

RESEARCH ARTICLE

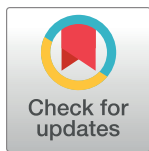
Mutational impact of APOBEC3A and APOBEC3B in a human cell line and comparisons to breast cancer

Michael A. Carpenter^{1,2,3,4,5}, Nuri A. Temiz^{2,6}, Mahmoud A. Ibrahim⁴, Matthew C. Jarvis^{1,2}, Margaret R. Brown^{1,2}, Prokopios P. Argyris^{1,2,3}, William L. Brown^{1,2}, Gabriel J. Starrett⁷, Douglas Yee^{2,8}, Reuben S. Harris^{1,2,3,4,5*}

1 Department of Biochemistry, Molecular Biology and Biophysics, University of Minnesota, Minneapolis, Minnesota, United States of America, **2** Masonic Cancer Center, University of Minnesota, Minneapolis, Minnesota, United States of America, **3** Howard Hughes Medical Institute, University of Minnesota, Minneapolis, Minnesota, United States of America, **4** Department of Biochemistry and Structural Biology, University of Texas Health San Antonio, San Antonio, Texas, United States of America, **5** Howard Hughes Medical Institute, University of Texas Health San Antonio, San Antonio, Texas, United States of America, **6** Institute for Health Informatics, University of Minnesota, Minneapolis, Minnesota, United States of America, **7** Center for Cancer Research, National Cancer Institute, NIH, Bethesda, Maryland, United States of America, **8** Department of Medicine, University of Minnesota Medical School, Minneapolis, Minnesota, United States of America

☞ These authors contributed equally to this work.

* rsh@uthscsa.edu



OPEN ACCESS

Citation: Carpenter MA, Temiz NA, Ibrahim MA, Jarvis MC, Brown MR, Argyris PP, et al. (2023) Mutational impact of APOBEC3A and APOBEC3B in a human cell line and comparisons to breast cancer. *PLoS Genet* 19(11): e1011043. <https://doi.org/10.1371/journal.pgen.1011043>

Editor: Sue Jinks-Robertson, Duke University, UNITED STATES

Received: July 31, 2023

Accepted: October 30, 2023

Published: November 30, 2023

Copyright: This is an open access article, free of all copyright, and may be freely reproduced, distributed, transmitted, modified, built upon, or otherwise used by anyone for any lawful purpose. The work is made available under the [Creative Commons CC0](https://creativecommons.org/licenses/by/4.0/) public domain dedication.

Data Availability Statement: All HAP1-TK-M9 clone alignment files (FASTQ and BAM format) are available through the Sequence Read Archive under the BioProject accession number PRJNA832427. S2 Table is a mutation annotation format (maf) file that provides a tabular list of all single base substitution mutations in the WGS described here. 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

Abstract

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

Genome Consortium (ICGC) were used in these analyses: <https://dcc.icgc.org/>.

Funding: These studies were supported by grants from the National Cancer Institute (NCI P01-CA234228 and NCI P50-CA247749 to RSH), and a Recruitment of Established Investigators Award from the Cancer Prevention and Research Institute of Texas (CPRIT RR220053 to RSH). Salary support was provided in part by grants from the National Cancer Institute (initially NCI T32-CA009138 to MCJ and subsequently NCI F31-CA243306 to MCJ). The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

Competing interests: I have read the journal's policy and the authors of this manuscript have the following competing interests: 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.

to help resolve a long-standing etiologic debate on the source of APOBEC3 signature mutations in cancer and indicate that future diagnostic and therapeutic efforts should focus on both APOBEC3A and APOBEC3B.

Author summary

A large source of mutation in cancer is attributable to single-stranded DNA cytosine to uracil (C-to-U) deamination by APOBEC3 family enzymes. Chromosomal DNA U's directly template the insertion of adenine (A), which results in C/G-to-T/A mutations. Chromosomal U's also trigger error-prone DNA repair processes, which less directly result in C/G-to-T/A mutations, C/G-to-G/C mutations, and more complex events such as insertion/deletion mutations. We report a human cell line for cancer mutation studies, where mutations can be identified selectively or by genomic DNA sequencing. Only APOBEC3A and APOBEC3B proved capable of inflicting the aforementioned types of single base substitution mutations. Interestingly, 70% and 50% of the single base substitution mutations attributable to APOBEC3A and APOBEC3B, respectively, occurred in TCA and TCT trinucleotide motifs preceded by a pyrimidine (T or C). This is significant because most APOBEC3-positive breast tumors exhibit an intermediate percentage of signature mutations in these broader tetranucleotide motifs, which supports the idea that both APOBEC3A and APOBEC3B contribute in a combinatorial manner to the overall mutation landscape in these tumors.

Introduction

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 pan-cancer analysis by [1] and reviews by [2–4]). Through these approaches and complementary wet lab 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, water-mediated 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 APOBEC3-catalyzed deamination is absent in most normal tissues, not associated with ageing, present in many primary tumors, and often enriched in metastases (APOBEC3 signature [1, 6, 8, 9, 18–23]).

The human APOBEC3 (A3) family is comprised of seven distinct 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 (herpesviruses, papillomaviruses, 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 contributing to the overall APOBEC3 mutation signature [46, 48, 49]. An important additional consideration is that A3A and A3B have been reported to exhibit broader tetranucleotide preferences upon expression in yeast, 5'-YTCW vs RTCW, respectively [50–52].

Here, we develop a human cellular system for mutation research and use it to compare 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 prior to undertaking additional experiments such as whole genome sequencing (WGS) and signature analysis. Using this system, only expression of A3A and A3B cause significant increases in Gan^{R} mutation frequencies. Sanger sequences from panels of individual *TK* mutant clones show a clear TC-biased mutation pattern including two hotspots with no obvious hairpin structure. WGS of independent clones demonstrates that both A3A and A3B can generate the APOBEC3 mutation signatures SBS2 and SBS13. A3A- and A3B-induced single base substitution mutations are mostly dispersed (non-clustered) throughout the genomes and both enzymes exhibit similar frequencies of mutation in TCW motifs in chromosomal DNA predicted to form non-hairpin versus hairpin structures. However, in comparison to catalytic mutant controls, both enzymes exhibit higher frequencies of APOBEC3 signature mutation in the single-stranded DNA loop regions of predicted hairpin structures. WGS also shows that A3A has a strong (slightly over 70%) preference for triggering cytosine mutations in YTCW sequence motifs, whereas A3B has a weaker (slightly under 50%) preference for the same motif. In comparison, APOBEC3 mutation signature-enriched primary breast cancers show predominantly intermediate frequencies of mutations in YTCW motifs, between 50% and 70%, suggesting involvement from both enzymes.

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 only recapitulate a subset of 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. Expression of the thymidine kinase (*TK*) protein confers exquisite sensitivity to the drug ganciclovir and, as for many antimicrobial agents, only *TK*-mutant, ganciclovir-resistant (Gan^{R}) cells survive selection by this drug. 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 **Fig 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 pre-existing chromosome aberrations), it cultures, engineers, and clones well (below), and it has a favorable *A3* expression profile (**Figs 1B and S1**). In particular, RT-qPCR measurements 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 (**Fig 1B**). Genomic DNA sequencing also revealed that this cell line’s only *A3H* allele is haplotype III (ΔAsn15), which is known to produce an unstable protein [58–60].

Functionality of human A3s expressed from MLV-based constructs

Plasmid constructs were assembled for these studies in which human *A3* expression is driven by a *MND* promoter from within an MLV-based retroviral construct. An additional feature of these constructs is a downstream puromycin resistance cassette, which facilitates selection of expressing cells. The functionality of each construct was assessed by transfection into 293T cells and, following 24–48 hrs incubation, protein analysis by immunoblotting and ssDNA deaminase activity assays (**Fig 1C–1E**; custom rabbit anti-human *A3A* monoclonal antibody validation described in **S2 Fig**). Each *A3* expressed at the expected kilodalton size and only the wildtype enzymes exhibited catalytic activity. *A3H*-I was expressed at lower levels than *A3H*-II and, accordingly, had less ssDNA deaminase activity in soluble extracts (**Fig 1E**), in agreement with prior reports [58–60].

We next used immunofluorescent (IF) microscopy to examine subcellular localization and DNA damage responses triggered by expression of each enzyme 3 days post-transduction into HAP1-TK-M9 cells (without puromycin selection). *A3A* appeared cell-wide and associated with a large increase in overall staining of the DNA damage marker γ -H2AX as well as an increase in individual γ -H2AX foci (representative images in **Fig 1F**, quantification in **Fig 1G and 1H**, and additional images in **S3 Fig**). Expression of an enzymatically inactive mutant, *A3A*-E72A, exhibited vector control levels of γ -H2AX staining. In comparison, neither expression of wildtype *A3B* nor *A3H* (haplotype I or II) was able to trigger statistically significant increases in γ -H2AX levels in HAP1-TK-M9 cells (representative images in **Fig 1F**, quantification in **Fig 1G and 1H**, and additional images in **S3 Fig**). These results were unexpected given that *A3B* localization is predominantly nuclear, and *A3H* (haplotypes I and II) can also access the nuclear compartment with some accumulation in nucleoli (**Fig 1F** and additional images in **S3 Fig**). These results differed from prior reports on *A3B* and *A3H* overexpression causing elevated γ -H2AX levels [9, 43, 48, 61], perhaps because expression levels here are lower with the *MND* promoter (versus strong Tet/Dox-inducible systems) and/or because the HAP1 system is somehow unique and more tolerant/adaptable to expression of these ssDNA deaminases. However, HAP1-TK-M9 cells are indeed capable of a canonical DNA damage response, as demonstrated by γ -H2AX accumulation following treatment with cisplatin (*cis*-diamminedichloroplatinum II; representative images in **Fig 1F**, quantification in **1G and 1H**, and additional images in **S3 Fig**).

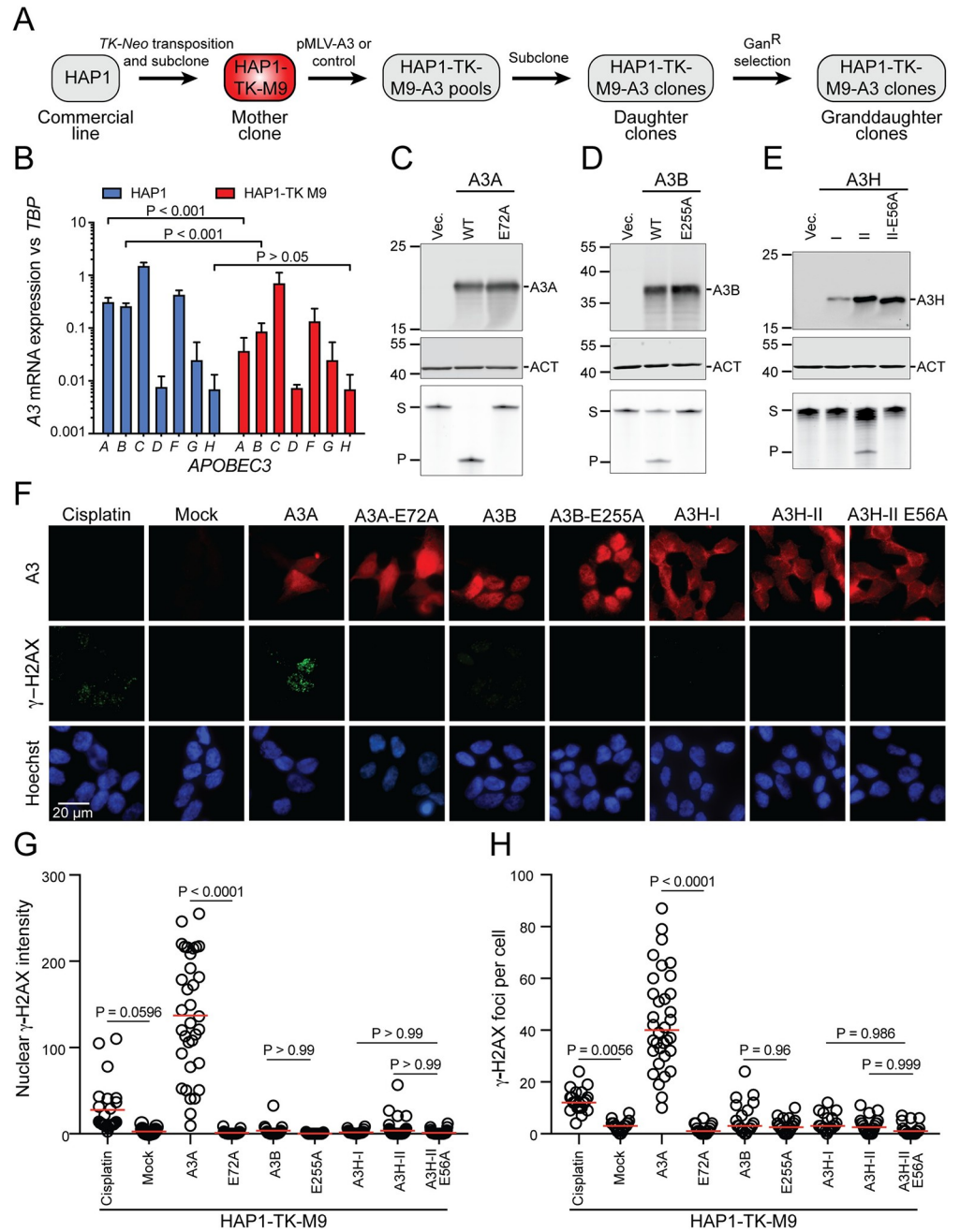


Fig 1. A3 activity in the HAP1-TK-M9 mutation reporter system. (A) Schematic of the construction of the HAP1-TK-M9 system and overall experimental workflow (see S1 Fig for HAP1-TK-M9 karyotype). (B) A3 mRNA levels in parental HAP1 cells in comparison to the HAP1-TK-M9 daughter clone by RT-qPCR (p-values from Welch's t-test). (C-E) Immunoblots of A3A, A3B, and A3H in WCE 24 hrs post-transfection of 293T cells. Primary antibodies are UMN-13, 5210-87-13, and P1-D8, respectively. Mouse anti-β-actin (ACT) is a loading control for the A3A and A3B blots, and rabbit anti-β-actin (ACT) for the A3H blot (see S2 Fig for anti-A3A UMN-13 mAb validation). The lower images show ssDNA deaminase activity of extracts from the same cell populations (S, substrate; P, product). (F) IF microscopy images of HAP1-TK-M9 cells expressing the indicated A3 enzymes or treated with 4 μM cisplatin for 24 hrs. A3 staining is red, γ-H2AX staining is green, and nuclei are blue from Hoechst (scale bar = 20 μm). (G-H) Quantification of pan-nuclear γ-H2AX intensity and discrete γ-H2AX foci, respectively, for each condition described in panel F (each data point represents an independent cell; n>25 cells per condition; p-values from one-way ANOVA test; red bar indicates mean for each group; representative of 2 biologically independent experiments).

