bioRxiv preprint doi: https://doi.org/10.1101/2020.02.28.970509; this version posted March 2, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

Induced mutagenesis by the DNA cytosine deaminase APOBEC3H Haplotype I protects against lung cancer

Mark A. Hix[1,#], Lai Wong[2,#], Ben Flath[2], Linda Chelico[2,*], G. Andrés Cisneros[1,3,*]

[1] Department of Chemistry, University of North Texas, Denton, TX 76201, USA

[2] Department of Biochemistry, Microbiology, and Immunology, University of Saskatchewan, Saskatoon, SK, S7N 5E5, Canada

[3] Center for Advanced Scientific Computing and Modeling (CASCaM), University of North Texas, Denton, TX 76201, USA

[#] Authors contributed equally

*To whom correspondence should be addressed: linda.chelico@usask.ca and andres@unt.edu

Summary (200 words)

The DNA cytosine deaminases APOBEC3A, APOBEC3B, and APOBEC3H Haplotype I can induce mutations in cells that lead to cancer evolution. The database of cancer biomarkers in DNA repair genes (DNArCdb) identified a single nucleotide polymorphism (rs139298) of APOBEC3H Haplotype I that is involved in lung cancer[1]. Here, we show that this single nucleotide polymorphism causes the destabilization of APOBEC3H Haplotype I. Computational analysis suggests that the resulting K121E change affects the structure of APOBEC3H leading to active site disruption and destabilization of the RNA-mediated dimer interface. A K117E mutation in a K121E background stabilized the APOBEC3H Haplotype I, enabled biochemical study, and showed that the K121E affected catalytic activity, single-stranded DNA binding, and oligomerization on single-stranded DNA. That the destabilization of a DNA mutator would be associated with lung cancer suggests that

bioRxiv preprint doi: https://doi.org/10.1101/2020.02.28.970509; this version posted March 2, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

too much mutation could result in immune recognition or death of tumor cells, suggesting that multiple APOBEC3s would not be expressed in the same tumor cells. In support of this hypothesis, stably expressed APOBEC3H Haplotype I caused a high amount of double-stranded DNA breaks in a lung cancer cell line that endogenously expressed APOBEC3B. Altogether, the data support the model that high APOBEC3 mutations in tumors are protective.

Introduction

The eleven member Apolipoprotein B editing enzyme (APOBEC) family in humans are comprised of RNA and single-stranded (ss) DNA cytosine deaminases with diverse biological functions[2-4]. Within this family are a subgroup of seven APOBEC3 family members that primarily deaminate cytosine in ssDNA, which forms promutagenic uracil. This C-to-U conversion is considered a DNA lesion and can result in a C-to-T mutation if uracil is used as a template during DNA replication or error prone uracil repair can nucleate other mutations, such as C-to-G transversions[5]. These enzymes are known for their ability to restrict the replication of various viruses, such as HIV-1, Epstein Barr Virus, and Hepatitis B virus, through the mutagenic fates of uracil in DNA[6-8]. The host genomic DNA is usually safe from cytosine deamination due to various protections such as cytoplasmic localization, low expression levels, or cytoplasmic RNA binding which both sequesters and inhibits enzyme activity[9-12]. However, when these checks are not in place, the APOBEC3 enzymes that can enter the nucleus, such as APOBEC3B (A3B), APOBEC3A (A3A), and APOBEC3H haplotype I (A3H Hap I)[9,13-18], can deaminate genomic DNA. This activity has linked these APOBEC3 enzymes to various cancers, such as lung, breast, bladder, cervical, and others. The activity related to cancer is on transient genomic ssDNA during replication or transcription and is

bioRxiv preprint doi: https://doi.org/10.1101/2020.02.28.970509; this version posted March 2, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

considered "off-target" activity[19-23]. Notably, the C-to-U conversion only occurs in a specific sequence context, e.g., 5' RTCA (A3B, "R" is A or G), 5' YTCA (A3A, "Y" is C or T), 5'TCT (A3H)[13,17,18]. This 5'TC sequence context has enabled the APOBEC3 mutation signature in at least 16 cancers and can be differentiated from chemical DNA damage, which may occur with lung cancer, for example[5,24]. Despite significant characterization of APOBEC3-induced mutations in tumors, it is not known if APOBEC3-induced mutations can initiate a cancer, cause evolution of a cancer or both[25]. The main hypothesis is that APOBEC3 enzymes create uracils randomly throughout the genome and that the majority of these are repaired by Base Excision Repair, but some may not be repaired, resulting in transition mutations, and some may be repaired in an error prone manner, resulting in transversion mutations[26-28]. Too few mutations and there is no effect, too many mutations and there is cell death or immune recognition, and somewhere in the middle, APOBEC3 enzymes can cause a "just right" rate of mutations for tumor evolution[28]. These fates depend on the cell cycle and the numbers of mutations. Although this mutagenic process can make up at least 60% of the mutations in a tumor genome, the mutagenesis is random and acts to increase the genetic diversity in the tumor cell population, facilitating the chance of a beneficial mutation arising that helps maintain the cell population[5,13,17,27,29,30]. The interplay between APOBEC3 enzymes in this process is not known. Although there have been multiple reports of A3B, A3A, or A3H Hap I individually being involved in a cancer, it is not known if more than one APOBEC3 is expressed in a cancer cell at the same time and what would be the effects[13,16-18,31]. What has recently been identified is that APOBEC3 mutations occur episodically in cancers[32]. This is thought to be because constant expression and mutagenesis would result in cell death or recognition by the immune system, rather than creating a "just right" level of mutagenesis for

bioRxiv preprint doi: https://doi.org/10.1101/2020.02.28.970509; this version posted March 2, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

cancer evolution. It is also known that for lung cancer, A3H Hap I predominantly causes early mutations and A3B predominantly causes late stage mutations[13].

Interestingly, there are seven haplotypes for A3H, defined by polymorphisms or deletions at positions 15, 18, 105, 121, and 178, but only A3H Hap I is primarily localized to the nucleus[33-35]. The other A3H haplotypes are primarily in the cytoplasm[35]. Only those APOBEC3 members that have access to the nucleus can contribute to cancer mutagenesis. In addition, not all A3H haplotypes are catalytically active. Some haplotypes are unstable (III, IV, VI) some are hypo-stable (I) and others are stable (II, V, VII)[33-35]. Despite A3H Hap I being hypo-stable/-active, it is the only A3H that has been implicated in cancer mutagenesis. Starrett *et al.* have shown that A3H Hap I is an enzymatic contributor to 'APOBEC signature' mutations in lung cancer and likely contributor to breast cancer[13]. This was a surprising result since A3H Hap I has a half-life in cells of approximately 30 min[33]. The short A3H Hap I half-life also precludes the purification of significant quantities of A3H Hap I for use in biochemical assays. When cell lysates are used and A3H Hap I is overexpressed, at equivalent protein levels to more stable haplotypes, the activity is comparable[13]. A3H Hap I has a G105 that causes the low stability and mutation of this to an R105 forms a stable enzyme and is A3H Hap VII, which has been used as an A3H Hap I proxy *in vitro*[36].

One single nucleotide polymorphism, rs139298, resulting in A3H Hap I K121E has been recently reported to be associated with lung cancer[1]. Lung cancer results in over 1.6 million deaths per year and nearly 2 million new cases annually, with pulmonary adenocarcinoma comprising about 40% of the total cases of non-smoking lung cancer[37-40]. Surprisingly, we found that this K121E mutation further destabilized A3H Hap I, giving the counter-intuitive result that the absence of A3H Hap I is associated with lung cancer. The data in the present work support the hypothesis that although APOBEC3 enzymes contribute to cancer mutagenesis, constant exposure to APOBEC3 mutagenesis or multiple APOBEC3s

bioRxiv preprint doi: https://doi.org/10.1101/2020.02.28.970509; this version posted March 2, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

