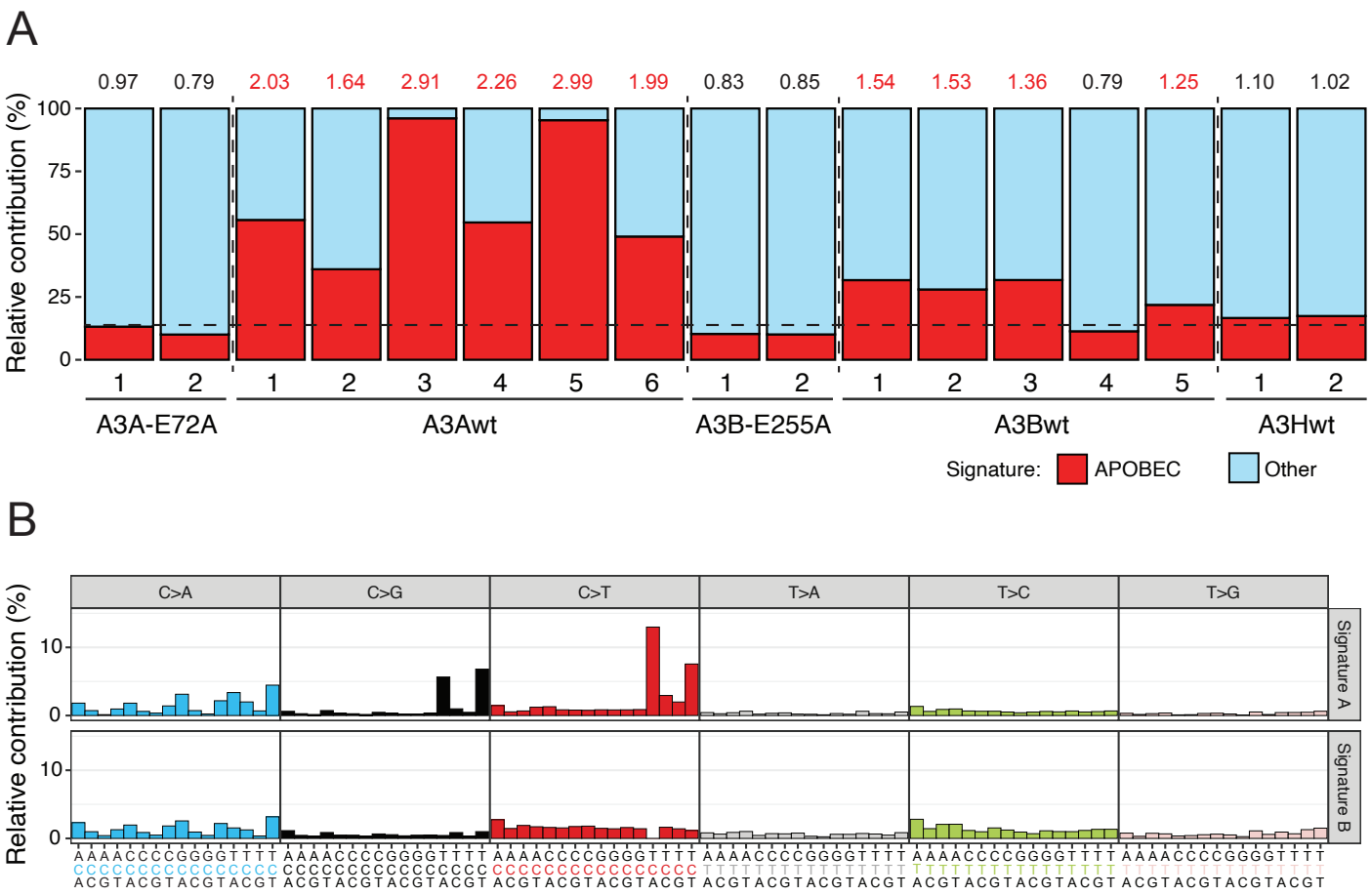


Figure 3 Supp. 6