<https://doi.org/10.1371/journal.pgen.1011043.g001>

To extend these results to a different cell line, our MLV-based A3 expression plasmids were transfected transiently into HeLa cells and, after 24 hrs incubation, subjected to additional analysis by IF microscopy. As with HAP1-TK-M9 cells, A3A appeared cell-wide (except nucleoli), A3B predominantly nuclear, and A3H haplotype II cytoplasmic with nucleolar accumulations (S4A Fig). A3H haplotype I was not analyzed here due to low expression and weak ssDNA deaminase activity in HAP1-TK-M9 cells. Interestingly, A3A caused a strong pan-nuclear increase in γ -H2AX without concomitant focus formation (representative images in S4A and quantification in S4B Fig). Similar results have been reported for A3A overexpression in other cell types [30, 62–64]. Interestingly, although a dose-responsive accumulation of γ -H2AX was expected given a wide range of transient transfection efficiencies for individual cells in each reaction, only a weak positive association was found because even low A3A staining cells exhibit high γ -H2AX levels (S4B and S4C Fig). In contrast, comparatively low levels of nuclear γ -H2AX were observed in cells expressing A3B, A3H, and catalytic mutant derivative proteins, although A3B expression uniquely triggered a modest elevation of nuclear γ -H2AX levels in comparison to the background observed in cells expressing the corresponding catalytic mutant E255A protein (S4A–S4D Fig). Moreover, as expected from the strong increase in γ -H2AX levels, only wildtype A3A caused significant increases in DNA breakage as quantified by alkaline comet assays (representative images in S4E and quantification in S4F Fig). Taken together, results with two different cell lines demonstrate that A3A expression induces a strong DNA damage response (high γ -H2AX and DNA breakage), A3B expression triggers a modest DNA damage response (low γ -H2AX and no overt DNA breakage), and A3H or catalytic mutant A3A/B/H expression is indistinguishable from background levels in negative vector control conditions.

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 (Methods, Table 1). 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, SEM = 0.56). 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 = 17, mean = 14, SEM = 2.5; A3B WT: median = 11, mean = 12, SEM = 1.8; Fig 2A). In comparison, expression of A3H-I or A3H-II or catalytic mutant derivatives of any of these DNA deaminases or empty vector controls failed to trigger increased Gan^R mutation frequencies (A3H-I WT: median = 3.0, mean = 3.6, SEM = 0.71; A3H-II WT: median = 3.0, mean = 3.3, SEM = 0.65; Fig 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 Sanger sequenced. C-to-T and C-to-G mutations in an APOBEC3-signature trinucleotide motif (APOBEC), all other single base substitution mutations (other SBS), and all insertion/deletion mutations (INDELs) were placed into groups for comparison (red, black, and blue tics in Fig 2B, respectively; individual sequence schematics in S5A Fig). *TK* sequences derived from A3A-expressing granddaughter clones harbored a greater number of APOBEC3 signature mutations [12/20 clones contained at least 1 APOBEC mutation, 22 T(C>T/G)W mutations

Table 1. List of key resources.

Reagent type (species) or resource	Designation	Source or reference	Identifiers	Additional information
Cell line (<i>homo sapiens</i> , male)	HAP1	Horizon	Cat#: C859	
Cell line (<i>homo sapiens</i> , male)	HAP1-TK-M9	Original clone, this study		Near diploid clone that expresses HSV-1 TK and is Gan ^S ; request by contacting RSH
Cell line (<i>homo sapiens</i> , female)	293T	ATCC	Cat#: CRL-3216 RRID: CVCL_0063	
Cell line (<i>homo sapiens</i> , male)	THP1	ATCC	Cat#: TIB-202 RRID: CVCL_0006	
Cell line (<i>Homo sapiens</i> , female)	HeLa	NIH ARP	ARP-153	
Antibody	Anti- β -actin (mouse monoclonal)	Millipore Sigma	Cat #: A1978	WB (1:5000)
Antibody	Anti- β -actin (rabbit monoclonal)	Cell Signaling Technology	Cat#:4970	WB (1:5000)
Antibody	Anti-A3A (rabbit monoclonal)	This study	UMN-13	WB (1:1500); request by contacting RSH
Antibody	Anti-A3B (rabbit monoclonal)	[101]	RRID:AB_2721202	WB (1:1,000) IF (1:300); request by contacting RSH
Antibody	Anti-A3H (mouse monoclonal)	[58]	Clone P1-D8	WB (1:1,000); request by contacting RSH
Antibody	Anti-A3H (rabbit polyclonal)	Novus ARP10	Cat#: NBP1-91682 RRID: AB_2057523	IF (1:300)
Antibody	Anti- γ -H2AX (S139) (rabbit polyclonal)	Millipore Sigma JW301	Cat#: H5912 RRID: AB_310406	IF (1:500)
Antibody	Anti-Tubulin (mouse monoclonal)	Sigma-Aldrich	Cat#: T5168 RRID: AB_477579	WB (1:20,000)
Recombinant DNA reagent	pQCXIP MLV vector	NovoPro	Cat#: V010396#	Parental transduction vector
Recombinant DNA reagent	pQCXIP-MND-A3Ai-IRES-Puro	This study	A3A with intron (pRH9978)	Request by contacting RSH
Recombinant DNA reagent	pQCXIP-MND-A3Ai-E72A-IRES-Puro	This study	A3A-E72A with intron (pRH9979)	Request by contacting RSH
Recombinant DNA reagent	pQCXIP-MND-A3Bi-IRES-Puro	This study	A3B with intron (pRH9980)	Request by contacting RSH
Recombinant DNA reagent	pQCXIP-MND-A3Bi-E255A-IRES-Puro	This study	A3B E255A with intron (pRH9981)	Request by contacting RSH
Recombinant DNA reagent	pQCXIP-MND-eGFP-IRES-Puro	This study	eGFP vector control (pRH9977)	Request by contacting RSH
Recombinant DNA reagent	pQCXIP-MND-A3H-Ii-IRES-Puro	This study	A3H hapI with intron (pRH9984)	Request by contacting RSH
Recombinant DNA reagent	pQCXIP-MND-A3H-Iii-IRES-Puro	This study	A3H hapII with intron (pRH9985)	Request by contacting RSH
Recombinant DNA reagent	pQCXIP-MND-A3H-Iii-E56A-IRES-Puro	This study	A3H hapII E56A with intron (pRH9986)	Request by contacting RSH
Software, algorithm	Burrows-Wheeler Aligner (BWA)		RRID: SCR_010910	
Software, algorithm	Fiji	Fiji	RRID: SCR_002285	

(Continued)

Table 1. (Continued)

Reagent type (species) or resource	Designation	Source or reference	Identifiers	Additional information
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

<https://doi.org/10.1371/journal.pgen.1011043.t001>

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 APOBEC3 signature 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]

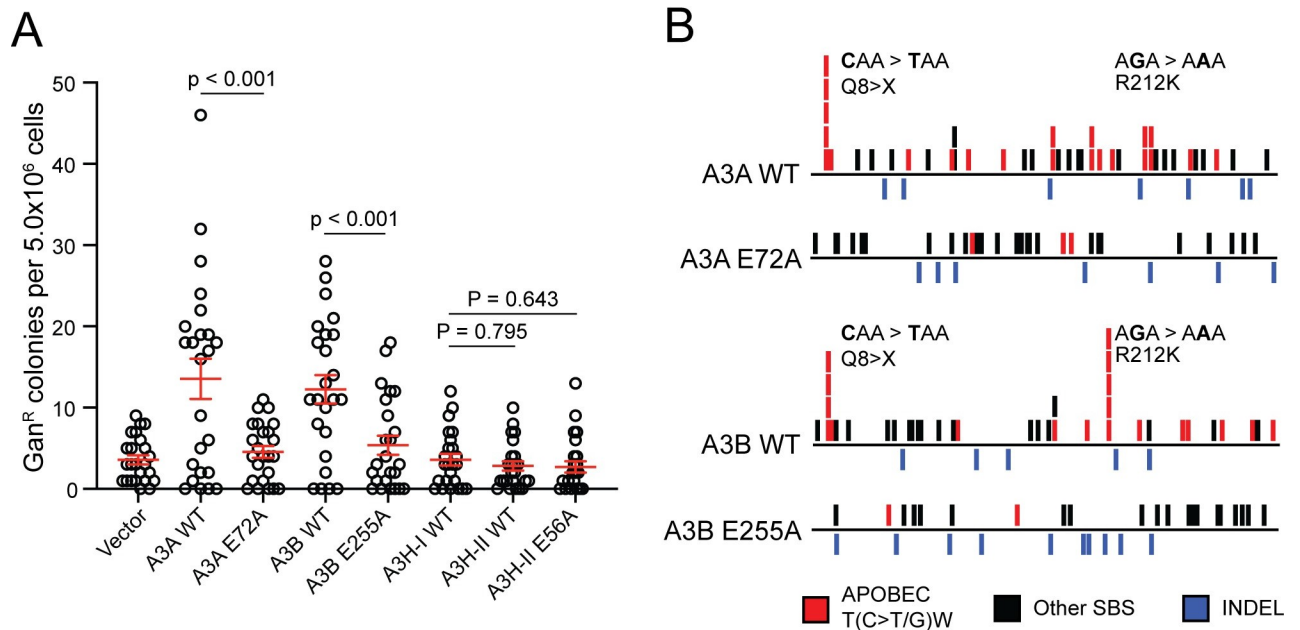


Fig 2. Characterization of *TK* mutations in ganciclovir-resistant clones. (A) A dot plot of Gan^R colonies generated under the indicated A3 expression or control conditions. Each data point represents the number of Gan^R mutants in a single clonal culture (mean +/- SD shown with p-values determined 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 S5 Fig for schematics of individual *TK* mutants for these and vector control conditions). Q8X and R212K mutation hotspots are labeled.

<https://doi.org/10.1371/journal.pgen.1011043.g002>

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] (Figs 2B and S5A). Additionally, TK cytosine nucleobases 22 (Q8X) and 635 (R212K) emerged as candidate mutational hotspots in both the A3A- and A3B-expressing clones (Fig 2B; considered further in Discussion). No significant 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 APOBEC3 signature mutations that occurred at 5'-TC dinucleotides in TK (S5B and S5C Fig). 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.

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

Independent granddaughter clones were selected using the workflow described above and RNA sequencing was done to confirm expression of each exogenously expressed A3 construct and compare mRNA levels relative to established breast cell lines and primary breast tumors. All expression values were determined relative to those of the conserved housekeeping gene *TBP* to be able to compare RNAseq data from different sources (e.g., unrelated cell lines and tumors). First, as expected from utilizing the same expression vector, mRNA levels of exogenously expressed A3A and A3B are similar (i.e., both averaging near 1 *TBP* equivalent; S6A Fig). However, these A3A and A3A-E72A mRNA levels in granddaughter clones are over 5-fold higher than the average endogenous A3A expression levels of APOBEC3 signature-enriched breast cancer cell lines BT474 and MDA-MD-453, breast cancer cell lines of the CCLE, or breast tumors of TCGA (S6A Fig). In comparison, the A3B and A3B-E255A mRNA levels in granddaughter clones are similar to averages reported for breast cancer cell lines of the CCLE and breast tumors of the TCGA, and approximately 2-fold lower than those of BT474 and MDA-MD-453 (S6A Fig). In other words, A3A is overexpressed in this system and A3B approximates levels observed in breast tumors and cell lines. A3H (haplotype I) mRNA levels showed greater variance but only two clones were analyzed by RNAseq and WGS due to negative results above with IF microscopy experiments and TK mutation analysis.

The mRNA expression levels of the other four A3 genes, as well as *AICDA* (*AID*), *APOBEC1*, *APOBEC2*, and *APOBEC4*, were also quantified and compared with those of A3A, A3B, and A3H (S6B Fig). Endogenous A3C was expressed at similarly high levels in all granddaughter clones, providing a relatively stable internal control. Endogenous A3F and A3G were expressed at lower but still detectable levels, and endogenous *AICDA*, *APOBEC1*, *APOBEC2*, *APOBEC4*, and A3D were expressed at very low or undetectable levels. As expected, levels of ectopically expressed A3A, A3B, and A3H mRNA exceeded those in the HAP1-TK-M9 parent clone as well as those in vector expressing granddaughter controls. In addition, protein

expression of A3A, A3B, and A3H was confirmed in granddaughter clones by immunoblotting and activity by ssDNA deamination assays (S7A and S7B Fig). Finally, the integration sites of the *TK* reporter and MLV-based A3 expression constructs were determined using WGS reads. Data from representative clones demonstrated a single *TK-Neo* integration site in chromosome 3 between nucleotides 143996100–143996500, which is ~3.5 kbp downstream of the nearest gene (*DIPK2A*), as well as unique MLV-A3 insertion sites as expected from the independent reactions and low multiplicities of infection that were used initially to establish the clones (S8 Fig).

To investigate mutational differences genome-wide, Illumina short-read WGS was done for randomly selected granddaughter clones. 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 A3 or control expression construct. Thus, all new SBS mutations had to be present in a significant proportion of reads from the granddaughter clones and absent from the reads from the original mother clone and, as such, must have occurred in the presence of an active A3 enzyme or a catalytic mutant control.