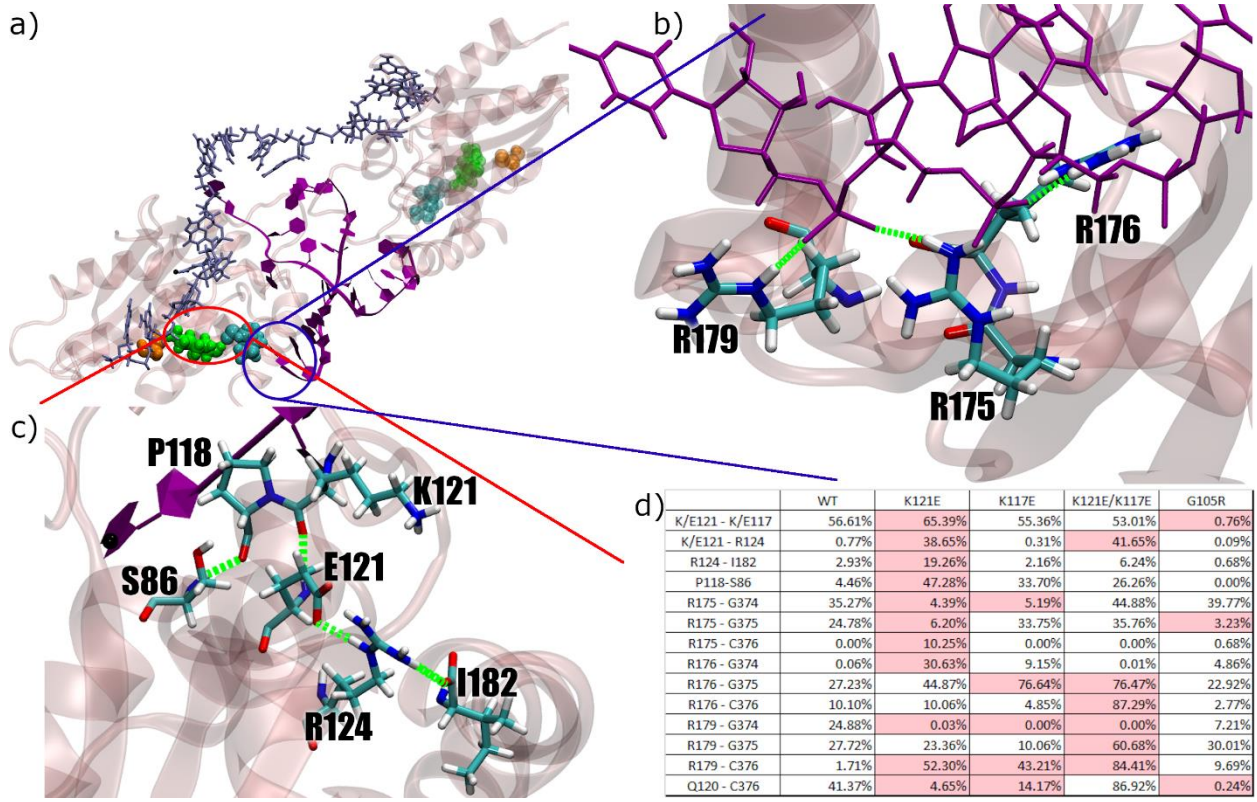
may be detrimental. In this study, we examine the A3H Hap I K121E variant and its stabilization by a novel A3H Hap I stabilizing mutation, K117E, using computational, biochemical, and cellular techniques. The combined results demonstrate that the timing of APOBEC3 mutations and the number of different APOBEC3s jointly expressed in cancer cells influences the fate of APOBEC3-induced somatic mutagenesis and provides the first cause and effect evidence of APOBEC3 enzymes and their role in cancer mutagenesis.

Results

A3H Hap I K121E results in the formation of a new hydrogen-bonding network

We used classical molecular dynamics (MD) to investigate the effect (if any) of the SNP resulting in the A3H Hap I K121E mutation on the protein structure and dynamics compared with A3H Hap I wild type (WT). Pairwise hydrogen bond analysis of the K121E indicates the formation of a new hydrogen bond (HBond) network across part of the protein surface compared with the WT. This HBond network induces strain on the active site suggesting a change in stability or activity (**Figure 1**). Hydrogen bond interactions between protein residues that differed by more than 30% of the simulation time were identified. The HBond network in the K121E system connects two existing networks (R124 to I182 and K117 to P118 to S87) from the WT into a single larger network in the K121E variant. This extended over-stabilized network effectively "freezes" the helix where the mutation occurs, disrupting both the active site and RNA-binding residues on the C-terminal helix. Changes in root mean squared fluctuation (RMSF) are also observed in the K121E system, with residues involved in the HBond network exhibiting reduced fluctuation while the rest of the protein exhibits increased fluctuation (**Figure 2**). Principle component analysis (PCA) shows that the first two modes of motion in the K121E system are constrained and tightly

bioRxiv preprint doi: https://doi.org/10.1101/2020.02.28.970509; this version posted March 2, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.



d)

	WT	K121E	K117E	K121E/K117E	G105R
K/E121 - K/E117	56.61%	65.39%	55.36%	53.01%	0.76%
K/E121 - R124	0.77%	38.65%	0.31%	41.65%	0.09%
R124 - I182	2.93%	19.26%	2.16%	6.24%	0.68%
P118-S86	4.46%	47.28%	33.70%	26.26%	0.00%
R175 - G374	35.27%	4.39%	5.19%	44.88%	39.77%
R175 - G375	24.78%	6.20%	33.75%	35.76%	3.23%
R175 - C376	0.00%	10.25%	0.00%	0.00%	0.68%
R176 - G374	0.06%	30.63%	9.15%	0.01%	4.86%
R176 - G375	27.23%	44.87%	76.64%	76.47%	22.92%
R176 - C376	10.10%	10.06%	4.85%	87.29%	2.77%
R179 - G374	24.88%	0.03%	0.00%	0.00%	7.21%
R179 - G375	27.72%	23.36%	10.06%	60.68%	30.01%
R179 - C376	1.71%	52.30%	43.21%	84.41%	9.69%
Q120 - C376	41.37%	4.65%	14.17%	86.92%	0.24%

Figure 1: **a)** APOBEC3H protein (transparent pink) with RNA interface (purple) and DNA substrate (iceblue). G105 is shown in orange, K117 in cyan, and K121 in green, with corresponding residues on opposite monomer shown in transparency. **b)** RNA-binding residues on terminal helix. **c)** Formed hydrogen bonding network as result of K121E mutation. **d)** Table of largest contributors in each hydrogen bonding interaction, shown as percentage of total simulation time.

correlated in comparison to the WT (**Figure 2**). Difference correlation matrices reveal that K121E has large regions of changing correlated movement, especially with respect to the region around the mutation point.

The simulation data suggest that a K117E mutation could rescue the K121E variant by restoring the two hydrogen bonding networks. The K121E/K117E variant removes the HBond between the side chains of 121 and 117, and partially restores the remaining interactions in the network towards their WT levels. (**Figure 1**). The K121E/K117E system shows less of a change in fluctuation, with residues in the network exhibiting near WT fluctuation, similarly to what is observed for G105R (**Figure 2**). The PCA

bioRxiv preprint doi: https://doi.org/10.1101/2020.02.28.970509; this version posted March 2, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

shows that the first two modes are similar to those of the WT, as does the comparatively minor change in correlated movement.