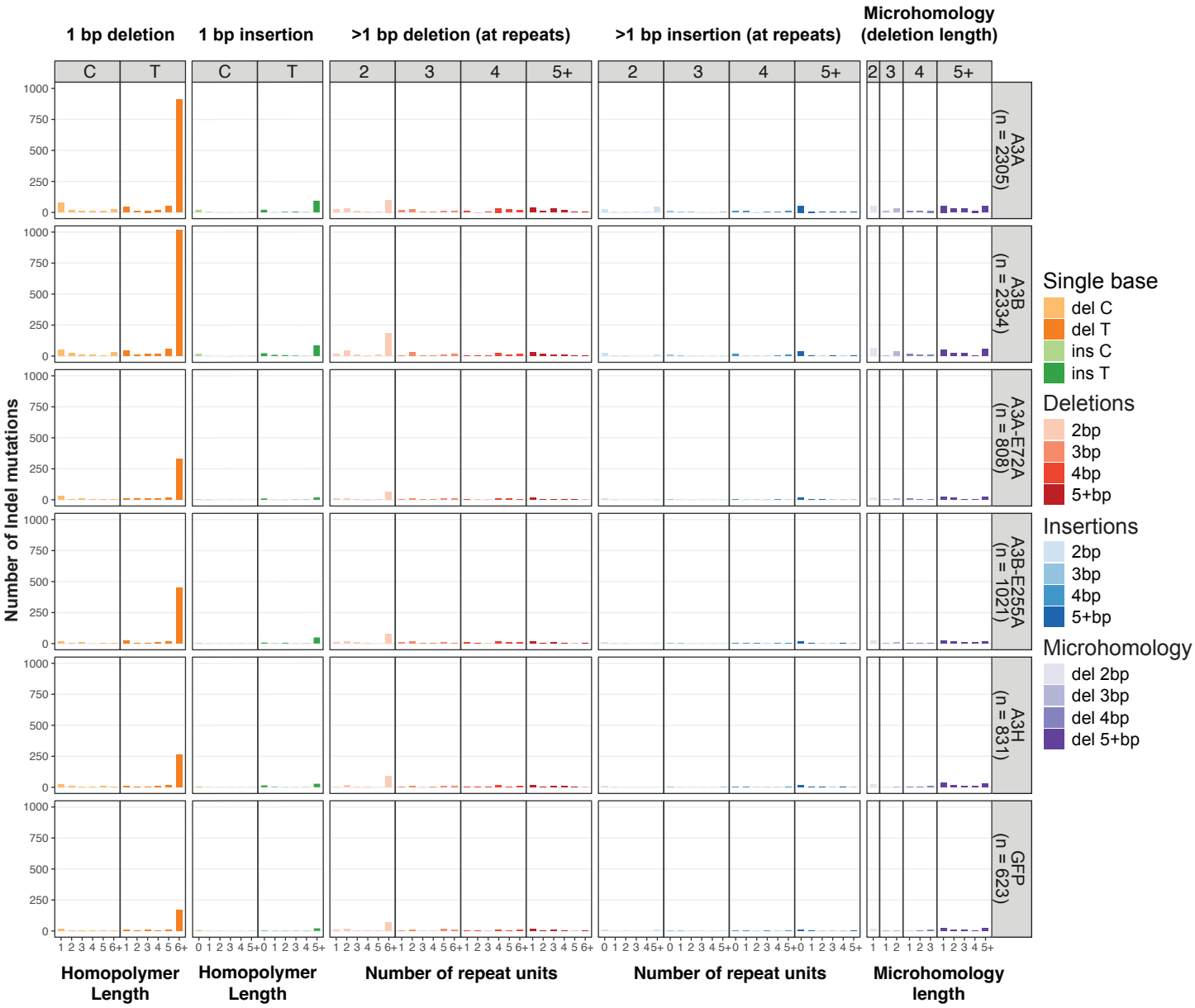


Figure 4

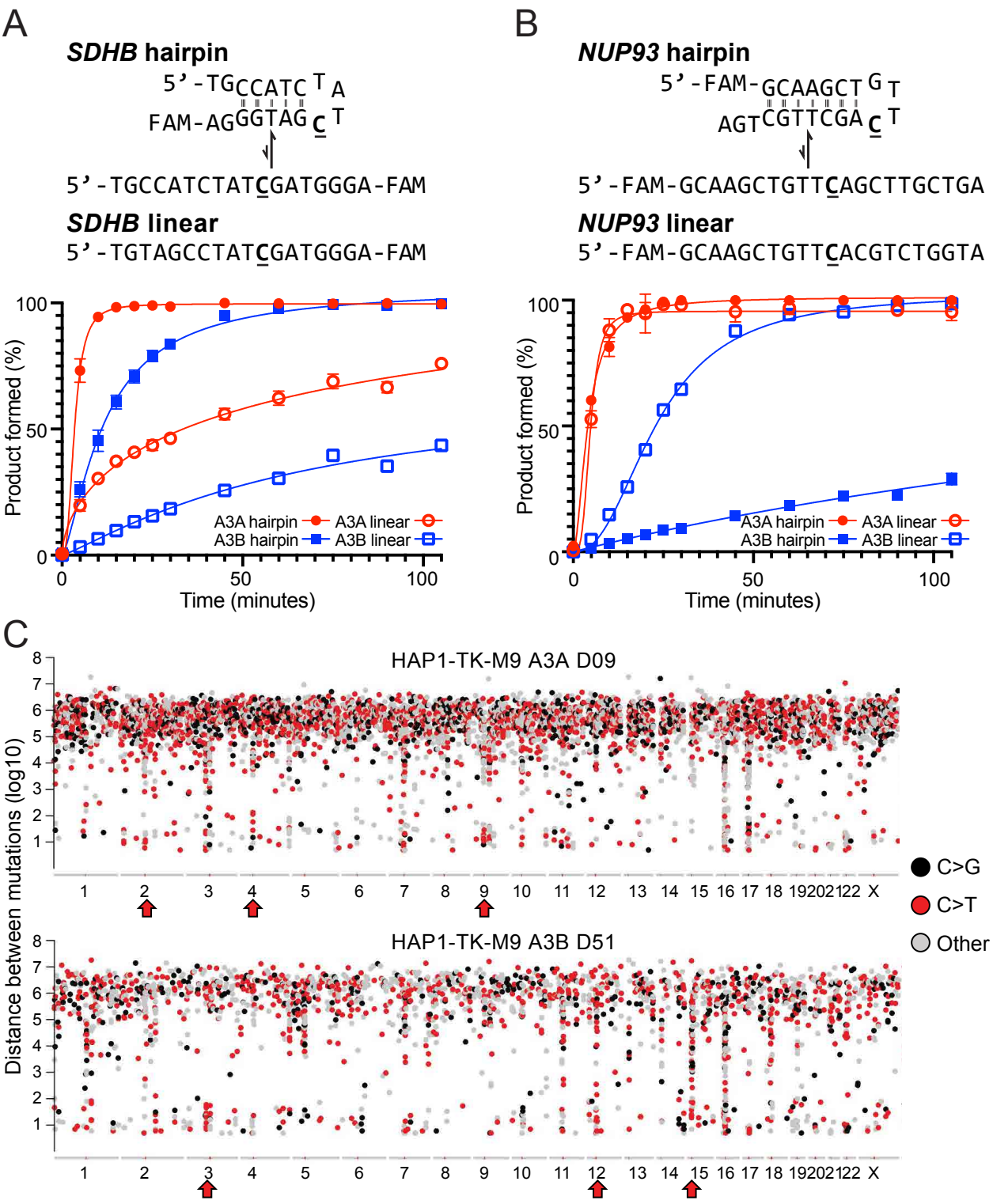