In A3A-expressing granddaughter clones, the total number of unique SBSs ranged from 2057 to 5256 ($n = 6$, median = 3129, mean = 3494, SD = 1358; S9A Fig). The total number of SBSs in A3B-expressing clones was lower, ranging from 1920 to 2652 ($n = 5$, median = 2346, mean = 2334, SD = 278). In comparison, the total number of SBSs in catalytic mutant and eGFP control granddaughter clones ranged from 1230 to 2204 ($n = 6$; median = 1748, mean = 1729, SD = 388). SBS mutations are further broken down into those occurring within NCN and TCW motifs in dot plots in S9B–S9E Fig. 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 (Figs 3A, S9C, S9E and S10). In other words, only A3A- and A3B-expressing granddaughter clones exhibited significant accumulations of APOBEC3 signature single base substitution mutations.

This result was confirmed by assessing APOBEC3 signature enrichment scores [50, 65], which indicated that 6/6 A3A-expressing clones and 4/5 A3B-expressing clones have significant enrichments of APOBEC3 signature mutations, whereas clones expressing catalytically inactive A3A or A3B, as well as clones expressing A3H-I or vector control have none (S11A Fig). A complementary bioinformatics analysis, non-negative matrix factorization (NMF [66]), 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 presence or functionality (S11B Fig). Signature B is comprised in part of a C-to-A mutation bias, which may be explained by the CA dinucleotide biased mutagenic activity of ganciclovir [67–69] but may also reflect a combination of mutational process and, in both cases, eclipse potential APOBEC3-instigated C-to-A mutational events; S11B Fig and also evident in trinucleotide breakdowns in Fig 3A). In comparison, genome-wide patterns of insertion/deletion mutations (INDELs) appeared largely unaffected by A3A or A3B (S12 Fig), in agreement with aforementioned *TK* mutation data where the majority of mutations are single base substitutions.

We next analyzed the broader contexts of the 5'-TC-focused C-to-T and C-to-G single base substitution mutations that accumulated in A3A- and A3B-expressing clones ($n = 7172$ and $n = 2033$, respectively) in comparison to those that accumulated in aggregate control clones ($n = 680$) as well as the overall distribution of 5'-TC in the human genome ($n = 339619283$) (Fig 3B and S1 Table). First, a bias for +1 A over +1 T emerged in A3A-expressing clones (48.3% > 36.3%), whereas the opposite bias was evident in A3B-expressing clones (33.9% <

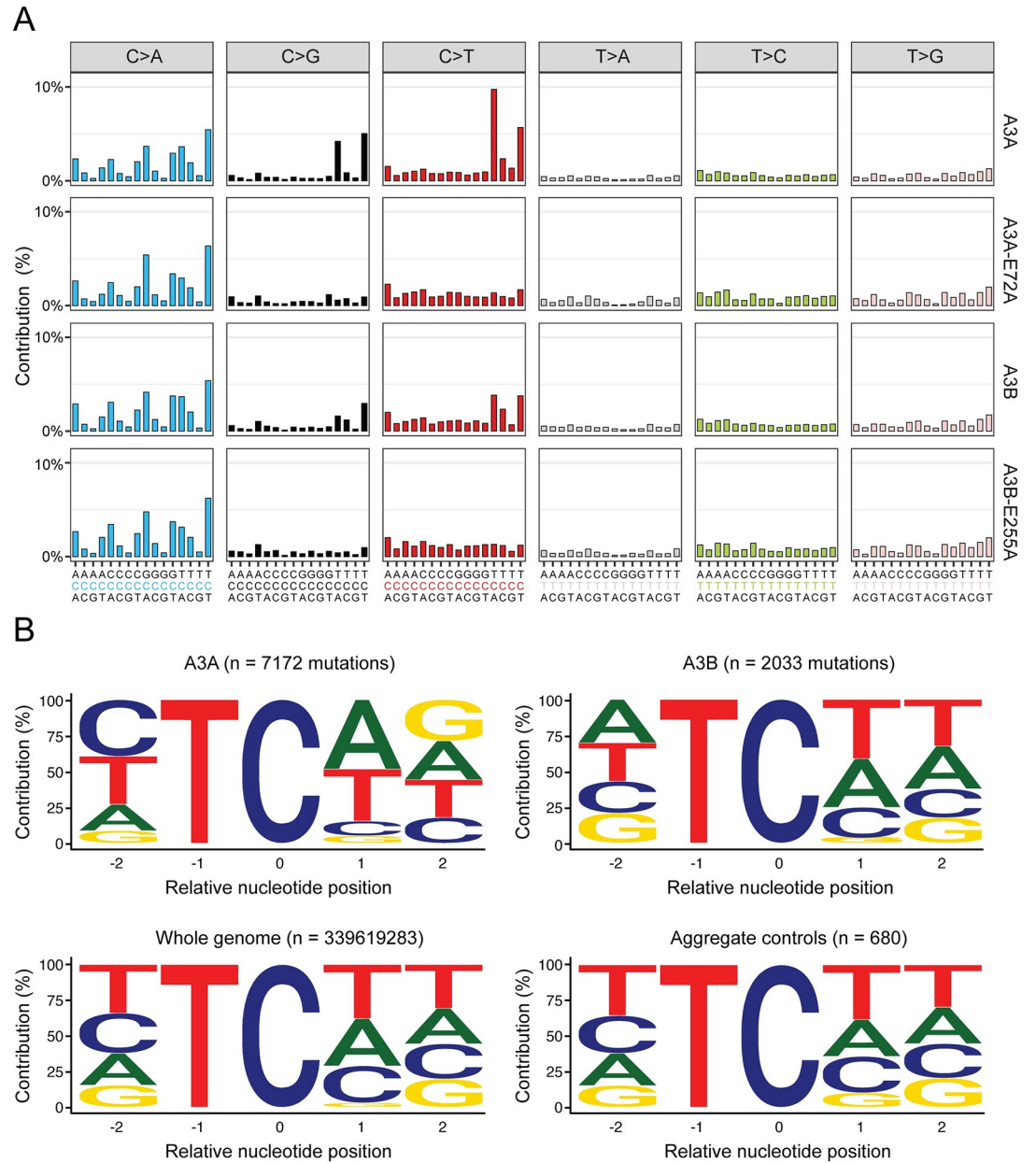


Fig 3. Single base substitution mutation signatures in granddaughter clone genomes. (A) Trinucleotide profiles of pooled SBSs across all clones sequenced for each listed experimental condition (A3A n = 6; A3A-E72A n = 2; A3B n = 5; A3B-E255A n = 2). See [S6](#) and [S7 Figs](#) for mRNA and protein level expression confirmation, respectively, and [S10 Fig](#) 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 and eGFP only conditions, which do not show evidence for APOBEC3 signature mutations; [S10 Fig](#)). The distribution of nucleobases flanking all TC motifs in the human genome is shown for comparison.

<https://doi.org/10.1371/journal.pgen.1011043.g003>

41.2%). However, for both enzymes, the percentage of +1 A and T (W) was similar (84.6% and 75.1%, respectively). Second, no significant bias was noted at the +2 position except that guanine is slightly over-represented in pentanucleotide motifs derived from A3A-expressing conditions compared to motifs derived from A3B-expressing clones. Third and most importantly, a strong bias for a pyrimidine nucleobase (C or T) occurred at the -2 position in A3A-

expressing clones (72.9% YTC/NTC), which was also evident in the broader APOBEC3 tetranucleotide context (72.7% YTCW/NTCW). This cytosine mutation preference resembles the strong -2 pyrimidine bias reported for human A3A in murine hepatocellular carcinomas [70], the chicken B cell line DT40 [71], and yeast [50, 51]. Fourth, in contrast, a slight bias for purine nucleobases (A or G) was apparent at the -2 position in A3B-expressing clones (51.1% RTC/NTC), which was also reflected in the broader APOBEC3 tetranucleotide context (52.8% RTCW/NTCW). This latter result also agreed with prior data from yeast [50, 51], but contrasted slightly with the mutation signature detected in tumors derived from human A3B-expressing mice (47% RTCW [72]). Possible explanations for this variability are considered in **Discussion**.

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

X-ray structures revealed a U-shaped bend in ssDNA substrates bound by A3A and A3B [16, 29], and other studies indicated that similarly bent ssDNA loop regions of hairpins (*i.e.*, DNA cruciform or stem-loop structures) can be preferred substrates for deamination by A3A [52, 73, 74]. To ask whether this preference extends 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 appeared to be part of predicted 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 base substitution mutations, the frequency of APOBEC3 signature TCW mutations was similar in ssDNA loop regions of predicted hairpin structures versus non-hairpin regions (*i.e.*, APOBEC3 signature mutation events were not enriched in the loop regions of stem-loop structures over those occurring in canonical ssDNA substrates; **S13 Fig**). Moreover, the frequency of APOBEC3 signature mutations in hairpin or non-hairpin structures appeared similar (not distinguishable statistically) in A3A- and A3B-expressing conditions. However, interestingly, the frequency of APOBEC3 signature mutations in predicted loop regions of hairpin structures appeared higher in A3A and A3B expressing clones in comparison to all non-catalytic control conditions (**S13 Fig**; $P = 0.0655$ and $P = 0.0367$ by Welch's t-test, respectively). The A3A versus control comparison likely failed to reach statistical significance due to the small number of clones with WGS ($n = 6$) and the large variance in numbers of TCW mutations in hairpin loop regions in the different clones (range = 1–16 TCW mutations). Nevertheless, these results combined to indicate that single-stranded loop regions of hairpin structures in human chromosomal DNA may be similarly susceptible to deamination by both A3A and A3B.

To assess relative rates of A3A and A3B-catalyzed deamination of experimental hairpin versus non-hairpin substrates, we used purified enzymes and ssDNA substrates representing two previously reported RNA editing hotspots of A3A –*SDHB* and *NUP93* [52, 75]. The *SDHB* hairpin is predicted to have a 5 bp stem and a 4 nt loop, and the *NUP93* hairpin a 7 bp stem and a 4 nt loop (**Fig 4A and 4B**). The control oligonucleotides have the same loop region sequences and a randomization of one-half of the nucleobases in the hairpin stem to reduce base-pairing potential. In each case, the linear substrates migrated similarly on native and denaturing PAGE while the hairpin substrates migrated faster by native PAGE and similar to the linear substrates when denatured, thereby confirming the integrity of both hairpins (**S14 Fig**). A3A and A3B were affinity-purified from human cells and incubated under single hit conditions with these oligonucleotide substrates over time. First, both A3A and A3B showed a strong preference for deaminating the *SDHB* hairpin substrate in comparison to a linear

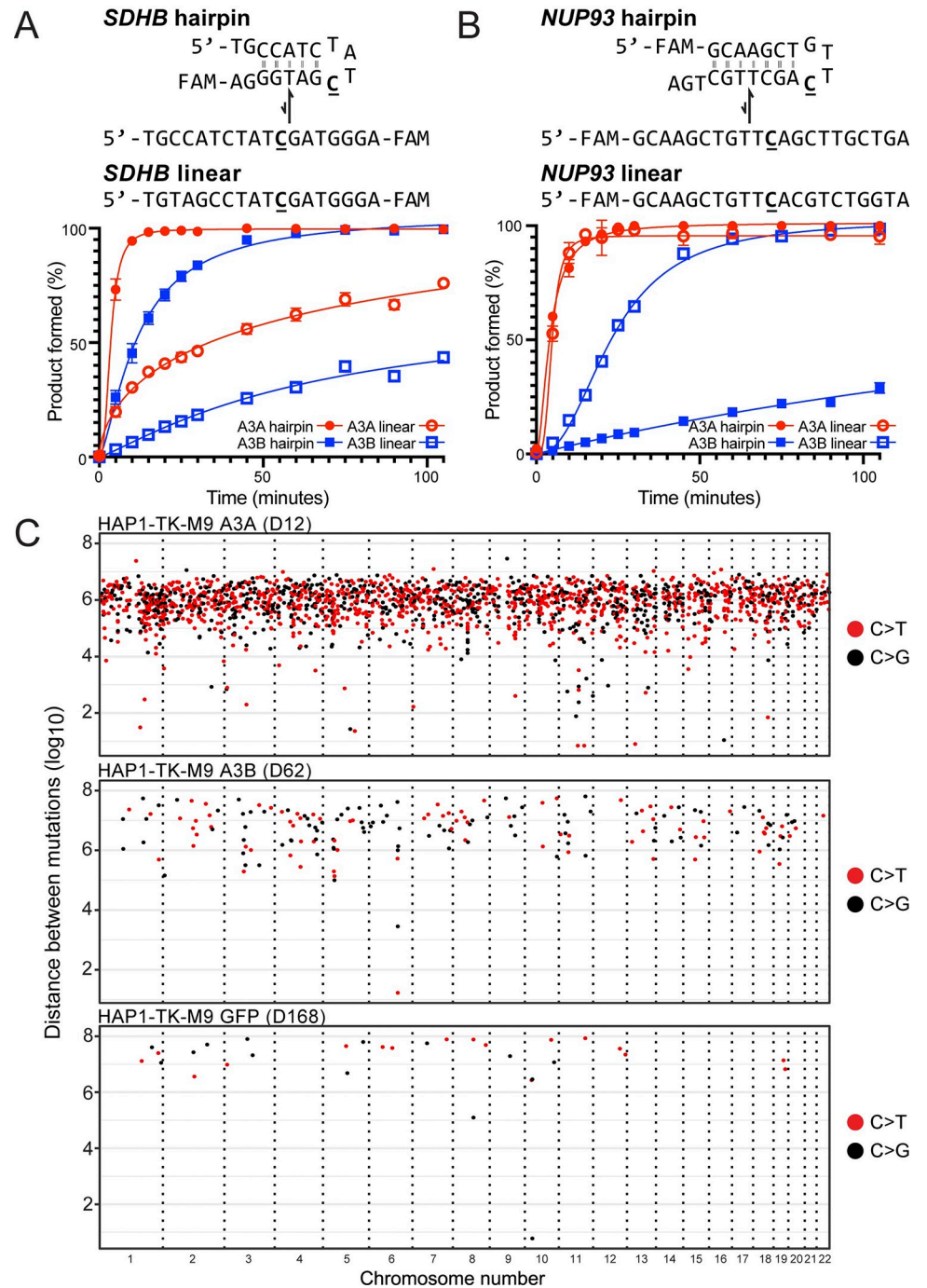


Fig 4. Mesoscale properties of A3A and A3B *in vitro* and in the HAP1-TK-M9 system. (A-B) Deamination kinetics of A3A and A3B using *SDHB* and *NUP93* DNA hairpin substrates in comparison to corresponding linear controls made by scrambling the 5' or 3' portion of the stem, respectively. See text for full description and [S13 Fig](#) for a genome-wide analysis and [S14 Fig](#) for a gel-based confirmation of DNA oligonucleotide integrity. (C) Rainfall plots of genome-wide intermutation distances for APOBEC3 signature single base substitution mutations in representative granddaughter clones (C>T mutations are red, C>G black, other SBS gray). See [S15 Fig](#) for an additional analysis of dispersed versus clustered single base substitution mutations.