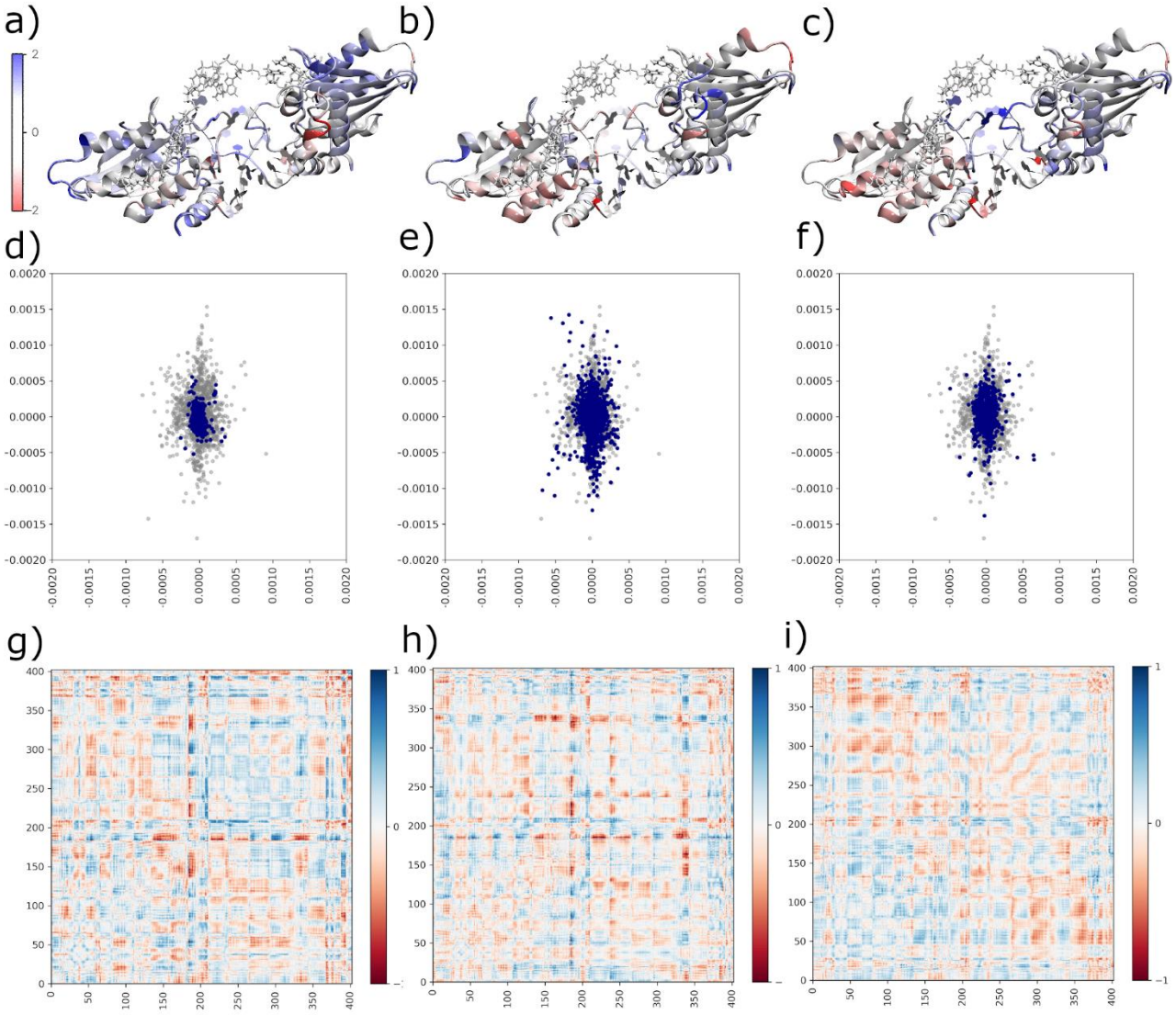


Figure 2: Difference RMSF with respect to wildtype overlaid on protein structures for **a)** K121E, **b)** K121E/K117E, and **c)** K117E. Principle component analyses for mutant systems (blue) over wildtype (grey) for **d)** K121E, **e)** K121E/K117E, and **f)** K117E. Difference correlation matrices against wildtype reference for **g)** K121E, **h)** K121E/K117E, and **i)** K117E.

K121E destabilizes A3H Hap I

To test the computational results we produced the A3H Hap I variants in 293T cells for cell-based

bioRxiv preprint doi: https://doi.org/10.1101/2020.02.28.970509; this version posted March 2, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

analysis. To determine the steady state expression levels of the A3H Hap I WT, K121E cancer variant, putative rescue mutant K117E/K121E, and rescue mutant control K117E, we transfected 3x HA-tagged A3H expression constructs into 293T cells. Immunoblotting showed that steady state expression levels for the cancer variant were not detectable by immunoblotting, indicating that the K121E mutation destabilized A3H Hap I WT (**Figure 3a**). We found that the K117E mutation stabilized A3H Hap I in the presence of the otherwise destabilizing G105 amino acid[33] and, as predicted, can stabilize the cancer variant (K117E/K121E) (**Figure 3a**). Since the destabilizing SNP was associated with lung cancer[1], these data suggested that A3H Hap I somatic mutations may be detrimental to cancer progression perhaps due to extensive mutagenesis leading to cell dysfunction or immune recognition. This is consistent with A3H Hap I mutations being identified early in cancer and the episodic nature of APOBEC3-induced mutations in cancer[1,13,32].

Combined expression of A3B and A3H Hap I in a lung cancer cell line induces high amounts of double-stranded DNA breaks

To directly test this hypothesis, we used lung cell lines from two sources representing normal (MRC-5, male fetal tissue) and cancer (NCI-H1563, adenocarcinoma from a male non-smoker) cells. We first determined if these cell lines expressed any endogenous A3s. We found that both the MRC-5 and NCI-H1563 cell lines expressed high amounts of A3B mRNA compared to the TBP mRNA control (**Figure 3b-c**). The two cell lines expressed 7-fold (MRC-5) and 5-fold (NCI-H1563) more *A3B* mRNA than *TBP* mRNA, which is equivalent to the amounts of *A3B* mRNA found in breast cancer cell lines and responsible for genomic mutations[17]. The endogenous *A3H* (not genotyped) was at an equivalent level to *TBP* and 7- to 5- fold less than A3B (**Figure 3b-c**). This provided an opportunity to transduce these cells with an A3H-Hap I-Flag expression construct to make doxycycline (dox)-inducible stable cell lines (**Figure SF6**) and test the effect of multiple A3s expressed in normal and cancer lung cells. A key marker

bioRxiv preprint doi: https://doi.org/10.1101/2020.02.28.970509; this version posted March 2, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

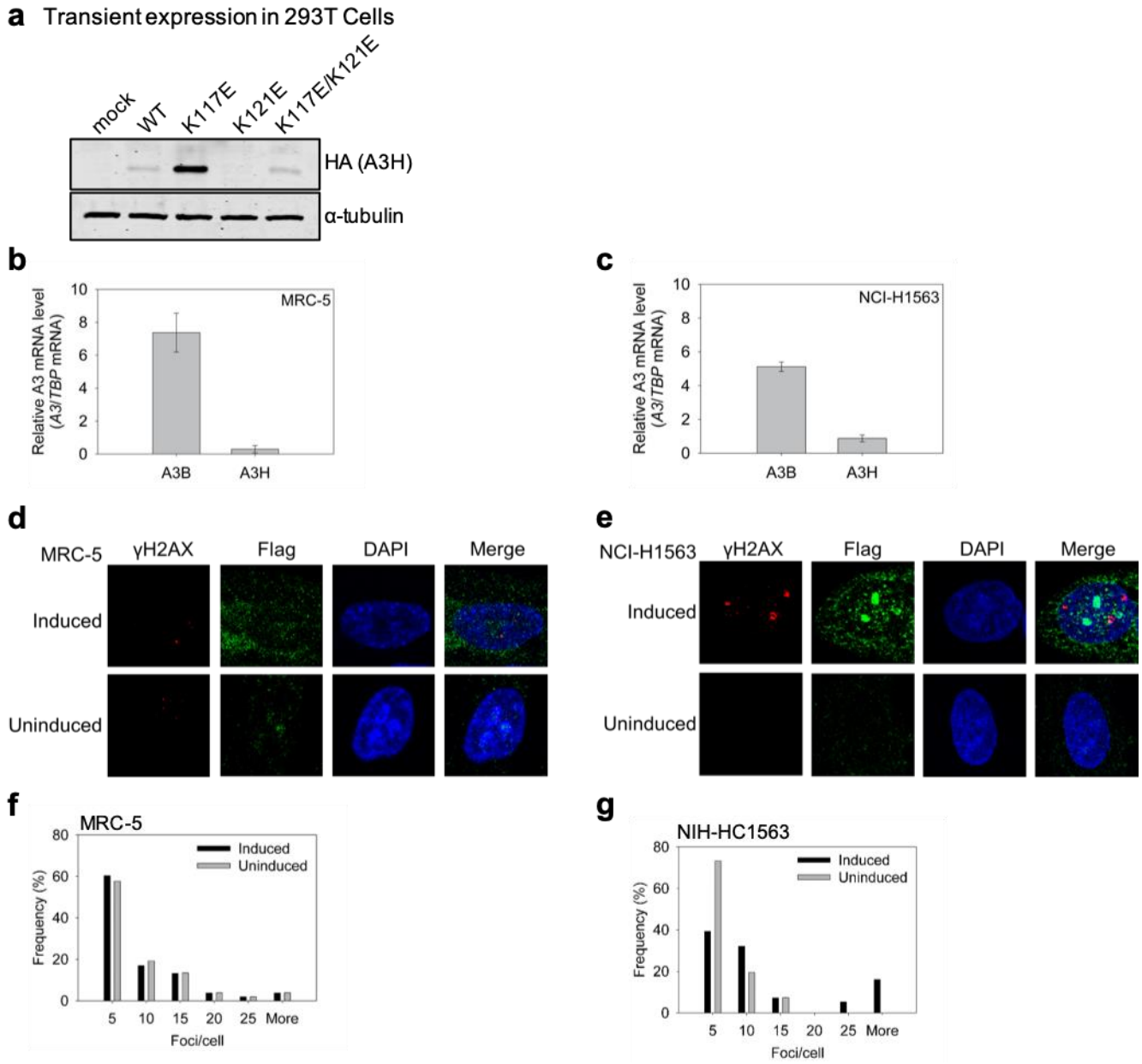


Figure 3: **Expression and mutagenic activity of A3H Hap I. a)** Transient expression of HA-tagged A3H Hap I expression constructs in 293T cells were detected by immunoblotting. The α-tubulin served as the loading control. The endogenous *A3B* and *A3H Hap I* mRNA levels were detected in **b)** MRC-5 and **c)** NCI-H1563 cells relative to *TBP* mRNA. The A3H Hap I mRNA levels were low, so doxycycline inducible A3H Hap I-Flag cell lines were produced (Figure SF6). **(d-g)** A3H Hap I induces high amounts of γH2AX foci in NCI-H1563 cells with endogenous A3B expression. The γH2AX foci were detected in induced and uninduced cells by immunofluorescence microscopy in **d)** MRC-5 and **e)** NCI-H1563 cells. Results were quantified and plotted as a histogram for **f)** MRC-5 and **g)** NCI-H1563 cells.