Figure 4 - Supp 1

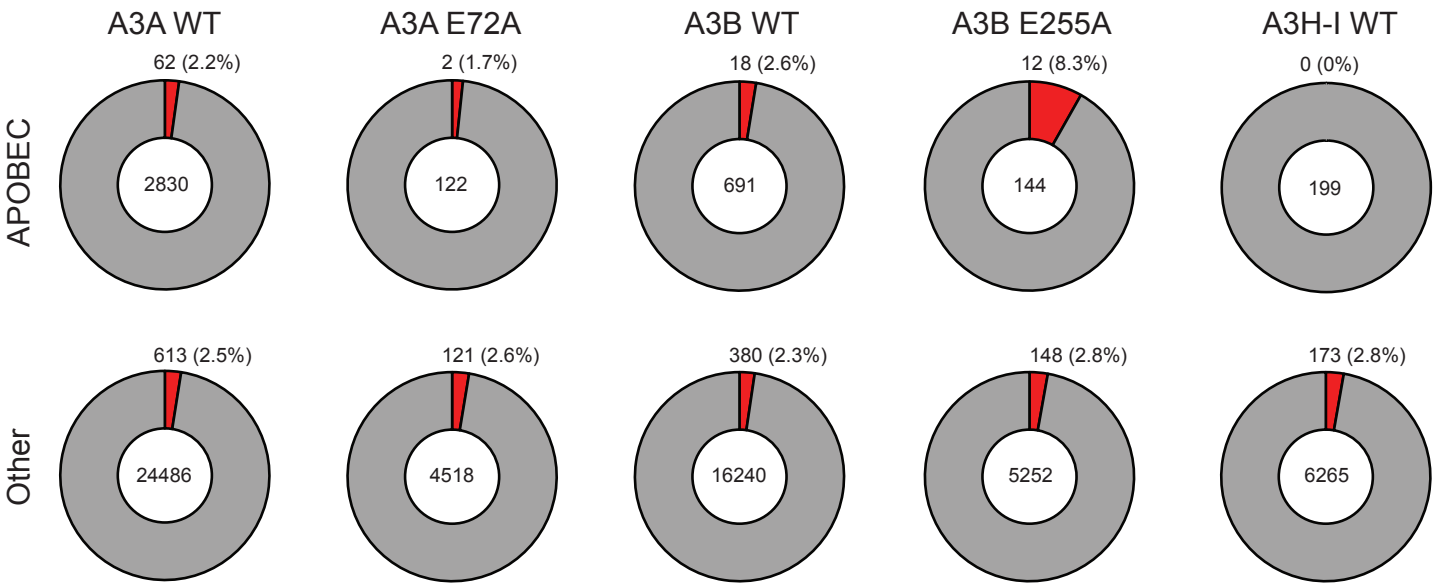


Figure 5