<https://doi.org/10.1371/journal.pgen.1011043.g004>

control with the same nucleobase content scrambled (~4- and ~8-fold preference, respectively; Fig 4A). A3A showed 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 were: 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 (Fig 4B). Rates of A3A-catalyzed deamination were high for both 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 combined to show that *both* A3A and A3B can deaminate hairpin and linear ssDNA substrates and, further, that it is possible to identify sites such as the *NUP93* hairpin (here with DNA and previously shown with RNA and DNA [52]) that are strongly (and perhaps even exclusively, in a few instances) preferred by a single A3 enzyme.

Another 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*; here defined as ≥ 2 strand-coordinated APOBEC3 signature mutations within a 10 kbp window) [5, 7, 66, 76]. No APOBEC3 signature *kataegis* events were observed in control conditions. However, clear APOBEC3 signature *kataegis* were evident in the genomic DNA of both A3A and A3B expressing granddaughter clones (e.g., Fig 4C). For instance, one A3A-attributable *kataegis* event was comprised of 5 T (C>T/G)W mutations, and an A3B-attributable *kataegis* event included 3 T(C>T/G)W mutations. Interestingly, however, the frequency of APOBEC3 signature *kataegis* did not differ significantly between A3A- and A3B-expressing granddaughter clones (A3A: 4 events; A3B:1 event; $p = 0.14$ by Welch's t-test). These results show that both A3A and A3B can cause *kataegis* in a human cell line (also see S15 Fig for additional analyses and depictions of dispersed versus clustered single base substitution mutations).

APOBEC signature etiology in primary breast tumors

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

Both the A3A-like and A3B-like groups were comprised of tumors that show significant levels of APOBEC3 signature mutations and correspondingly high enrichment scores (Fig 5 and S1 Table). However, only a small proportion of the A3A-like group of breast tumors exhibit a

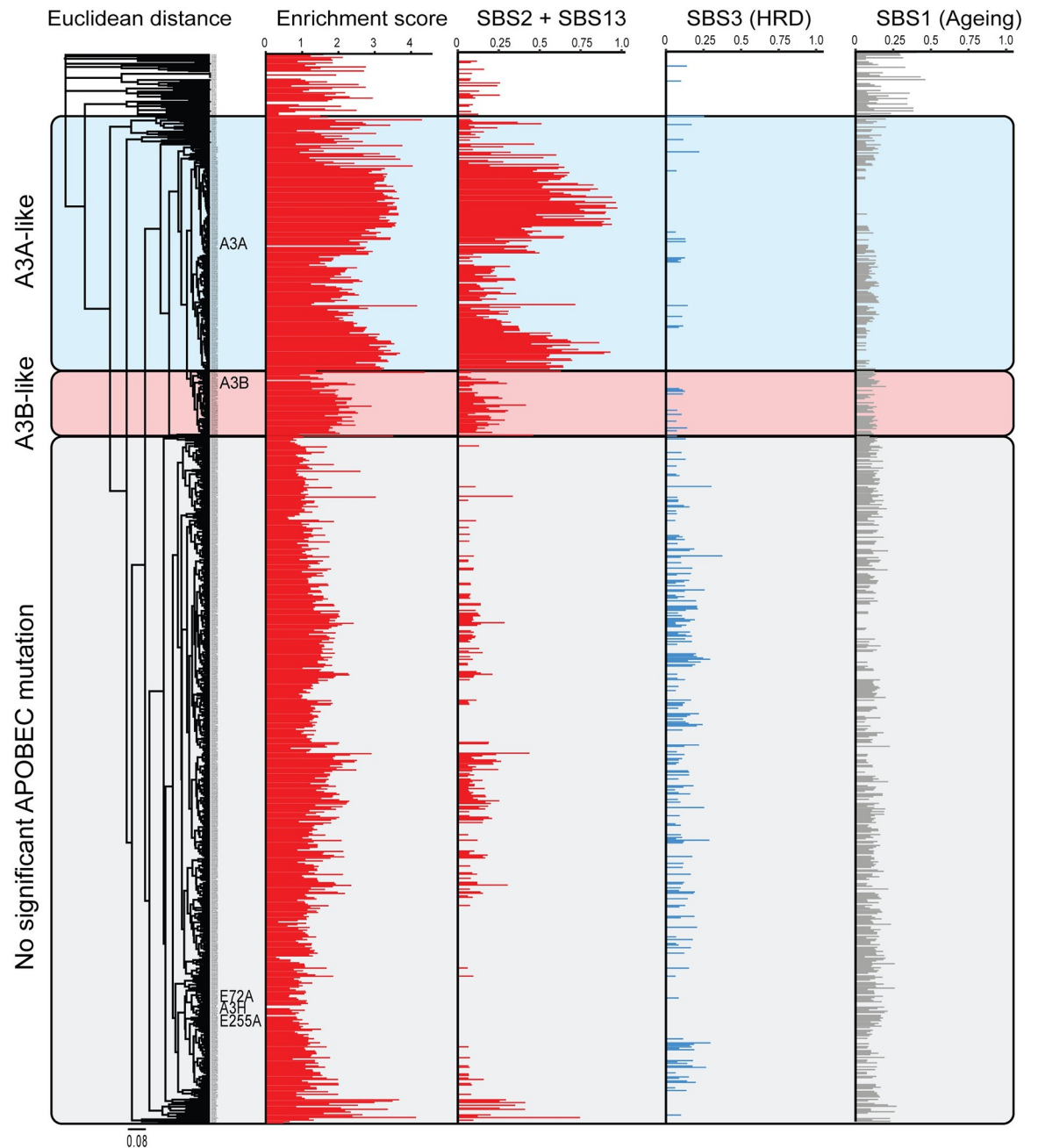


Fig 5. A composite origin of APOBEC3 signature mutations in breast cancer. An unsupervised clustering analysis of similarity between the pentanucleotide SBS profiles from WGSs of the A3A, A3B, and control granddaughter clones described here versus those from primary breast tumor whole-genome sequencing data sets (ICGC, $n = 784$). The APOBEC3 mutation signature is represented by both enrichment score and SBS2+13 (red), HRD signature as SBS3 (blue), and ageing signature as SBS1 (gray).

<https://doi.org/10.1371/journal.pgen.1011043.g005>

percentage of APOBEC3 signature mutations in YTCW motifs that approaches the level inflicted by A3A here in the HAP1-TK-M9 system (*i.e.*, $\geq 70\%$). Similarly strong YTCW preferences have also been reported recently for A3A expression in the chicken B cell line DT40 [71] and for A3A-driven murine hepatocellular carcinomas [70]. The lower percentages of APOBEC3 signature mutations in YTCW motifs in most breast tumors therefore suggests that

A3A alone may not account for the observed composite APOBEC3 mutation signature and, moreover, that many tumors in A3A-like group may include contributions from another A3 enzyme, most likely A3B. As regards the A3B-like group, most tumors have higher percentages of C-to-T and C-to-G mutations in YTCW motifs than can be explained by A3B alone. For instance, most tumors manifest a greater proportion of YTCW mutations than observed here for A3B in the HAP1-TK-M9 system (47.2%) and for A3B-driven tumors in mice (53%) [72]. These intrinsic preferences therefore combined to suggest that the observed percentages of APOBEC3 signature mutations in most breast tumors may be a composite resulting from the combinatorial activities of both A3A and A3B activity. In comparison, the homologous recombination repair deficiency (HRD) signature (SBS3) appeared underrepresented in both the A3A- and A3B-like tumor groups, and the ageing signature (SBS1) occurred in all three groups regardless of the presence or absence of an APOBEC3 mutation signature enrichment (Fig 5).

Discussion

Here we report the development and implementation of a genetic system to investigate mutational processes in the human HAP1 cell line. 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 demonstrated a strong shift in the mutational pattern from a variety of different base substitution mutations in control conditions to a strongly 5'TC-biased pattern (Fig 2). Interestingly, the *TK* mutation spectra inflicted by A3A and A3B were very similar including two shared hotspots (Q8X, R212K) and no obvious mesoscale features such as palindromic sequences capable of hairpin formation. This result may be due to the limited number of mutable cytosines and TC motifs in *TK* that confer resistance to ganciclovir (local base composition) and/or to selective pressure. Regardless of the precise molecular explanation, an analogy can be drawn with the mutational spectrum of the *PIK3CA* gene in breast, head/neck, cervical, and others cancers, which has two prominent APOBEC3 mutation hotspots (E542K, E545K) and no obvious hairpin structures [52, 77]. These observations combined to suggest that selective pressure has the potential to overshadow the intrinsic preferences of individual 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, has been 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 and, importantly, rates can vary dramatically between different substrates (e.g., Fig 4 and prior biochemical studies [52, 73, 74]). 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 most mutations are unselected and larger mutation data sets enable broader analyses. For instance, our unbiased WGS analysis of A3A- and A3B-expressing clones indicated that both enzymes are capable of deaminating the ssDNA loop regions of hairpin substrates at similar frequencies (S13 Fig). This result was somewhat anticipated by structural studies where both A3A and a mutant A3B catalytic domain were shown to bind to ssDNA in a U-shaped conformation [16, 29], but it was also unexpected due to recent reports describing strong RNA and DNA hairpin biases for A3A [52, 73]. Additional biochemical and WGS studies will be required to confirm these results as well as extend them to additional cell lines and experimental systems. One potential drawback of the HAP1-TK-M9 system is that an exclusive focus on selected granddaughter clones might overestimate the mutational impact of a given process. However, earlier work also used HAP1 cells to

successfully characterize a variety of different mutation sources, but these studies did not leverage the power of a lethal genetic selection nor did they address the enzyme(s) responsible for APOBEC3 signature mutations [78, 79].

Taken together, the HAP1-TK-M9 studies here have demonstrated unambiguously that both A3A and A3B can inflict an APOBEC3 mutation signature in human genomic DNA with, in both instances, ssDNA deamination events immortalizing predominantly as C-to-T and C-to-G mutations in TCA and TCT trinucleotide motifs (Figs 2 and 3, and S1 Table). Over identical month-long timeframes, A3A causes 4-fold more APOBEC3 signature mutations in comparison to A3B (6070 vs 1528 mutations from 6 and 5 subclone WGSs, respectively). This difference may be explained in part by super-pathological A3A expression levels in the HAP1-TK-M9 system (at least 5-fold higher than levels in breast tumors or cell lines) and in part by the higher intrinsic activity of this enzyme in comparison to A3B. Regardless, both A3A and A3B inflicted thousands of TC-focused APOBEC3 signature mutations, which enabled comparisons between extended intrinsic preferences. Most importantly, A3A has a strong preference for a pyrimidine at the -2 position relative to the target cytosine (72.7% YTCW). This -2 pyrimidine bias mirrors original results from human A3A expression in yeast [50, 51], as well as recent WGS results from human A3A expression in the chicken B cell line DT40 (~70% YTCW [71]) and from human A3A-induced murine hepatocellular carcinomas (70% YTCW [70]). Also, similar to original studies in yeast [50, 51], human A3B showed a slight negative preference for YTCW motifs (47.2% YTCW) and a corresponding enrichment for RTCW motifs (52.8% RTCW; Figs 3B and S9, and S1 Table). In comparison, recent studies showed that human A3B driven tumors in mice (hepatocellular carcinomas and B cell lymphomas) exhibit an opposite APOBEC3 mutation signature bias with 53% YTCW and 47% RTCW [72]. These differences in APOBEC3B local preferences in the different systems may be due to genetic and/or epigenetic factors including but not limited to different base content, chromatin states, DNA repair processes, and/or post-translational regulatory mechanisms. They could also be due to simple stochastic variation and, accordingly, we hypothesize that A3B is non-discriminatory at the -2 position relative to the target cytosine and that this may relate to the amino acid composition of catalytic domain loop 1 residues relative to those of A3A (loop 3, loop 7, and most other active site residues are identical or nearly identical).

Regardless of the precise molecular explanation(s) for differences between experimental systems, the fact that both A3A and A3B can inflict YTCW mutations in human cells helps to inform interpretations of the APOBEC3 enzyme responsible for the overall APOBEC3 mutation program in cancer. For instance, based on comparisons of A3A- and A3B-attributable single base substitution mutation signatures observed here in the HAP1-TK-M9 system and extracted APOBEC3 mutation signatures from 784 breast cancer WGS, it is likely that neither enzyme's preferred motif fully explains the composite signature in most individual tumors (Fig 5 and S1 Table). As regards A3A, the majority of breast tumors have $\leq 72.7\%$ APOBEC3 signature mutations in YTCW motifs, which is the overall preference of A3A here in the HAP1-TK-M9 system suggesting the involvement of at least one other A3 enzyme. As regards A3B, most individual breast tumors have a larger proportion of mutations in YTCW motifs than can be explained by A3B alone ($\geq 47.2\%$). Thus, it is likely that both A3A and A3B contribute to the composite single base substitution mutation signature observed in individual APOBEC3 signature enriched breast tumors. However, we cannot exclude the possibility that some breast tumors may be mutated exclusively by A3A or A3B due to factors listed above, or that some breast tumors may have small mutagenic contributions from other APOBEC3 enzymes. Of course, A3B is not a direct factor in A3B-null breast tumors, although the inherited deletion that removes all A3B coding sequences may dysregulate A3A expression [45, 46]. The ICGC breast tumor data set lacks sufficient A3B-null tumors for a robust SBS signature

analysis but a retrospective analysis of 17 null tumors in the TCGA breast cancer cohort indicates a high proportion of APOBEC3 signature YTCW mutations indicative of (but perhaps not exclusively due to) A3A activity (67% YTCW from a total of $n = 538$ exomic SBS mutations) [46]. Although genomic analyses here were focused on breast cancer, the WGS data sets for A3A- and A3B-expressing HAP1-TK-M9 clones including the extracted mutation signatures may also be useful for comparing with APOBEC3 attributable events in other tumor types.