bioRxiv preprint doi: https://doi.org/10.1101/2020.02.28.970509; this version posted March 2, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

of high A3 activity is the formation of double-stranded (ds) DNA breaks that are induced by base excision of closely located uracils. To avoid any possible cell death, we only induced cells with dox for 24 h, then cells were fixed and antibodies used to detect γH2AX foci as a marker of dsDNA breaks.

In MRC-5 cells the endogenous A3B alone and A3B with A3H-Hap I-Flag were not significantly different and showed that ~60% of cells had five γH2AX foci/cell, ~20% of cells had ten γH2AX foci/cell, and ~13% of cells had fifteen γH2AX foci/cell. Larger numbers of foci per cell were observed in less than 10% of the population (**Figure 3d, f**). These data support that there is sufficient DNA repair to process A3-induced dsDNA breaks from multiple A3s in normal cells.

In NCI-H1563, the lung adenocarcinoma cell line, there are mutations in KRAS, PIK3R1, SMARCA4, and p53[41], indicating that it has common cancer mutations that dysregulate cell cycle, cell division and chromatin remodelling. This makes cancer cells more susceptible to A3-induced mutagenesis. However, with only endogenous A3B expressed, the number of γH2AX foci/cell were approximately the same as both conditions tested for MRC-5 cells, i.e., ~60 to 70% of cells having five γH2AX foci/cell (**Figure 3e, g**). Nonetheless, when A3H Hap I-Flag expression was induced the number of γH2AX foci/cell increased dramatically with 16% of cells having over twenty-five γH2AX foci/cell (range was between 33 to 61 per cell). The other major populations were 32% and 39% of cells with ten and five γH2AX foci/cell, respectively (**Figure 3e, g**). These data demonstrate that more than one A3 expressed in a cancer cell can induce a large amount of dsDNA breaks, which may result in cell death or chromosomal aberrations. This supports the genetic data that the loss of A3 activity, specifically A3H Hap I, can be associated with cancer[1].

A G105R mutation results in a more active and stable A3H than a K117E mutation

Although a K117E mutation can stabilize A3H Hap I, it is not a naturally occurring variant. The

bioRxiv preprint doi: https://doi.org/10.1101/2020.02.28.970509; this version posted March 2, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

naturally occurring stabilization of A3H Hap I is a G105R mutation, which is known as A3H Hap VII. The reason for the instability induced by G105 is not known. Computational models of the G105R system exhibit dynamic motion and hydrogen bonding patterns similar to the WT, indicating that this mutation does not negatively affect the structural stability of the protein (see SI for G105R computational data). The G105R may simply increase the propensity of the protein to become polyubiquitinated and degraded[42], in contrast to K121E that destabilizes the structural integrity of A3H Hap I (**Figure 2**). To test if the longer steady state half-life of A3H K117E and K117E/K121E in cells correlates with increased catalytic activity similar to A3H Hap VII, we conducted an *in vitro* deamination assay using protein purified from *Sf*9 insect cells to obtain a quantitative measurement. A3H deaminase activity was observed by adding a 118 nt ssDNA with two A3H deamination motifs (5'CTC) and an internal fluorescein label to equivalent amounts of the cell lysates. The two 5'CTC motifs account for the preference of some A3H variants to preferentially deaminate one cytosine motif more than the other[43]. We found that despite recovery of protein stability the A3H K117E and K117E/K121E were 4-fold less active than A3H Hap VII (relative to Hap I, G105R) or A3H Hap II (relative to Hap I, G105R, E178D) (**Figure 4a**). These data suggest that the WT hydrogen bonding network is important for catalytic activity and compensatory mutations can correct the instability, but not fully reinstate catalytic activity. Despite the decreased specific activity, the mutants were also processive, similar to what has been reported for A3H Hap VII and A3H Hap II (**Figure 4b**)[43].

bioRxiv preprint doi: https://doi.org/10.1101/2020.02.28.970509; this version posted March 2, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

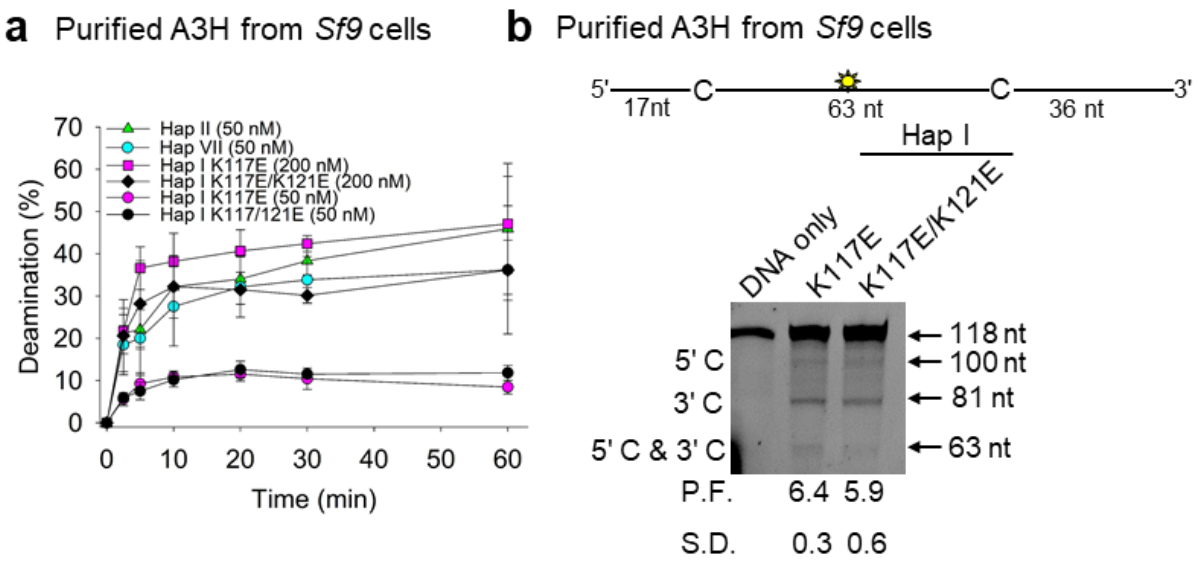


Figure 4. Deamination activity of A3H WT and mutants. a) Time course of deamination of A3H Hap II, A3H Hap VII, and A3H Hap I WT and mutants. Deamination was tested on a 118 nt ssDNA (100 nM) and gels analyzed for each plot are shown in Figure SF5. **b)** Processivity of A3H Hap I K117E and K117E/K121E. The processivity factor (P.F.) measures the likelihood of a processive deamination over a nonprocessive deamination and is described further in the Methods. Both K117E and K117E/K121E are approximately 6-fold more likely to undergo processive deamination than a nonprocessive deamination. The standard deviation for three independent experiments is shown as (**a**) error bars or (**b**) below the gel.