In addition to selective pressures and mesoscale features, additional factors are likely to influence the APOBEC3-attributable fraction 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 were 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 combined 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.*, Fig 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, 61, 80–85]. *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 a fruitless endeavor (except in A3B-null tumors). Rather, it may be more prudent to focus on developing strategies to simultaneously diagnose and treat the mutagenic contributions of both enzymes.

Independent whole genome sequencing experiments have provided additional information on the APOBEC3 mutation process. Initial studies induced overexpression of A3B in 293-derived cell lines, documented the resulting DNA damage responses, and performed WGS to assess genome-wide associations [86, 87]. However, an unambiguous APOBEC3 mutation signature was difficult to extract from these whole genome sequences due to large numbers of mutations attributable to defective mismatch repair [86, 87]. A more recent study compared *de novo* mutations occurring in APOBEC3 signature positive cell lines during multiple generations of clonal outgrowth [88]. An intriguing finding from this work is that APOBEC3 signature mutations may be able to occur in an episodic manner, accumulating in some generations and not others, consistent with evidence discussed above that A3A and A3B expression can be induced by multiple signal transduction pathways. Episodic mutagenesis, however, is unexpected in cell-based systems in which continuous and relatively stochastic mutagenesis should predominate given defined media and well-controlled growth conditions. These studies were followed-up more recently by WGS comparisons of subclones of the same cancer cell lines CRISPR-engineered to lack A3A, A3B, or both genes [89]. The results of over 250 WGS combined to indicate that A3A may be the source of a significant fraction of observed APOBEC3 signature SBS mutations, A3B a smaller fraction, and another as-yet-undefined APOBEC3 enzyme an additional minor fraction. These data are complementary to

the major results here, with both A3A and A3B proving capable of generating genome-wide APOBEC3 signature single base substitution mutations. Differences in the overall magnitude of A3A vs A3B mutagenesis may be due to cell line selection and factors described above including differential intrinsic activity, protein expression levels, genomic DNA accessibility, cell culture conditions, and importantly durations of mutagenesis. Both studies were necessarily done in model cellular systems, each with obvious strengths, but neither capable of fully recapitulating the wide repertoire of factors that impact the actual pre- and post-transformation environments *in vivo*, including anti-tumor immune responses, which are further likely to vary between different tissue types, tumor types, and patients.

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 [89]. 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 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, 90–96]. 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, the whole genome sequences, and the A3A-specific rabbit monoclonal antibody described here will help to expedite the achievement of these goals.

Materials and methods

Cell lines and culture conditions

All cell lines were cultured at 37°C under 5% CO₂ (Table 1). HAP1 cells (Horizon) and derivatives (this study) were grown in IMDM (Invitrogen) supplemented with 10% fetal bovine serum (Sigma), penicillin (100 U/mL), and streptomycin (100 µg/mL). 293T and HeLa 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 [97]. Puromycin (ThermoFisher) and G418 (Gold Biotechnology) were used at 1 µg/mL and 1 mg/mL, respectively. Ganciclovir (ThermoFisher) was used at 5 µM to select *TK* mutant clones.

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, 98]. Semi-confluent cells in 6 well plates were transfected, treated 24 hrs later with G418 (1 mg/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 Sanger sequencing exon-specific PCR amplicons [99, 100] and further confirmed by WGS (below).

Standard molecular cloning procedures were used to create derivatives of MLV pQCXIP for expressing each A3 protein (Table 1). 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-containing A3 coding sequences [70]. 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 and functional assays as described in the main text.

Each MLV-based construct was co-transfected into 293T cells with appropriate packaging vectors and 48 hrs later the resulting viral supernatants were filtered (0.45 μm) and used to transduce semi-confluent HAP1-TK-M9 cells (MOI < 0.1). After 48 hrs incubation, transduced cells were selected with puromycin (1 $\mu\text{g}/\text{mL}$) and subcloned by limiting dilution to obtain A3 expressing daughter clones. Each daughter clone expressed only a single, integrated construct, which was anticipated by low MOI transduction and verified by mapping insertion sites for 5 representative daughter clones (S8 Fig). 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 5×10^6 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 granddaughter clones were counted using a light microscope and expanded and characterized as described in the main text.

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 (UMN-13, IB: 1:1500, IF: 1:300 [this study]), anti-A3A/B (5210-87-13, IB: 1:1,000, IF: 1:300 [101]), anti-A3H (P1-D8, IB: 1:1000 [58]), anti-Tubulin (Sigma Aldrich, 1:20,000), and anti- β -Actin (Millipore Sigma, mouse mAb, 1:5000 or Cell Signaling Technology, rabbit mAb, 1:5000, as indicated) (Table 1). 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) (Table 1). 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 μ L HED lysis buffer per 1 million cells (25 mM HEPES, 15 mM EDTA, 10% Glycerol, 1 mM DTT, and 1 protease inhibitor tablet [Roche]). Samples were sonicated in a water bath sonicator to ensure complete lysis. A3-containing lysates were incubated at 37°C for 1 hr (A3A and A3B) or 4 hrs (A3H) with purified human UNG2 and a ssDNA substrate containing either a single TCA or a single TCT trinucleotide motif (5'-ATTATTATTATTC AAATGGATTTATTTATTTATTTATTTATTTATTTATTT-FAM [A3A and A3B]; 5'-ATTATTATTA TTCTAATGGATTTATTTATTTATTTATTTATTTATTTATTT-FAM [A3H]) following established protocols [9, 102]. After this initial incubation, the reaction was treated with 100 mM NaOH for 5 min at 95°C. The reaction was run out on a 15% TBE-urea acrylamide gel to separate substrate oligo from cleaved product oligo and imaged on a Typhoon FLA 7000 with ImageQuant TL 8.2.0 software (GE Healthcare).

Recombinant enzyme assays: A3A- and A3B-mycHis were prepared from transfected 293T cells as reported [14, 16, 103–105]. Single hit kinetics were ensured by incubating 25 nM of each protein with 800 nM substrate in reaction buffer (25 mM HEPES pH 7.4, 50 mM NaCl, 5 mM imidazole) for the indicated times at 37°C. Reactions were stopped by freezing in liquid nitrogen and then were heated to 95°C to denature the enzymes. Reactions were then treated with 0.5 U/reaction uracil DNA glycosylase (NEB, USA) for 10 min at 37°C. The resulting abasic sites were cleaved by incubation with 100 mM NaOH and heating to 95°C for 5 min. Products were separated by 20% TBE-Urea PAGE, imaged on a Typhoon FLA-7000 (GE Healthcare, USA), and quantified using ImageQuant TL 8.2.0 software (GE Healthcare). Deamination of the target cytosine was calculated by dividing the total reaction product by the total amount of starting substrate. The oligonucleotide substrates were analyzed by both denaturing and native 20% PAGE to determine the extent of hairpin formation. The oligos (and a previously reported NUP93-noHP oligo) were heated to 70°C and slowly cooled to 37°C in HEPES buffer as above. 1 pmol of each oligo was then mixed with agarose gel loading dye (30% Ficoll 400 in 1x TAE, xylene cyanol, bromophenol blue) or with DNA PAGE loading dye (80% formamide in 1x TBE, xylene cyanol, bromophenol blue) and separated by native and denaturing PAGE, respectively.

DNA content by flow cytometry

Propidium iodide (PI) staining was used to assess the ploidy of HAP1 clones relative to THP1 (ATCC, Cat#: TIB-202) as a confirmed diploid control [106]. Cells were trypsinized and suspended in 100 μ L of 1X PBS per 1 million cells. Then, 500 μ L ice-cold ethanol was added to cell suspensions and incubated at -20°C for 1 hr to fix the cells. After fixation, cells were pelleted and washed in 1X PBS. Cells were finally suspended in 500 μ L of FxCycle PI stain (Invitrogen) and incubated for 30 min at RT in the dark to stain the cells. Cells were spun down and resuspended in 300 μ L of the PI stain solution and placed in a 96 well round bottom plate for flow cytometry analysis using a BD LSRFortessa flow cytometer (with high-throughput 96 well adapter system). A minimum of ten thousand events were acquired for each condition.

RT-qPCR

Total RNA was extracted using the High Pure RNA isolation kit (Roche). cDNA was synthesized using SuperScript First-Strand RT (ThermoFisher). Quantification of mRNA was done using validated primer sets for all human A3 genes relative to the housekeeping gene *TBP* [9, 81, 98, 99]. All RT-qPCR reactions were performed using SsoFast SYBR Green mastermix (Bio-

Rad) in 384-well plates on a LightCycler 480 (Roche) following the manufacturer's protocol. Statistical analyses were done using GraphPad Prism 6 and R.

Immunofluorescent (IF) microscopy

IF microscopy was conducted as described [107, 108]. As a positive control for DNA damage, 4 μ M cisplatin (*cis*-diamminedichloroplatinum II, Selleck Chemical) was incubated with cells for 24 hrs. Cells were grown at low density in 4-chamber, tissue culture-treated glass slides (Falcon) prior to fixation. The cells were then fixed with 4% paraformaldehyde in PBS for 15 min at room temperature. Permeabilization followed using PBS containing 0.2% Triton X-100 (Sigma Aldrich), before rinsing with PBS. The cells were blocked using an IF blocking solution (0.1% triton X-100, 5% goat serum in PBS) for 1 hr at room temperature, then incubated with primary antibody overnight at 4°C. The primary antibodies used were anti-A3A/B/G (5210-87-13, 1:300) [101], anti-A3A (UMN-13, 1:300 [this study]), anti-A3H (Novus, 1:300) [58], and anti- γ -H2AX (JBW301, Millipore Sigma, 1:500). Following this, cells were washed with PBS and incubated with a fluorophore-conjugated secondary antibody for 1 hr at room temperature in the dark. The secondary antibodies used were Alexa Fluor 488 goat anti-mouse (Invitrogen 1:1,000) and Alexa Fluor 594 goat anti-rabbit (Invitrogen, 1:1,000). Both primary and secondary antibodies were diluted in blocking buffer. Hoechst 33342 (Mirus) was used at a final concentration of 1 μ g/mL to stain nuclei for another 15 min, before the slides were washed three times with PBS and mounted using antifade mounting media (Cell Signaling 9071). Images were captured at 60x magnification using a Nikon ECLIPSE Ti2 microscope. The number of γ -H2AX foci per cell were counted for at least 50 cells per condition for each experiment. The nuclear intensity of the γ -H2AX signal and the cellular intensity of each A3 were measured by ImageJ2 software (2.9.0/1.53t). Statistical analyses were conducted in GraphPad Prism 9.

Alkaline comet assays

HeLa cells transfected with individual A3 or control constructs were harvested 24 hrs post-transfection and resuspended in ice-cold 1X phosphate-buffered saline (PBS, Ca²⁺ and Mg²⁺ free, at a density of 10⁵ cells/mL). As a positive control, mock-transfected cells were treated with 2 μ M camptothecin (CPT, Millipore Sigma) for 2 hrs prior to harvesting as above. The CometAssay ESII kit was used for all alkaline Comet assays, following the manufacturer's protocol (BioTechne). Cells were resuspended in low-melt agarose and spread at low density on a glass slide to secure the cells in place for lysis. DNA unwinding and electrophoresis were performed according to the manufacturer's alkaline protocol. Comet tail moments were measured for at least 50 cells per condition using the OpenComet plugin for Image J [109]. Statistical analyses were done using GraphPad Prism 9.

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'-ATCTTGGTGGCGT-GAAACTC and 5'-CTTCCGGTATTGTCTCCTTCC. 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, 5'-GGTCA TGCTGCCATAAAGGTA, 5'-CCGTTCTGGCTCCTCATATC, and 5'-CTTCCGGTATTG TCTCCTTCC).

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 SpeedSeq [110], which relies on the Burrows-Wheeler Aligner, BWA (version 0.7.17). PCR duplicates were removed using Picard (version 2.18.16). Reads were locally realigned around INDELS using GATK3 (version 3.6.0) tools RealignerTargetCreator to create intervals, followed by IndelRealigner on the aligned bam files. Single base substitutions and small INDELS were called in each clone relative to the bam file generated from the HAP1-TK-M9 mother clone using Mutect2 from GATK3 (version 3.6.0). SBSs that passed the internal GATK3 filter with minimum 4 reads supporting each variant, minimum 20 total reads at each variant site and a variant allele frequency over 0.05 were used for downstream analysis. SBSs were analyzed in R (version 4.0.5) using the MutationalPatterns [111] and deconstructSigs R packages (version 1.8.0 [112]). All visualizations were generated using the ggplot2 package (version 3.3.5). The indel landscapes were generated using the MutationalPatterns R package [111] following PCAWG definitions [1]. All individual clone data from each condition were pooled for presentation. **S2 Table** (maf format) provides a tabular list of all single base substitution mutations in the WGS described here.

COSMIC single base substitution mutation signatures (v3 –May 2019 <https://cancer.sanger.ac.uk/cosmic/signatures/SBS/>) were obtained from <https://www.synapse.org/#!/Synapse:syn11738319>. *De-novo* non-negative matrix factorization of mutational signatures was performed with the “extract_signature” command from the MutationalPatterns package, with a rank of 2 and 100 iterations. TCW mutation enrichment scores were calculated as described [50, 65]. Sequence logos of -2 to +2 sequence surrounding C-to-T mutations were created using the ggseqlogo (version 0.1) package. Putative hairpin structures were predicted using the nBMST tool [113] and human genome GRCH38 with a minimum stem length of 6 bases. The loop criteria established by Buisson *et al.* with ssDNA loop lengths of 3 to 11 nucleotides were used to search for APOBEC3 signature mutations in these regions [52].

Mutation clusters (kataegis) were identified using katdetectr [114]. An APOBEC3 kataegis event is so defined if it is comprised of 2 or more APOBEC3 signature mutations within 10 kbp of each other (examples in the context of an intermutation distance plot in **Fig 4C**). In addition, SigProfilerClusters was used to quantify clusters of single base substitutions [115]. This approach for mutation cluster analysis includes all base substitution mutations and, as such, may also include sequencing and bioinformatic errors, DNA polymerase mistakes, and/or other mutagenic processes in addition to APOBEC [1, 116, 117].

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

TCGA primary breast tumors represented by both RNA-seq and whole exome sequencing were downloaded from the Firehose GDAC resource through the Broad Institute pipeline (<http://gdac.broadinstitute.org/>) for multiple tumor tissue types. APOBEC3 mutation signatures were determined as described [5, 65] using the deconstructSigs R package [112]. APOBEC3 mutation enrichment scores were calculated using the hg19 reference genome and published methods [50]. Enrichment score significance was assessed using a Fisher exact test with Benjamini-Hochberg false discovery rate (FDR) correction. All data analyses and visualizations were conducted using R and the ggplot2 package (<https://www.R-project.org/>).

TK integration site determination

To determine the integration site of the single copy *TK-Neo* construct, a *TK* reference sequence was provided as an additional chromosome during alignment of the WGS reads to the reference genome (hg38). Reads that mapped to this region were then categorized as discordant and realigned to hg38 using GRIDSS (v2.2) [118] to determine the site of integration.

Clustering analysis

All primary breast tumor whole genome sequencing variant information from International Cancer Genome Consortium (ICGC) was downloaded from the ICGC data portal (<https://dcc.icgc.org>). SBSs used in these analyses included only C-to-T variants in TC dinucleotide contexts (TCA, TCC, and TCT) and excluded all mutations in CG motifs due to potential overlap with spontaneous water-mediated methyl-C deamination. SBSs meeting these inclusion criteria from all clones expressing A3A, A3B, A3A-E72A, A3B-E255A, and A3H-I were pooled per condition for this analysis. A matrix comprised of the number of mutations within a pentanucleotide across all samples within a cancer type was generated, and counts were normalized to frequency within each cancer type. The resulting matrix was then clustered using the `hclust` function in R with the classical Euclidean distance as the distance method for clustering, which was then plotted as dendrograms. Mutation signatures were calculated using `deconstructSigs` as described above.

Supporting information

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

(XLSX)

S2 Table. Mutation allele frequencies. This mutation allele frequencies (`maf`) file provides a tabular list of all single base substitution mutations in the WGS described here.

(MAF)

S1 Fig. 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 for the KBM7 parent line of HAP1.

(TIF)

S2 Fig. Validation of UMN-13 as a custom rabbit anti-human APOBEC3A mAb. (A) Schematic of human A3A, A3B, and A3G indicating the unique N-terminal epitope used here to generate the 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 with C-terminal HA tags. The blot was probed first with our

custom rabbit anti-human A3A mAb UMN-13 (left) and, after stripping, 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 ($\Delta A-G$), each treated with DMSO as a control or 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), which is absent in the deletion mutant, 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), which are all absent in the deletion mutant. (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 A3A-mCherry signal.

(TIF)

S3 Fig. Additional IF microscopy images of HAP1-TK-M9 cells expressing A3A, A3B, and catalytic mutant derivatives. (A-B) Additional IF-images of HAP1-TK-M9 cells mock-transduced and DMF or cisplatin treated, respectively, and stained as indicated in parallel with cells in panels C-F (scale = 20 μm). (C-F) Additional IF-images of HAP1-TK-M9 cells transduced with the indicated A3 expression constructs and stained with the 5210-87-13 rabbit anti-human A3A/B mAb, γ -H2AX, and Hoechst (scale = 20 μm).

(TIF)

S4 Fig. A3 localization, γ -H2AX staining, and DNA breakage in HeLa cells. (A) IF-images of HeLa cells transfected with the indicated A3 expression constructs and stained for A3A/B (5210-87-13 rabbit anti-human A3A/B mAb) or A3H (Novus ARP10 rabbit anti-human A3H pAb), γ -H2AX, and Hoechst (scale = 20 μm). (B) Quantification of nuclear γ -H2AX staining intensity in the different A3 expressing conditions in panel A ($n > 50$ cells per condition; red bars indicate mean expression levels; p-values by Welch's t-test). (C) Dot plot of nuclear γ -H2AX staining intensity versus A3A staining levels for A3A and A3A-E72A expressing cells shown in panel B. Correlation coefficients (R-values) indicate that A3A expression levels fail to correlate with nuclear γ -H2AX staining intensity. The p-value by Welch's t-test reflects the difference between that A3A and A3A-E72A data sets. (D) Dot plot of nuclear γ -H2AX staining intensity versus A3B staining levels for A3B and A3B-E255A expressing cells shown in panel B. Correlation coefficients (R-values) indicate that A3B but not A3B-E255A expression levels associate with nuclear γ -H2AX staining intensity. The p-value by Welch's t-test reflects the difference between that A3B and A3B-E255A data sets. (E) Representative comets from HeLa cells mock transduced, treated with 2 μM camptothecin (CPT), or transduced with expression constructs for A3A, A3B, or catalytic mutant derivatives (scale = 50 μm). (F) Quantification of tail moment for > 50 cells per condition indicated in panel E (red bars indicate mean tail moment; p-values by one-way ANOVA).

(TIF)

S5 Fig. Mutations in TK derived from ganciclovir-resistant clones. (A) Schematics of individual TK mutations in ganciclovir-resistant clones. T[C>G/T]W mutations are shown in red, other SBSs in black, and INDELs in blue. Composite mutation schematics are shown below for each condition. (B-C) Pentanucleotide logos depicting the -2 and +2 sequence preferences flanking all T(C>T/G)W mutations that accumulated in TK during expression of the indicated constructs.

(TIF)

S6 Fig. 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 $n \geq 2$ clones for each condition; mean \pm 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 ($n = 1093$) for comparison (mean \pm SD). (B) A heatmap depicting mean expression levels of all 7 human *APOBEC3* family members, in addition to *AICDA*, *APOBEC1*, *APOBEC2*, and *APOBEC4*, relative to those of the housekeeping gene *TBP* (RNA-seq values are FKPM; $n \geq 2$ for each condition to provide matching data sets for clones subjected to WGS). Endogenous *A3C* provides a consistent internal control.

(TIF)

S7 Fig. A3 protein expression in granddaughter clones. (A) Immunoblots of *A3A*, *A3B*, and *A3H* in the indicated clones. Tubulin (TUB) is a loading control. (B) Deaminase activity of WCE on ssDNA from the same clones (S, substrate; P, product).

(TIF)

S8 Fig. Representative MLV-A3 insertion sites in HAP1-TK-M9 clones. A table indicating chromosomal locations of representative MLV-A3 insertion sites in HAP-1-TK-M9 clones. In each clone, a single MLV-A3 insertion is positioned in the window between the indicated nucleotides.

(TIF)

S9 Fig. SBS mutation summary. (A-C) Dot plots showing total numbers of SBS mutations and cytosine mutations in NCN and TCW motifs, respectively, in WGS from individual granddaughter clones (p-values using Welch's t-test). (D-E) Dot plots showing total numbers of cytosine mutations in NCN and TCW motifs, respectively, in WGS from individual *A3A* and *A3B* expressing granddaughter clones in comparison to all non-catalytic controls combined (p-values using Welch's t-test).

(TIF)

S10 Fig. Single base substitution mutation profiles of individual granddaughter clones by whole genome sequencing. Trinucleotide profiles of all SBS mutations in WGSs from the indicated granddaughter clones (conditions, clone names, and total SBS numbers are indicated to the right in each profile). Aggregate profiles for *A3A*, *A3B*, and corresponding catalytic mutant controls are shown in Fig 3.

(TIF)

S11 Fig. APOBEC3 mutation signature analysis using an alternative NMF-based workflow. (A) Mutation signature profiles extracted from granddaughter clone WGS using an NMF-based approach. The SBSs of each clone resolved into 2 signatures—an APOBEC3-like mutation 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 (Benjamini-Hochberg false discovery rate-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 mutation profiles of signatures A and B derived using NMF with the former exhibiting an APOBEC3 SBS signature (C-to-T and C-to-G in TCA and TCT motifs).

(TIF)

S12 Fig. Indel landscape of sequenced granddaughter clones. Bar plots showing the percentage of each of the indicated indel types occurring in WGS from HAP1-TK-M9 granddaughter

clones. Total numbers of single T deletions at homopolymers of 6 or more are too numerous to plot on the same axis and are therefore listed here (A3A, 39.5%; A3B, 44.2%; A3A-E72A, 40.6%; A3B-E255A, 40.0%; A3H, 31.8%; and eGFP, 27.4%). The cosine similarity of the indel landscape across all conditions is over 0.96 indicating no significant differences. (TIF)

S13 Fig. Hairpin mutation analysis. A list of the total number of SBS mutations in each HAP1-TK-M9 clone and numbers and frequencies mutations predicted to occur in non-hairpin regions or hairpin loop regions of the genome (see [Methods](#) for additional information). Columns 1 and 2: The A3 or control condition and clone number as listed in [S5 Fig](#). Column 3: The total number of SBS mutations in each clone by WGS (human genome: 3000 mbp). Columns 4 and 5: The total number and frequency of non-hairpin APOBEC3 signature TCW mutations per clone (estimated non-hairpin genomic DNA: 2991.375 mbp). Columns 6 and 7: Total number and frequency of APOBEC3 signature TCW mutations in 3–11 nucleotide loop regions of predicted chromosomal DNA hairpin structures (estimated ssDNA loop region genomic DNA: 8.625 mbp). For A3A data (red boxes) and A3B data (green boxes), the non-hairpin versus hairpin APOBEC3 TCW mutation frequencies are not significantly different ($P = 0.25$ and $P = 0.19$, respectively, by Welch's t-test). APOBEC3 TCW mutation frequencies are also not significantly different between the A3A and A3B data sets (red versus green boxes) for both non-hairpin regions ($P = 0.087$) or predicted ssDNA loop regions of hairpins ($P = 0.149$). However, in comparisons of A3A (red) and A3B (blue) APOBEC3 TCW mutation frequencies in ssDNA loop regions of hairpins and equivalent data sets from aggregate controls (blue boxes: the catalytic mutant of each protein and eGFP), the A3A data set approaches statistical significance ($P = 0.0655$) and the A3B data are significantly different ($P = 0.0367$) by Welch's t-test. (TIF)

S14 Fig. Validation of *NUP93* and *SDHB* DNA hairpins and linear structures. Native PAGE (left) and denaturing PAGE (right) analysis of the indicated oligonucleotide substrates. The hairpin substrates migrate faster under native conditions, and their mobility is similar to the linear derivatives under denaturing conditions. The only oligonucleotide not used in biochemical experiments in [Fig 4](#) is the *NUP93*-noHP (no hairpin), which has half of the stem replaced by adenines (5'-6-carboxyfluorescein-GCAAGCTGTTCAAAAAATGA) and is included here as an additional control. (TIF)

S15 Fig. Single base substitution mutation signatures and distributions in A3A, A3B, and eGFP expressing clones. (A) Trinucleotide profiles of total, clustered, non-clustered single base substitution mutations in representative A3A (top), A3B (middle), and eGFP (bottom) clones. The intermutation distance (IMD) is indicated, together with the expected number number of kataegis events and the actual range of kataegis events observed for clones of each condition determined using SigProfilerClusters (Materials and Methods). (B) IMDs of A3A (top), A3B (middle), and eGFP (bottom) clones. Green lines are based on actual intermutation distances, and red lines are simulated distributions of intermutation distances. 95% confidence intervals are shown in pink for the simulated IMD distributions. (TIF)

Acknowledgments

We thank Justin Leung, Eloise Dray, Weixing Zhao, and Wenjing Li for expert assistance with IF microscopy and comet assays, Arad Moghadasi and Sofia Moraes for initial technical

support, and Kyle Richards for helping to set-up spectral karyotype analysis. We also thank the UMN Cytogenetics Core for spectral karyotyping and Scott McIvor for sharing a plasmid construct with the *MND* promoter.

Author Contributions

Conceptualization: Michael A. Carpenter, Nuri A. Temiz, Matthew C. Jarvis, Reuben S. Harris.

Data curation: Nuri A. Temiz, Mahmoud A. Ibrahim, Matthew C. Jarvis.

Formal analysis: Michael A. Carpenter, Nuri A. Temiz, Mahmoud A. Ibrahim, Matthew C. Jarvis, Gabriel J. Starrett, Douglas Yee, Reuben S. Harris.

Funding acquisition: Matthew C. Jarvis, Douglas Yee, Reuben S. Harris.

Investigation: Michael A. Carpenter, Nuri A. Temiz, Mahmoud A. Ibrahim, Matthew C. Jarvis, Margaret R. Brown, Prokopios P. Argyris, William L. Brown.

Methodology: Michael A. Carpenter, Nuri A. Temiz, Mahmoud A. Ibrahim, Matthew C. Jarvis, Margaret R. Brown, Prokopios P. Argyris, William L. Brown, Reuben S. Harris.

Project administration: Douglas Yee, Reuben S. Harris.

Resources: Reuben S. Harris.

Supervision: Michael A. Carpenter, Prokopios P. Argyris, Douglas Yee, Reuben S. Harris.