A3H cancer variant destabilizes dimer interface

From monkeys to greater apes, the A3 family has lost activity. Although A3B activity can be lost through a gene deletion[44], other A3s have lost activity because of decreased catalytic activity due to mutation of residues near the active site, e.g., A3D or loss of dimerization, e.g., A3C[45,46]. Based on recent crystal structures, the association of A3H with RNA is concomitant with enzyme stability and dimerization[9,47-50]. This is a unique feature of A3H compared to other A3s, in that it uses a double-stranded RNA molecule to form a dimer interface without any protein-protein contacts and that this imparts stability to the enzyme. As a result, we hypothesized that the reason for the A3H cancer variant instability (**Figure 3a**) may be due to the inability to bind RNA and dimerize. RNA-mediated dimerization can be detected by treating a purified protein preparation with RNase A and observing if a

bioRxiv preprint doi: https://doi.org/10.1101/2020.02.28.970509; this version posted March 2, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

~12 nt RNA is protected from digestion. We confirmed that recombinant A3H Hap I mutants K117E and K117E/K121E purified from *Sf*9 cells with RNA and RNase A treatment resulted in a ~12 nt RNA being protected from degradation by the protein (**Figure 5a**). The A3H Hap VII and A3H Hap II also bound cellular RNA and protected a ~12 nt RNA in the presence of RNase A. Accordingly, all four A3H had the same Size Exclusion Chromatography (SEC) profile and formed primarily a dimer (**Figure 5b-d**).

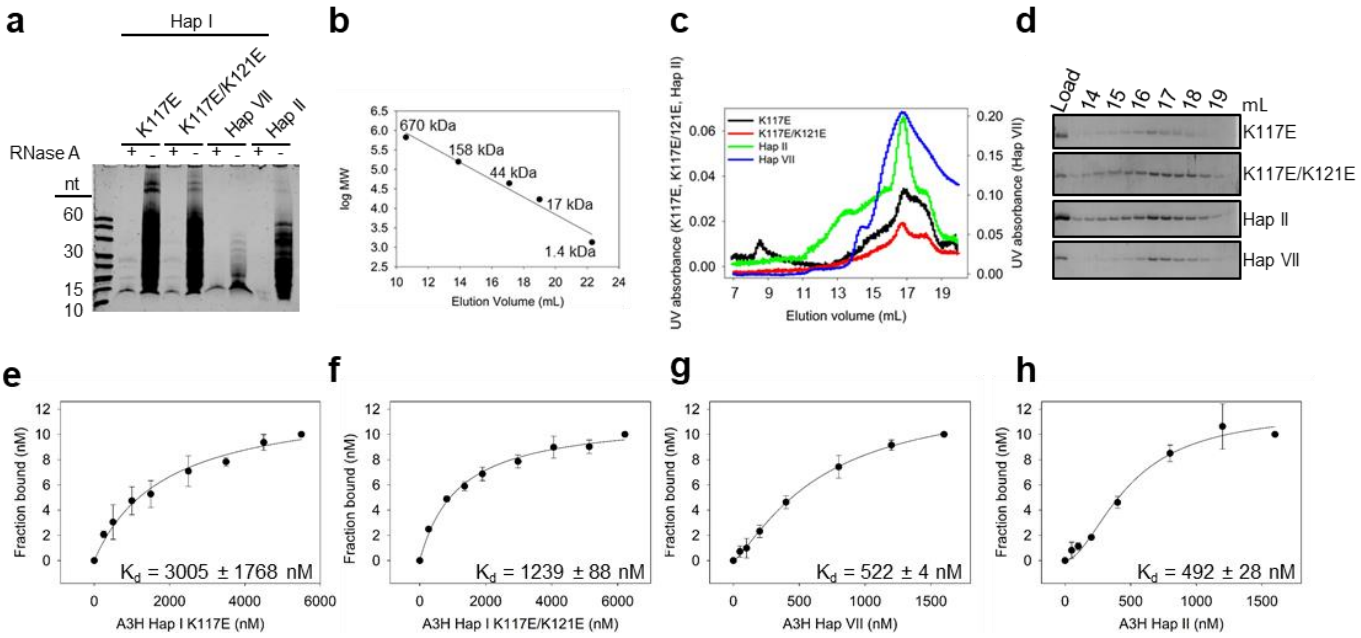


Figure 5: **A3H Hap I mutant dimerization, oligomerization, and ssDNA binding. a)** A3H dimerization is mediated by RNA. **a)** Purified A3H was denatured in formamide buffer and samples were resolved by urea denaturing PAGE. The gel was stained with SYBR Gold to detect nucleic acids. The A3H Hap I K117E, A3H Hap I K117E/K121E, A3H Hap VII, and A3H Hap II protect an approximately 12-15 nt RNA. Representative image shown from three independent experiments. **b)** Standard curve and **c-d)** SEC profile for A3H Hap I K117E, A3H Hap I K117E/K121E, A3H Hap VII, and A3H Hap II demonstrating similar elution profiles that are composed primarily of a dimer peak (44 kDa, 17 mL elution volume). **(e-h)** The apparent K_d of A3H enzymes from a 118 nt ssDNA was analyzed by steady-state rotational anisotropy. The **e)** A3H Hap I K117E (3005 ± 1768 nM) and **f)** A3H Hap I K117E/K121E (1239 ± 88 nM) bind the ssDNA with different affinities than **g)** A3H Hap VII (522 ± 4 nM) and **h)** A3H Hap II (492 ± 28 nM). The apparent K_d was calculated by determining the best fit by least squares analysis which was a hyperbolic fit for **e-f** and a sigmoidal fit for **g-h**. Error bars represent the S.D. from three independent experiments.

bioRxiv preprint doi: https://doi.org/10.1101/2020.02.28.970509; this version posted March 2, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

Since we cannot purify the K121E variant, we used computational analysis to predict whether the change at amino acid 121 would destabilize dimerization. RMSF analysis for WT, K121E and K117E/K121E (**Figure 2**) shows that residues in the K121E system exhibited increased fluctuation compared with WT, except those residues involved in the new hydrogen bonding network and many that interact with the RNA interface. These RNA-binding residues (R175, R176, R179) exhibit changing hydrogen bonding interactions with the RNA interface. In the simulations, these changes cause the RNA to change positions, affecting additional protein-RNA interaction at loop 1 residues. Correlation matrices were calculated for all systems, with the K121E variant system showing increased correlation between the monomer subunits and decreased correlation within each monomer when compared to the WT (**Figure SF2**), suggesting that the dimer interface is destabilized by the newly formed hydrogen bonding network. The K117E system shows qualitatively similar zones of correlation and anticorrelation at the dimer interface with respect to the WT. The K117E/K121E mutant correlation matrices reveal that this system more closely resembles the K121E than either the K117E or WT dimer systems, with qualitatively similar correlation matrices. These data are consistent with poor expression in cells (**Figure 3a**). Further, consistent with the biochemical data, the K117E/K121E variant system exhibits fewer changes in fluctuation, and there is not a general trend of increased fluctuation as seen in the K121E variant.

A3H variants have weakened interactions with the substrate

We observed that both the A3H Hap I K117E and K117E/K121E had ~6-fold (K117E) and ~2.5-fold (K117E/K121E) higher apparent dissociation constants from ssDNA than A3H Hap VII or A3H Hap II (**Figure 5e-h**). Further, both the A3H Hap I K117E and K117E/K121E binding curves best fit to a rectangular hyperbola by least squares analysis, in contrast to A3H Hap VII and A3H Hap II that best fit to a sigmoid. The sigmoidal binding relationship for A3H Hap II has been shown to be due to further

bioRxiv preprint doi: https://doi.org/10.1101/2020.02.28.970509; this version posted March 2, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

oligomerization of A3H dimers on ssDNA[48]. This does not appear to occur with A3H Hap I K117E or K117E/K121E (**Figure 5e-f**), although they can form dimers in solution (**Figure 5b-d**). The hindered oligomerization on ssDNA that may be due to a faster off-rate from the substrate likely is what affected the deamination activity (**Figure 4a**). Altogether, the data indicate that the SNP that inactivates A3H Hap I results in an inability to use RNA for dimerization, emphasizing the importance of cellular RNA to the structural stability of A3H. Further, the hydrogen bonding network is important for interactions with ssDNA and oligomerization on ssDNA, providing a reasoning for the decreased catalytic activity in A3H Hap I K117E and K117E/K121E.

Discussion

A3H Hap I was identified to induce genomic mutations specifically in lung cancer[13]. Interestingly, the activity of A3H Hap I was found because researchers wanted to investigate how genomic mutations with a cytidine deaminase signature were still occurring in people with an *A3B-/-* deletion[51-54], which occurs in the world population at 22.5%, although it primarily occurs in Oceanic populations[44]. Although A3A is also involved in cancer mutagenesis, it appears to be strongly associated with breast cancer[31,55]. The association of A3H Hap I expression and mutation signature were strong in early clonal variants of lung cancers. Further supporting a role of A3H Hap I in lung cancer was a computational study that identified the association of SNP rs139298 in A3H Hap I in lung cancer[1]. However, we demonstrate in this work that the SNP makes the A3H Hap I completely unstable, suggesting that loss of this enzyme specifically promoted cancer. While the effects arising from the combination of environmental, e.g., smoking, and A3 mutations remain unknown, we did assess if multiple A3 enzymes would cause too many mutations and be detrimental to cancer evolution. Recent reports support the seemingly contradictory idea that A3 enzymes, specifically APOBEC3B can both contribute to cancer and contribute to success of

bioRxiv preprint doi: https://doi.org/10.1101/2020.02.28.970509; this version posted March 2, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

cancer immune therapy due to A3-mutation created neoepitopes[56] or efficiency of platinum-based drugs due to contributions to DNA damage[57].

In this report we came to an equivalent conclusion since we determined that the previously identified rs139298 SNP that creates a K121E mutation in A3H Hap I resulted in a loss of enzyme stability in cells. While based on previous evidence, it appears that A3H Hap I, in an A3B null background, can promote genome mutations that contribute to lung cancer[13], it had not been investigated what would occur in the presence of another A3. Surprisingly, we found that two lung cell lines, one normal and the other cancerous, expressed similar and high amounts of endogenous A3B. The A3H Hap I in the cancer cell line, co-expressed with A3B increased the number of dsDNA breaks significantly compared to A3B alone. These dsDNA breaks could lead to cell death, destructive chromosomal relocations or high amounts of neoantigens. While in many A3/cancer studies it is difficult to determine the fate of A3 mutations, the genetic data combined with computational and biochemical study of the K121E mutation demonstrates that A3H Hap I has these detrimental effects on cell growth and proliferation in lung cancer.

The inactivation of A3 enzymes through SNPs is not unique to A3H, but A3H does have the highest number of inactivating SNPs in the A3 family. In particular, A3H has been lost multiple times in primate evolution, presumably due to nuclear localization and acquisition of genomic mutations[33]. While it is thought that A3H is kept active at a certain level in the population due to the antiviral effects of the enzyme, there are strong population stratifications in A3H active/hypo-active/not active forms that correlate with the historical levels of HIV infection, a pathogen that active A3Hs can restrict efficiently[58,59]. The antiviral properties of these enzymes are likely why they are maintained despite the negative effect of their off-target activity. The destabilizing SNPs for A3H other than the one studied here, are not thought to destabilize protein structure, but to promote A3H ubiquitination and proteosomal degradation[42]. Thus, the SNP studied in this work for A3H Hap I is unique since it appears to destabilize

bioRxiv preprint doi: https://doi.org/10.1101/2020.02.28.970509; this version posted March 2, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

the enzyme by disrupting a hydrogen bonding network. While this disruption decreased catalytic activity, it did not cause protein unfolding. Rather, it decreased the interaction strength with a dsRNA molecule that A3H acquires from the cell and uses to dimerize[9,47-50]. This dimerization using an RNA intermediate is essential to A3H thermodynamic stability[47,48]. While the nature of dimerization for different A3s is unique, e.g., only A3H uses an RNA intermediate, a SNP that regulates activity through dimerization has also been found for A3C[46,60]. This type of functional inactivation enables rapid evolutionary toggling of activity for times when enzyme activity is needed again[61].

Since the identification of distinct mutational signatures and high A3 expression in cancers[5,16,17], a number of research groups have contributed to our understanding of how A3 enzymes access ssDNA, which A3 enzymes are involved, and specifically for A3B, the clinical effects of the induced mutations[18,19,21,23,62-64]. However, there are no studies examining multiple A3 expression in a cancer cell, although this potentially could happen with three A3 enzymes involved in cancer mutagenesis. The data presented here supports the idea that the expression of multiple A3s in a single cell may benefit the host and be detrimental to the cancer since the A3H Hap I K121E is associated with lung cancer, and as shown here, is destabilized in cells. Based on analysis of dsDNA breaks, it is likely that at least A3H Hap I and A3B are not compatible for promoting cancer due to the accumulation of too many mutations. In combination with the genetic data[1], our data provide direct evidence for such a cause and effect relationship of A3s and cancer.

Methods

Computational Methods

Molecular dynamics (MD) simulations were performed with the ff14SB[65], OL15[66], YIL[67], and TIP3P[68] force fields using the pmemd.cuda[69,70] and OpenMM[71] programs. All systems were constructed

bioRxiv preprint doi: https://doi.org/10.1101/2020.02.28.970509; this version posted March 2, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

from pdbid 5W45, corresponding to A3H-HapI based on the UniProt consensus sequence[72,73]. A search of 12 possible binding orientations of the DNA substrate on one single monomer was performed to determine the most probable orientation for the DNA substrate (see SI for coordinates of dimer in holo form)[74].

The most probable orientation was subsequently used to construct the biologically relevant dimer systems using the macaque-APOBEC3H dimer structure (pdbid 5W3V) as a template for UCSF Chimera and taking into account the polarity of the substrate[75-77]. The monomer crystal of human A3H was overlaid on the dimer crystal of the macaque A3H and aligned. The protonation states for the dimer systems were determined using H++, neutralized with K^+ (**Table ST2**) and solvated in a box of TIP3P water of approximately 143Å on each side to allow free movement without self-interaction across the periodic boundary[78-80]. Particle mesh Ewald was employed for long-range electrostatics, with a cutoff distance of 10 Å and an error tolerance of 10^{-4} kJ mol^{-1} $Å^{-2}$.[70]

All simulations were performed in the NPT ensemble at 1.0 bar and 300 K using a Monte Carlo barostat, and a Langevin thermostat with a 1 fs timestep[71,81,82]. Simulation frames were saved at 10 ps intervals. To maintain active site geometry in the simulations, distance restraints of 20 kcal/mol· $Å^2$ at 2.0 Å were applied between the Zn^{2+} atoms and their respective coordinating histidines, with additional 10 kcal/mol· $Å^2$ restraints at 5.0 Å between the Zn^{2+} and water-coordinating glutamate to prevent incursion into the active site. Each system was minimized for 2,000 steps and equilibrated for 20ns before beginning production. Additional simulations using distance constraints of 2.0 Å between the Zn^{2+} and coordinating atoms were also run and found to give quantitatively similar results to the restrained simulations (data not shown).

Five systems were constructed including A3H Hap I wildtype (WT), A3H Hap I K121E (cancer variant), A3H Hap I K117E/K121E, A3H Hap I K117E, and A3H Hap I G105R. Every system was run

bioRxiv preprint doi: https://doi.org/10.1101/2020.02.28.970509; this version posted March 2, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

for 250 ns in triplicate (750ns total simulation time per variant). Energy decomposition analysis (EDA) was performed using an in-house FORTRAN90 program[83-85]. RMSF, RMSD, correlation, hydrogen bonding analysis, clustering, normal mode analysis, and distance calculations were performed using cpptraj in the AmberTools suite[86,87].

Immunoblotting

Transfection of 1 x 10^5 293T cells per well of a 6-well plate was carried out using GeneJuice (Novagen) transfection reagent as per manufacturer's protocol. Cells were maintained in DMEM with 10% FBS. Plasmids transfected were empty pcDNA3.1 (mock) and the following C-terminally tagged A3H-3x HA constructs in pcDNA3.1: A3H Hap I, A3H Hap I K117E, A3H Hap I K121E, and A3H Hap I K117E/K121E. After 48 h, cells were lysed using 2x Laemmli Buffer and 30 µg total protein was used. A3H was detected using an anti-HA mouse antibody (Sigma) and loading control for cell lysate detected α-tubulin using an anti- α-tubulin rabbit antibody (Invitrogen). Secondary detection was performed using Licor IRDye antibodies produced in goat (IRDye 680-labeled anti-rabbit and IRDye 800-labeled anti mouse).

Purification of A3H from *Sf9* insect cells

The pFAST-bac1vectors were used to produce recombinant baculovirus according to the protocol for the Bac-to-Bac system (Life Technologies). Recombinant GST-A3H Hap I WT and mutant baculovirus (K117E, K117E/K121E) was used to infect *Sf9* cells at a multiplicity of infection of 1 and cells were harvested after 72 h. Cells lysates were treated with 100 µg/mL of RNase A (Qiagen). Lysates were cleared by centrifugation and then incubated with Glutathione-Sepharose 4B resin (GE Healthcare) at 4 °C overnight and subjected to a series of salt washes (0.25M to 1M NaCl) as described previously[88]. On-column cleavage from the GST tag with Thrombin (GE Healthcare) was performed at 21 °C for 18 h

bioRxiv preprint doi: https://doi.org/10.1101/2020.02.28.970509; this version posted March 2, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

in thrombin digestion buffer (20mM HEPES, pH 7.5, 150 mM NaCl, 10% glycerol, and 2 mM DTT). Proteins were assessed to be 90% pure by SDS-PAGE (**Figure SF7**).