Validation: Michael A. Carpenter, Nuri A. Temiz, Mahmoud A. Ibrahim, Margaret R. Brown, Prokopios P. Argyris, William L. Brown.

Visualization: Michael A. Carpenter, Nuri A. Temiz, Mahmoud A. Ibrahim, Matthew C. Jarvis, Reuben S. Harris.

Writing – original draft: Matthew C. Jarvis, Reuben S. Harris.

Writing – review & editing: Michael A. Carpenter, Nuri A. Temiz, Mahmoud A. Ibrahim, Margaret R. Brown, Prokopios P. Argyris, William L. Brown, Gabriel J. Starrett, Douglas Yee, Reuben S. Harris.

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. <https://doi.org/10.1038/s41586-020-1943-3> PMID: 32025018; PubMed Central PMCID: PMC7054213.
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*. 2018; 29(3):563–72. <https://doi.org/10.1093/annonc/mdy003> 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. <https://doi.org/10.1016/j.dnarep.2020.102905> PMID: 32818816; PubMed Central PMCID: PMC7494591.
4. Koh G, Degasperis 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. <https://doi.org/10.1038/s41568-021-00377-7> 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. Epub 2013/08/16. <https://doi.org/10.1038/nature12477> PMID: 23945592.
6. Roberts SA, Lawrence MS, Klimczak LJ, Grimm SA, Fargo D, Stojanov P, et al. An APOBEC cytidine deaminase mutagenesis pattern is widespread in human cancers. *Nat Genet*. 2013; 45(9):970–6. <https://doi.org/10.1038/ng.2702> PMID: 23852170; PubMed Central PMCID: PMC3789062.

7. Nik-Zainal S, Alexandrov LB, Wedge DC, Van Loo P, Greenman CD, Raine K, et al. Mutational processes molding the genomes of 21 breast cancers. *Cell*. 2012; 149(5):979–93. Epub 2012/05/23. <https://doi.org/10.1016/j.cell.2012.04.024> PMID: 22608084.
8. Nik-Zainal S, Davies H, Staaf J, Ramakrishna M, Glodzik D, Zou X, et al. Landscape of somatic mutations in 560 breast cancer whole-genome sequences. *Nature*. 2016; 534(7605):47–54. <https://doi.org/10.1038/nature17676> PMID: 27135926; PubMed Central PMCID: PMC4910866.
9. Burns MB, Lackey L, Carpenter MA, Rathore A, Land AM, Leonard B, et al. APOBEC3B is an enzymatic source of mutation in breast cancer. *Nature*. 2013; 494(7437):366–70. Epub 2013/02/08. <https://doi.org/10.1038/nature11881> PMID: 23389445.
10. Leonard B, Hart SN, Burns MB, Carpenter MA, Temiz NA, Rathore A, et al. APOBEC3B upregulation and genomic mutation patterns in serous ovarian carcinoma. *Cancer Res*. 2013; 73(24):7222–31. Epub 2013/10/25. <https://doi.org/10.1158/0008-5472.CAN-13-1753> PMID: 24154874; PubMed Central PMCID: PMC3867573.
11. Ito F, Fu Y, Kao SA, Yang H, Chen XS. Family-wide comparative analysis of cytidine and methylcytosine deamination by eleven human APOBEC proteins. *J Mol Biol*. 2017; 429(12):1787–99. <https://doi.org/10.1016/j.jmb.2017.04.021> PMID: 28479091; 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. <https://doi.org/10.1038/s41598-018-21955-0> 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. <https://doi.org/10.1128/JVI.05238-11> 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. <https://doi.org/10.1074/jbc.M112.385161> M112.385161 [pii]. 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 cytosine deaminase. *PloS one*. 2013; 8(6):e63461. Epub 2013/07/11. <https://doi.org/10.1371/journal.pone.0063461> 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. <https://doi.org/10.1038/nsm.3344> 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. <https://doi.org/10.1093/nar/gkx345> PMID: 28472485; PubMed Central PMCID: PMC5570014.
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. <https://doi.org/10.1038/ng.3441> PMID: 26551669; PubMed Central PMCID: PMC4783858.
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. <https://doi.org/10.1038/ng.2701> 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. <https://doi.org/10.1371/journal.pone.0171343> 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. <https://doi.org/10.1007/s12672-014-0196-8> 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. <https://doi.org/10.1038/s41588-019-0507-7> PMID: 31570896; PubMed Central PMCID: PMC6858873.
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. <https://doi.org/10.1038/s41586-021-03822-7> PMID: 34433962.

24. Harris RS, Dudley JP. APOBECs and virus restriction. *Virology*. 2015; 479-480C:131–45. <https://doi.org/10.1016/j.virol.2015.03.012> 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. <https://doi.org/10.1038/ni.3156> PMID: 25988886.
26. Silvas TV, Schiffer CA. APOBEC3s: DNA-editing human cytidine deaminases. *Protein Sci*. 2019; 28(9):1552–66. Epub 2019/06/27. <https://doi.org/10.1002/pro.3670> PMID: 31241202; PubMed Central PMCID: PMC6699113.
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. <https://doi.org/10.1371/journal.pgen.1008545> 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. <https://doi.org/10.1128/JVI.01318-17> PMID: 29367246; PubMed Central PMCID: PMC5972886.
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. <https://doi.org/10.1038/ncomms15024> PMID: 28452355; PubMed Central PMCID: PMC5414352.
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. <https://doi.org/10.1080/15384101.2016.1152426> PMID: 26918916.
31. Chen H, Lilley CE, Yu Q, Lee DV, Chou J, Narvaiza I, et al. APOBEC3A is a potent inhibitor of adeno-associated virus and retrotransposons. *Curr Biol*. 2006; 16(5):480–5. <https://doi.org/10.1016/j.cub.2006.01.031> PMID: 16527742.
32. Faden DL, Thomas S, Cantalupo PG, Agrawal N, Myers J, DeRisi J. Multi-modality analysis supports APOBEC as a major source of mutations in head and neck squamous cell carcinoma. *Oral Oncology*. 2017; 74:8–14. <https://doi.org/10.1016/j.oraloncology.2017.09.002> PMID: 29103756.
33. Warren CJ, Xu T, Guo K, Griffin LM, Westrich JA, Lee D, et al. APOBEC3A functions as a restriction factor of human papillomavirus. *J Virol*. 2015; 89(1):688–702. <https://doi.org/10.1128/JVI.02383-14> PMID: 25355878; PubMed Central PMCID: PMC4301161.
34. Taylor BJ, Nik-Zainal S, Wu YL, Stebbings LA, Raine K, Campbell PJ, et al. DNA deaminases induce break-associated mutation showers with implication of APOBEC3B and 3A in breast cancer kataegis. *Elife*. 2013; 2:e00534. Epub 2013/04/20. <https://doi.org/10.7554/eLife.00534> PMID: 23599896; PubMed Central PMCID: PMC3628087.
35. Verhalen B, Starrett GJ, Harris RS, Jiang M. Functional upregulation of the DNA cytosine deaminase APOBEC3B by polyomaviruses. *J Virol*. 2016; 90(14):6379–86. <https://doi.org/10.1128/JVI.00771-16> PMID: 27147740.
36. Vieira VC, Leonard B, White EA, Starrett GJ, Temiz NA, Lorenz LD, et al. Human papillomavirus E6 triggers upregulation of the antiviral and cancer genomic DNA deaminase APOBEC3B. *MBio*. 2014; 5(6). <https://doi.org/10.1128/mBio.02234-14> PMID: 25538195; PubMed Central PMCID: PMC4278539.
37. Starrett GJ, Serebrenik AA, Roelofs PA, McCann JL, Verhalen B, Jarvis MC, et al. Polyomavirus T antigen induces APOBEC3B expression using an LXCXE-dependent and TP53-independent mechanism. *MBio*. 2019; 10(1). <https://doi.org/10.1128/mBio.02690-18> PMID: 30723127.
38. Walker BA, Wardell CP, Murison A, Boyle EM, Begum DB, Dahir NM, et al. APOBEC family mutational signatures are associated with poor prognosis translocations in multiple myeloma. *Nat Commun*. 2015; 6:6997. <https://doi.org/10.1038/ncomms7997> PMID: 25904160; PubMed Central PMCID: PMC4568299.
39. Roelofs PA, Goh CY, Chua BH, Jarvis MC, Stewart TA, McCann JL, et al. Characterization of the mechanism by which the RB/E2F pathway controls expression of the cancer genomic DNA deaminase APOBEC3B. *Elife*. 2020; 9. Epub 2020/09/29. <https://doi.org/10.7554/eLife.61287> PMID: 32985974; PubMed Central PMCID: PMC7553775.
40. Salamango DJ, McCann JL, Demir O, Brown WL, Amaro RE, Harris RS. APOBEC3B nuclear localization requires two distinct N-terminal domain surfaces. *J Mol Biol*. 2018; 430(17):2695–708. <https://doi.org/10.1016/j.jmb.2018.04.044> PMID: 29787764.
41. de Bruin EC, McGranahan N, Mitter R, Salm M, Wedge DC, Yates L, et al. Spatial and temporal diversity in genomic instability processes defines lung cancer evolution. *Science*. 2014; 346(6206):251–6. <https://doi.org/10.1126/science.1253462> PMID: 25301630.
42. Argyris PP, Wilkinson PE, Jarvis MC, Magliocca KR, Patel MR, Vogel RI, et al. Endogenous APOBEC3B overexpression characterizes HPV-positive and HPV-negative oral epithelial dysplasias and

- head and neck cancers. *Modern Pathology*. Epub 2020/07/08. <https://doi.org/10.1038/s41379-020-0617-x> PMID: 32632179.
43. Yamazaki H, Shirakawa K, Matsumoto T, Hirabayashi S, Murakawa Y, Kobayashi M, et al. Endogenous APOBEC3B overexpression constitutively generates DNA substitutions and deletions in myeloma cells. *Sci Rep*. 2019; 9(1):7122. Epub 2019/05/11. <https://doi.org/10.1038/s41598-019-43575-y> PMID: 31073151; PubMed Central PMCID: PMC6509214.
 44. Yamazaki H, Shirakawa K, Matsumoto T, Kazuma Y, Matsui H, Horisawa Y, et al. APOBEC3B reporter myeloma cell lines identify DNA damage response pathways leading to APOBEC3B expression. *PLoS One*. 2020; 15(1):e0223463. Epub 2020/01/09. <https://doi.org/10.1371/journal.pone.0223463> PMID: 31914134; PubMed Central PMCID: PMC6948746.
 45. Nik-Zainal S, Wedge DC, Alexandrov LB, Petrijak M, Butler AP, Bolli N, et al. Association of a germline copy number polymorphism of APOBEC3A and APOBEC3B with burden of putative APOBEC-dependent mutations in breast cancer. *Nat Genet*. 2014; 46(5):487–91. <https://doi.org/10.1038/ng.2955> PMID: 24728294; PubMed Central PMCID: PMC4137149.
 46. Starrett GJ, Luengas EM, McCann JL, Ebrahimi D, Temiz NA, Love RP, et al. The DNA cytosine deaminase APOBEC3H haplotype I likely contributes to breast and lung cancer mutagenesis. *Nat Commun*. 2016; 7:12918. <https://doi.org/10.1038/ncomms12918> PMID: 27650891.
 47. Chen TW, Lee CC, Liu H, Wu CS, Pickering CR, Huang PJ, et al. APOBEC3A is an oral cancer prognostic biomarker in Taiwanese carriers of an APOBEC deletion polymorphism. *Nat Commun*. 2017; 8(1):465. <https://doi.org/10.1038/s41467-017-00493-9> PMID: 28878238; PubMed Central PMCID: PMC5587710.
 48. Hix MA, Wong L, Flath B, Chelico L, Cisneros GA. Single-nucleotide polymorphism of the DNA cytosine deaminase APOBEC3H haplotype I leads to enzyme destabilization and correlates with lung cancer. *NAR Cancer*. 2020; 2(3):zcaa023. Epub 2020/09/29. <https://doi.org/10.1093/narcan/zcaa023> PMID: 32984821; PubMed Central PMCID: PMC7503452.
 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. <https://doi.org/10.1074/jbc.M115.666065> 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. <https://doi.org/10.1038/ng.3378> 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. <https://doi.org/10.1016/j.celrep.2016.01.021> 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. <https://doi.org/10.1126/science.aaw2872> PMID: 31249028; PubMed Central PMCID: PMC6731024.
 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. <https://doi.org/10.1038/nature00862> 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. [https://doi.org/10.1016/s1097-2765\(02\)00742-6](https://doi.org/10.1016/s1097-2765(02)00742-6) 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. <https://doi.org/10.1186/1471-2172-6-10> 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. <https://doi.org/10.1073/pnas.0501694102> 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. <https://doi.org/10.1371/journal.pgen.1003149> 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. <https://doi.org/10.1371/journal.pgen.1004761> 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. <https://doi.org/10.1038/s41467-018-06594-3> 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. <https://doi.org/10.3390/v12040378> PMID: 32235597; PubMed Central PMCID: PMC7232234.
 61. 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. <https://doi.org/10.1016/j.celrep.2015.08.066> PMID: 26411678; PubMed Central PMCID: PMC4597099.
 62. Green AM, Budagyan K, Hayer KE, Reed MA, Savani MR, Wertheim GB, et al. Cytosine deaminase APOBEC3A sensitizes leukemia cells to inhibition of the DNA replication checkpoint. *Cancer Res.* 2017; 77(17):4579–88. <https://doi.org/10.1158/0008-5472.CAN-16-3394> PMID: 28655787; PubMed Central PMCID: PMC5581702.
 63. Landry S, Narvaiza I, Linfesty DC, Weitzman MD. APOBEC3A can activate the DNA damage response and cause cell-cycle arrest. *EMBO Rep.* 2011; 12(5):444–50. Epub 2011/04/05. <https://doi.org/10.1038/embor.2011.46> PMID: 21460793.
 64. Land AM, Law EK, Carpenter MA, Lackey L, Brown WL, Harris RS. Endogenous APOBEC3A DNA cytosine deaminase is cytoplasmic and nongenotoxic. *J Biol Chem.* 2013; 288(24):17253–60. Epub 2013/05/04. <https://doi.org/10.1074/jbc.M113.458661> PMID: 23640892; PubMed Central PMCID: PMC3682529.
 65. Jarvis MC, Ebrahimi D, Temiz NA, Harris RS. Mutation signatures including APOBEC in cancer cell lines. *JNCI Cancer Spectr.* 2018; 2(1). <https://doi.org/10.1093/jncics/pky002> PMID: 29888758; PubMed Central PMCID: PMC5993214.
 66. 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. <https://doi.org/10.1016/j.celrep.2012.12.008> PMID: 23318258; PubMed Central PMCID: PMC3588146.
 67. Fang H, Yan HHN, Bilardi RA, Flensburg C, Yang H, Barbour JA, et al. Ganciclovir-induced mutations are present in a diverse spectrum of post-transplant malignancies. *Genome Medicine.* 2022; 14(1):124. Epub 2022/10/31. <https://doi.org/10.1186/s13073-022-01131-w> PMID: 36316687; PubMed Central PMCID: PMC9620652.
 68. Starrett GJ, Yu K, Golubeva Y, Lenz P, Piaskowski ML, Petersen D, et al. Evidence for virus-mediated oncogenesis in bladder cancers arising in solid organ transplant recipients. *Elife.* 2023; 12. Epub 2023/03/24. <https://doi.org/10.7554/eLife.82690> PMID: 36961501; PubMed Central PMCID: PMC10446826.
 69. de Kanter JK, Peci F, Bertrums E, Rosendahl Huber A, van Leeuwen A, van Roosmalen MJ, et al. Antiviral treatment causes a unique mutational signature in cancers of transplantation recipients. *Cell Stem Cell.* 2021; 28(10):1726–39 e6. Epub 2021/09/07. <https://doi.org/10.1016/j.stem.2021.07.012> PMID: 34496298; PubMed Central PMCID: PMC8516432.
 70. 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. <https://doi.org/10.1084/jem.20200261> PMID: 32870257.
 71. 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. <https://doi.org/10.1016/j.celrep.2022.110555> PMID: 35320711.
 72. Durfee C, Temiz NA, Levin-Klein R, Argyris PP, Alsoe L, Carracedo S, et al. Human APOBEC3B promotes tumor development in vivo including signature mutations and metastases. *Cell Rep Med.* 2023; 4(10):101211. Epub 2023/10/04. <https://doi.org/10.1016/j.xcrm.2023.101211> PMID: 37797615; PubMed Central PMCID: PMC10591044.
 73. Jalili P, Bowen D, Langenbacher 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. <https://doi.org/10.1038/s41467-020-16802-8> PMID: 32532990; PubMed Central PMCID: PMC7293259.
 74. Langenbacher 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. <https://doi.org/10.1038/s41467-021-21891-0> PMID: 33707442; PubMed Central PMCID: PMC7952602.