Deamination assays

The 118 nt ssDNA substrate was previously described[43]. The 118 nt ssDNA substrate had the following sequence: 5′GAA TAT AGT TTT TAG CTC AAA GTA AGT GAA GAT AAT [Fluorescein-dT] TAG AGA GTT GTA ATG TGA TAT ATG TGT ATG AAA GAT ATA AGA CTC AAA GTG AAA AGT TGT TAA TGT GTG TAG ATA TGT TAA. The ssDNA substrate (100 nM), which contained two A3H deamination motifs (5′CTC, with the underlined C being deaminated), was incubated with 200 nM of A3H Hap I mutants (K117E, K117E/K121E) or 50 nM of A3H Hap II, Hap VII and Hap I mutants for 60 min at 37 °C in buffer that contained (50 mM Tris pH 7.5, 40 mM KCl, 10 mM $MgCl_2$, and 1 mM DTT). For determination of enzyme processivity, additional reactions were carried out under single-hit conditions, i.e., <15% substrate usage, to ensure that deaminations on each ssDNA were catalyzed by a single enzyme[89]. Under these conditions, a processivity factor can be determined by comparing the total amount of deaminations occurring at two sites on same ssDNA to a calculated theoretical value of deaminations at these two sites if the deamination events were uncorrelated (not processive) [90]. A3H catalyzed deaminations were detected by treating the ssDNA with Uracil DNA Glycosylase (New England Biolabs) and heating under alkaline conditions before resolving the Fluorescein dT-labeled ssDNA on a 10% (v/v) denaturing polyacrylamide gel. Gel images were obtained using a Chemidoc-MP imaging system (Bio-Rad) and integrated gel band intensities were analyzed using ImageQuant (GE Healthcare).

Detection of RNA bound to A3H

To examine the RNA present in purified A3H, 3.5 μg A3H was or was not treated with RNase A

bioRxiv preprint doi: https://doi.org/10.1101/2020.02.28.970509; this version posted March 2, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

(Roche Applied Science) for 15 min at room temperature. At the end of the reactions, samples were mixed with an equal volume of formamide containing 5 mM EDTA and resolved on a 20% (v/v) denaturing polyacrylamide gel. The resolved nucleic acids were stained with SYBR-GOLD (Invitrogen) and gel images were obtained using a Chemidoc-MP imaging system (Bio-Rad).

Size exclusion chromatography

The oligomerization states of A3H Hap I mutants were determined by loading 120 to 300 μg of the purified enzymes onto a Superdex 200 10/300 Increase (GE Healthcare). The running buffer contained 20 mM Tris pH 8.0, 300 mM NaCl, 10% (v/v) glycerol and 1 mM DTT. The Bio-Rad gel filtration standard set was used to generate a calibration curve from which the apparent molecular masses and oligomerization states of the enzymes were determined.

Steady-state fluorescence polarization

The apparent dissociation constant (K_d) values of A3H Hap II, A3H Hap VII, and A3H Hap I mutants K117E and K117E/K121E for fluorescein labeled ssDNA (118 nt) were determined using steady state fluorescence polarization to measure rotational anisotropy. Reactions (50 μL) were conducted in deamination buffer (50 mM Tris, pH 7.5, 40 mM KCl, 10 mM $MgCl_2$, and 2 mM DTT) and contained 50 nM fluorescein labeled DNA or RNA and increasing amounts of A3H. A QuantaMaster QM-4 spectrofluorometer (Photon Technology International) with a dual emission channel was used to collect data and calculate anisotropy. Measurements were performed at 21°C. Samples were excited with vertically polarized light at 495 nm (6 nm band pass) and vertical and horizontal emissions were measured at 520 nm (6 nm band pass). The K_d was obtained by fitting to a rectangular hyperbola or sigmoidal curve equation using SigmaPlot 11.2 software.

bioRxiv preprint doi: https://doi.org/10.1101/2020.02.28.970509; this version posted March 2, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

Cell culture and generation of stable cell lines

NCI-H1563 and MRC-5 were obtained from ATCC and cultured in RPMI-1460 supplemented with 10% FBS, 10 mM HEPES, 1mM sodium pyruvate and DMEM supplemented with 10% FBS, respectively. Lentivirus was generated by co-transfecting psPAX2, pVSV-G, and the pLVX lentiviral vector containing A3H hap I-Flag with GeneJuice (EMD Millipore) in HEK-293T cells. Media was changed after 16 hr and viral particles were harvested 48 hr post transfection. NCI-H1563 and MRC-5 were transduced with the resulting lentivirus by incubation for 16 hr in medium containing 8 ug/mL polybrene. Transduced cells were selected with 1 ug/mL puromycin for a week and maintained with 0.25 ug/mL puromycin.

Immunofluorescence microscopy

NCI-H1563 and MRC-5 cells were treated with 2 ug/mL doxycycline for 24 h to induce A3H Hap I-Flag and fixed with 100% cold methanol for 10 min. Cells were permeabilized with 100% cold acetone for 1 min and 0.5% Triton X-100 for 10 min. Anti-Flag and anti- γH2AX antibodies (Invitrogen) in 3% BSA in 4x SSC buffer were incubated 1 hr. Primary antibodies were detected with Alexafluor-594 and Alexafluor-488- conjugated secondary antibodies (Invitrogen). Nuclei were stained with DAPI and cells were imaged using the Zeiss LSM700 system.

qPCR

The RNA preparation, cDNA synthesis and qPCR was carried out according to Refsland et al.[91]

Supplementary Information

RMSD, RMSF, PCA, and correlation matrices for individual replicates, coordinate, and parameter files for each system, deamination activity, expression and purity are provided. Source code and R scripts for energy decomposition analysis are provided as additional SI material.

bioRxiv preprint doi: https://doi.org/10.1101/2020.02.28.970509; this version posted March 2, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

References

1 Silvestrov, P., Maier, S. J., Fang, M. & Cisneros, G. A. DNArCdb: A database of cancer biomarkers in DNA repair genes that includes variants related to multiple cancer phenotypes. *DNA Repair (Amst)* **70**, 10-17, doi:10.1016/j.dnarep.2018.07.010 (2018).

2 Adolph, M. B., Love, R. P. & Chelico, L. Biochemical Basis of APOBEC3 Deoxycytidine Deaminase Activity on Diverse DNA Substrates. *ACS Infect Dis* **4**, 224-238, doi:10.1021/acsinfecdis.7b00221 (2018).

3 Smith, H. C. RNA binding to APOBEC deaminases; Not simply a substrate for C to U editing. *RNA Biol* **14**, 1153-1165, doi:10.1080/15476286.2016.1259783 (2017).

4 Venkatesan, S. *et al.* Perspective: APOBEC mutagenesis in drug resistance and immune escape in HIV and cancer evolution. *Ann Oncol* **29**, 563-572, doi:10.1093/annonc/mdy003 (2018).

5 Alexandrov, L. B. *et al.* Signatures of mutational processes in human cancer. *Nature* **500**, 415-421, doi:10.1038/nature12477 (2013).

6 Harris, R. S., Hultquist, J. F. & Evans, D. T. The restriction factors of human immunodeficiency virus. *J Biol Chem* **287**, 40875-40883, doi:R112.416925 [pii]
10.1074/jbc.R112.416925 (2012).

7 Cheng, A. Z. *et al.* Epstein-Barr virus BORF2 inhibits cellular APOBEC3B to preserve viral genome integrity. *Nat Microbiol* **4**, 78-88, doi:10.1038/s41564-018-0284-6 (2019).

8 Suspene, R. *et al.* Extensive editing of both hepatitis B virus DNA strands by APOBEC3 cytidine deaminases in vitro and in vivo. *Proc Natl Acad Sci U S A* **102**, 8321-8326, doi:10.1073/pnas.0408223102 (2005).

9 Shaban, N. M. *et al.* The Antiviral and Cancer Genomic DNA Deaminase APOBEC3H Is Regulated by an RNA-Mediated Dimerization Mechanism. *Mol Cell* **69**, 75-86 e79, doi:10.1016/j.molcel.2017.12.010 (2018).