75. 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. <https://doi.org/10.1038/ncomms7881> PMID: 25898173; PubMed Central PMCID: PMC4411297.
76. 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. <https://doi.org/10.1016/j.molcel.2012.03.030> PMID: 22607975; PubMed Central PMCID: PMC3361558.
77. 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. <https://doi.org/10.1016/j.celrep.2014.05.012> PMID: 24910434.
78. 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. <https://doi.org/10.1038/s43018-021-00200-0> PMID: 34164627; PubMed Central PMCID: PMC7611045.
79. 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. <https://doi.org/10.1038/s41467-018-04052-8> PMID: 29717121; PubMed Central PMCID: PMC5931590.
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). <https://doi.org/10.3390/v9080233> 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. <https://doi.org/10.1158/0008-5472.CAN-15-2171-T> PMID: 26420215; PubMed Central PMCID: PMC4631676.
82. 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. <https://doi.org/10.1093/nar/gkx721> PMID: 28977491; PubMed Central PMCID: PMC5737468.
83. 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). <https://doi.org/10.1126/scitranslmed.aas9668> PMID: 30135250.
84. 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. <https://doi.org/10.1038/s41467-021-25203-4> PMID: 34389714; PubMed Central PMCID: PMC8363607.
85. 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. <https://doi.org/10.1016/j.bbrc.2016.08.148> PMID: 27577680.
86. Akre MK, Starrett GJ, Quist JS, Temiz NA, Carpenter MA, Tutt AN, et al. Mutation processes in 293-based clones overexpressing the DNA cytosine deaminase APOBEC3B. *PLoS One.* 2016; 11(5): e0155391. <https://doi.org/10.1371/journal.pone.0155391> PMID: 27163364; PubMed Central PMCID: PMC4862684.
87. 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. <https://doi.org/10.1038/bjc.2017.133> PMID: 28535155; PubMed Central PMCID: PMC5520199.
88. 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. <https://doi.org/10.1016/j.cell.2019.02.012> PMID: 30849372; PubMed Central PMCID: PMC6424819.
89. 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. <https://doi.org/10.1038/s41586-022-04972-y> PMID: 35859169; PubMed Central PMCID: PMC9329121.
90. 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. <https://doi.org/10.1126/sciadv.1601737> PMID: 27730215.
91. 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. <https://doi.org/10.1186/s12920-019-0579-3> PMID: 31533728; PubMed Central PMCID: PMC6751822.
92. 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. <https://doi.org/10.18632/oncotarget.23344> PMID: 29435122; PubMed Central PMCID: PMC5796993.
 93. 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. <https://doi.org/10.1038/ng.3670> PMID: 27643540; PubMed Central PMCID: PMC6583788.
 94. 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. <https://doi.org/10.1038/s41388-018-0657-6> PMID: 30647454; PubMed Central PMCID: PMC6499643.
 95. 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. <https://doi.org/10.1038/s41467-020-14568-7> PMID: 32034147; PubMed Central PMCID: PMC7005822.
 96. 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. <https://doi.org/10.1007/s12282-015-0641-8> PMID: 26476745.
 97. Uphoff CC, Drexler HG. Detecting mycoplasma contamination in cell cultures by polymerase chain reaction. *Methods Mol Biol*. 2011; 731:93–103. Epub 2011/04/26. https://doi.org/10.1007/978-1-61779-080-5_8 PMID: 21516400.
 98. Stenglein MD, Burns MB, Li M, Lengyel J, Harris RS. APOBEC3 proteins mediate the clearance of foreign DNA from human cells. *Nat Struct Mol Biol*. 2010; 17(2):222–9. <https://doi.org/10.1038/nsmb.1744> PMID: 20062055.
 99. Refsland EW, Stenglein MD, Shindo K, Albin JS, Brown WL, Harris RS. Quantitative profiling of the full APOBEC3 mRNA repertoire in lymphocytes and tissues: implications for HIV-1 restriction. *Nucleic Acids Res*. 2010; 38:4274–84. <https://doi.org/10.1093/nar/gkq174> PMID: 20308164.
 100. Wang J, Becker JT, Shi K, Lauer KV, Salamango DJ, Aihara H, et al. The role of RNA in HIV-1 Vif-mediated degradation of APOBEC3H. *J Mol Biol*. 2019; 431(24):5019–31. Epub 2019/10/20. <https://doi.org/10.1016/j.jmb.2019.09.014> PMID: 31628948; PubMed Central PMCID: PMC6948013.
 101. Brown WL, Law EK, Argyris PP, Carpenter MA, Levin-Klein R, Ranum AN, et al. A rabbit monoclonal antibody against the antiviral and cancer genomic DNA mutating enzyme APOBEC3B. *Antibodies*. 2019; 8(3). Epub 2019/09/24. <https://doi.org/10.3390/antib8030047> PMID: 31544853.
 102. Lackey L, Demorest ZL, Land AM, Hultquist JF, Brown WL, Harris RS. APOBEC3B and AID have similar nuclear import mechanisms. *J Mol Biol*. 2012; 419(5):301–14. Epub 2012/03/27. <https://doi.org/10.1016/j.jmb.2012.03.011> PMID: 22446380; PubMed Central PMCID: PMC3368237.
 103. Shi K, Carpenter MA, Kurahashi K, Harris RS, Aihara H. Crystal structure of the DNA deaminase APOBEC3B catalytic domain. *J Biol Chem*. 2015; 290(47):28120–30. <https://doi.org/10.1074/jbc.M115.679951> PMID: 26416889; PubMed Central PMCID: PMC4653671.
 104. Shaban NM, Yan R, Shi K, Moraes SN, Cheng AZ, Carpenter MA, et al. Cryo-EM structure of the EBV ribonucleotide reductase BORF2 and mechanism of APOBEC3B inhibition. *Science Advances*. 2022; 8:eabm2827. <https://doi.org/10.1126/sciadv.abm2827> PMID: 35476445
 105. McCann JL, Cristini A, Law EK, Lee SY, Tellier M, Carpenter MA, et al. APOBEC3B regulates R-loops and promotes transcription-associated mutagenesis in cancer. *Nat Genet*. 2023; 55(10):1721–34. Epub 2023/09/21. <https://doi.org/10.1038/s41588-023-01504-w> PMID: 37735199; PubMed Central PMCID: PMC10562255.
 106. Ikeda T, Shimizu R, Nasser H, Carpenter MA, Cheng AZ, Brown WL, et al. APOBEC3 degradation is the primary function of HIV-1 Vif for virus replication in the myeloid cell line THP-1. *bioRxiv*. 2023. Epub 2023/03/29. <https://doi.org/10.1101/2023.03.28.534666> PMID: 37034786; PubMed Central PMCID: PMC10081227.
 107. McNamara RP, McCann JL, Gudipaty SA, D'Orso I. Transcription factors mediate the enzymatic disassembly of promoter-bound 7SK snRNP to locally recruit P-TEFb for transcription elongation. *Cell Rep*. 2013; 5(5):1256–68. Epub 2013/12/10. <https://doi.org/10.1016/j.celrep.2013.11.003> PMID: 24316072; PubMed Central PMCID: PMC3882317.
 108. Lackey L, Law EK, Brown WL, Harris RS. Subcellular localization of the APOBEC3 proteins during mitosis and implications for genomic DNA deamination. *Cell Cycle*. 2013; 12(5):762–72. Epub 2013/

- 02/08. <https://doi.org/10.4161/cc.23713> [pii]. PMID: 23388464; PubMed Central PMCID: PMC3610724.
109. Gyori BM, Venkatachalam G, Thiagarajan PS, Hsu D, Clement MV. OpenComet: an automated tool for comet assay image analysis. *Redox Biol.* 2014; 2:457–65. Epub 2014/03/14. <https://doi.org/10.1016/j.redox.2013.12.020> PMID: 24624335; PubMed Central PMCID: PMC3949099.
 110. Chiang C, Layer RM, Faust GG, Lindberg MR, Rose DB, Garrison EP, et al. SpeedSeq: ultra-fast personal genome analysis and interpretation. *Nat Methods.* 2015; 12(10):966–8. Epub 2015/08/11. <https://doi.org/10.1038/nmeth.3505> PMID: 26258291; PubMed Central PMCID: PMC4589466.
 111. Blokzijl F, Janssen R, van Boxtel R, Cuppen E. MutationalPatterns: comprehensive genome-wide analysis of mutational processes. *Genome Medicine.* 2018; 10(1):33. Epub 2018/04/27. <https://doi.org/10.1186/s13073-018-0539-0> PMID: 29695279; PubMed Central PMCID: PMC5922316.
 112. Rosenthal R, McGranahan N, Herrero J, Taylor BS, Swanton C. DeconstructSigs: delineating mutational processes in single tumors distinguishes DNA repair deficiencies and patterns of carcinoma evolution. *Genome Biol.* 2016; 17:31. Epub 2016/02/22. <https://doi.org/10.1186/s13059-016-0893-4> PMID: 26899170; PubMed Central PMCID: PMC4762164.
 113. Cer RZ, Bruce KH, Donohue DE, Temiz NA, Mudunuri US, Yi M, et al. Searching for non-B DNA-forming motifs using nBMST (non-B DNA motif search tool). *Curr Protoc Hum Genet.* 2012; Chapter 18: Unit 18 7 1–22. <https://doi.org/10.1002/0471142905.hg1807s73> PMID: 22470144; PubMed Central PMCID: PMC3350812.
 114. Hazelaar DM, van Riet J, Hoogstrate Y, van de Werken HJG. Katdetect: an R/bioconductor package utilizing unsupervised changepoint analysis for robust kataegis detection. *Gigascience.* 2022; 12. <https://doi.org/10.1093/gigascience/giad081> PMID: 37848617; PubMed Central PMCID: PMC10580377.
 115. Bergstrom EN, Kundu M, Tbeileh N, Alexandrov LB. Examining clustered somatic mutations with Sig-ProfilerClusters. *Bioinformatics.* 2022; 38(13):3470–3. <https://doi.org/10.1093/bioinformatics/btac335> PMID: 35595234; PubMed Central PMCID: PMC9237733.
 116. Consortium ITP-CAoWG. Pan-cancer analysis of whole genomes. *Nature.* 2020; 578(7793):82–93. Epub 2020/02/05. <https://doi.org/10.1038/s41586-020-1969-6> PMID: 32025007; PubMed Central PMCID: PMC7025898.
 117. Wang Q, Pierce-Hoffman E, Cummings BB, Alfoldi J, Francioli LC, Gauthier LD, et al. Landscape of multi-nucleotide variants in 125,748 human exomes and 15,708 genomes. *Nat Commun.* 2020; 11(1):2539. Epub 2020/05/27. <https://doi.org/10.1038/s41467-019-12438-5> PMID: 32461613; PubMed Central PMCID: PMC7253413.
 118. Cameron DL, Schroder J, Penington JS, Do H, Molania R, Dobrovic A, et al. GRIDSS: sensitive and specific genomic rearrangement detection using positional de Bruijn graph assembly. *Genome Res.* 2017; 27(12):2050–60. Epub 2017/11/04. <https://doi.org/10.1101/gr.222109.117> PMID: 29097403; PubMed Central PMCID: PMC5741059.