10 Salamango, D. J. *et al.* APOBEC3H Subcellular Localization Determinants Define Zipcode for Targeting HIV-1 for Restriction. *Mol Cell Biol* **38**, doi:10.1128/MCB.00356-18 (2018).

11 Bennett, R. P., Presnyak, V., Wedekind, J. E. & Smith, H. C. Nuclear Exclusion of the HIV-1 host defense factor APOBEC3G requires a novel cytoplasmic retention signal and is not dependent on RNA binding. *J Biol Chem* **283**, 7320-7327, doi:M708567200 [pii]
10.1074/jbc.M708567200 (2008).

12 Soros, V. B., Yonemoto, W. & Greene, W. C. Newly synthesized APOBEC3G is incorporated into HIV virions, inhibited by HIV RNA, and subsequently activated by RNase H. *PLoS Pathog* **3**, e15, doi:06-PLPA-RA-0322R2 [pii]
10.1371/journal.ppat.0030015 (2007).

13 Starrett, G. J. *et al.* The DNA cytosine deaminase APOBEC3H haplotype I likely contributes to breast and lung cancer mutagenesis. *Nat Commun* **7**, 12918, doi:10.1038/ncomms12918 (2016).

14 Zhu, M. *et al.* The eQTL-missense polymorphisms of APOBEC3H are associated with lung cancer risk in a Han Chinese population. *Sci Rep* **5**, 14969, doi:10.1038/srep14969 (2015).

15 Salter, J. D., Bennett, R. P. & Smith, H. C. The APOBEC Protein Family: United by Structure, Divergent in Function. *Trends in Biochemical Sciences* **41**, 578--594, doi:10.1016/j.tibs.2016.05.001 (2016).

16 Burns, M. B., Temiz, N. A. & Harris, R. S. Evidence for APOBEC3B mutagenesis in multiple human cancers. *Nat Genet* **45**, 977-983, doi:10.1038/ng.2701 (2013).

17 Burns, M. B. *et al.* APOBEC3B is an enzymatic source of mutation in breast cancer. *Nature* **494**, 366-370, doi:10.1038/nature11881 (2013).

18 Chan, K. *et al.* An APOBEC3A hypermutation signature is distinguishable from the signature of background mutagenesis by APOBEC3B in human cancers. *Nat Genet* **47**, 1067-1072, doi:10.1038/ng.3378 (2015).

19 Hoopes, J. I. *et al.* APOBEC3A and APOBEC3B Preferentially Deaminate the Lagging Strand Template during DNA Replication. *Cell Rep* **14**, 1273-1282, doi:10.1016/j.celrep.2016.01.021 (2016).

20 Kazanov, M. D. *et al.* APOBEC-Induced Cancer Mutations Are Uniquely Enriched in Early-Replicating,

bioRxiv preprint doi: https://doi.org/10.1101/2020.02.28.970509; this version posted March 2, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

Gene-Dense, and Active Chromatin Regions. *Cell Rep* **13**, 1103-1109, doi:10.1016/j.celrep.2015.09.077 (2015).

21 Kanu, N. *et al.* DNA replication stress mediates APOBEC3 family mutagenesis in breast cancer. *Genome Biol* **17**, 185, doi:10.1186/s13059-016-1042-9 (2016).

22 Haradhvala, N. J. *et al.* Mutational Strand Asymmetries in Cancer Genomes Reveal Mechanisms of DNA Damage and Repair. *Cell* **164**, 538-549, doi:10.1016/j.cell.2015.12.050 (2016).

23 Seplyarskiy, V. B. *et al.* APOBEC-induced mutations in human cancers are strongly enriched on the lagging DNA strand during replication. *Genome Res* **26**, 174-182, doi:10.1101/gr.197046.115 (2016).

24 Jarvis, M. C., Ebrahimi, D., Temiz, N. A. & Harris, R. S. Mutation Signatures Including APOBEC in Cancer Cell Lines. *JNCI Cancer Spectr* **2**, doi:10.1093/jncics/pky002 (2018).

25 Swanton, C., McGranahan, N., Starrett, G. J. & Harris, R. S. APOBEC Enzymes: Mutagenic Fuel for Cancer Evolution and Heterogeneity. *Cancer Discov* **5**, 704-712, doi:10.1158/2159-8290.CD-15-0344 (2015).

26 Adolph, M. B., Ara, A. & Chelico, L. APOBEC3 Host Restriction Factors of HIV-1 Can Change the Template Switching Frequency of Reverse Transcriptase. *J Mol Biol* **431**, 1339-1352, doi:10.1016/j.jmb.2019.02.015 (2019).

27 Roberts, S. A. & Gordenin, D. A. Clustered and genome-wide transient mutagenesis in human cancers: Hypermutation without permanent mutators or loss of fitness. *Bioessays*, doi:10.1002/bies.201300140 (2014).

28 Burns, M. B., Leonard, B. & Harris, R. S. APOBEC3B: pathological consequences of an innate immune DNA mutator. *Biomed J* **38**, 102-110, doi:10.4103/2319-4170.148904 (2015).

29 Roberts, S. A. *et al.* An APOBEC cytidine deaminase mutagenesis pattern is widespread in human cancers. *Nat Genet* **45**, 970-976, doi:10.1038/ng.2702 (2013).

30 Stephens, P. J. *et al.* The landscape of cancer genes and mutational processes in breast cancer. *Nature* **486**, 400-404, doi:10.1038/nature11017 (2012).

31 Cortez, L. M. *et al.* APOBEC3A is a prominent cytidine deaminase in breast cancer. *PLoS Genet* **15**, e1008545, doi:10.1371/journal.pgen.1008545 (2019).

32 Petljak, M. *et al.* Characterizing Mutational Signatures in Human Cancer Cell Lines Reveals Episodic APOBEC Mutagenesis. *Cell* **176**, 1282-1294 e1220, doi:10.1016/j.cell.2019.02.012 (2019).

33 OhAinle, M., Kerns, J. A., Li, M. M., Malik, H. S. & Emerman, M. Antiretroelement activity of APOBEC3H was lost twice in recent human evolution. *Cell Host Microbe* **4**, 249-259, doi:10.1016/j.chom.2008.07.005 (2008).

34 Wang, X. *et al.* Analysis of human APOBEC3H haplotypes and anti-human immunodeficiency virus type 1 activity. *J Virol* **85**, 3142-3152, doi:10.1128/JVI.02049-10 (2011).

35 Li, M. M. & Emerman, M. Polymorphism in human APOBEC3H affects a phenotype dominant for subcellular localization and antiviral activity. *J Virol* **85**, 8197-8207, doi:10.1128/JVI.00624-11 (2011).

36 Adolph, M. B., Love, R. P., Feng, Y. & Chelico, L. Enzyme cycling contributes to efficient induction of genome mutagenesis by the cytidine deaminase APOBEC3B. *Nucleic Acids Res* **45**, 11925-11940, doi:10.1093/nar/gkx832 (2017).

37 *World Cancer Report 2014*. (World Health Organization, 2014).

38 MacKinnon, A. C., Jens, K. & Tariq, S. The molecular and cellular biology of lung cancer: Identifying novel therapeutic strategies. *British Medical Bulletin* **95**, 47--61, doi:10.1093/bmb/ldq023 (2010).

39 *Fact Sheets by Cancer*, <http://gco.iarc.fr/today/data/factsheets/cancers/15-Lung-fact-sheet.pdf> (2018).

40 Kerr, K. M. Pulmonary adenocarcinomas: classification and reporting. *Histopathology* **54**, 12--27, doi:10.1111/j.1365-2559.2008.03176.x (2009).

41 Tate, J. G. *et al.* COSMIC: the Catalogue Of Somatic Mutations In Cancer. *Nucleic Acids Research* **47**, D941-D947, doi:10.1093/nar/gky1015 (2018).

42 Chesarino, N. M. & Emerman, M. Polymorphisms in human APOBEC3H differentially regulate ubiquitination and antiviral activity. *bioRxiv*, 2020.2002.2007.939439, doi:10.1101/2020.02.07.939439 (2020).

43 Feng, Y. *et al.* Natural Polymorphisms and Oligomerization of Human APOBEC3H Contribute to Single-

bioRxiv preprint doi: https://doi.org/10.1101/2020.02.28.970509; this version posted March 2, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

stranded DNA Scanning Ability. *J Biol Chem* **290**, 27188-27203, doi:10.1074/jbc.M115.666065 (2015).

44 Kidd, J. M., Newman, T. L., Tuzun, E., Kaul, R. & Eichler, E. E. Population stratification of a common APOBEC gene deletion polymorphism. *PLoS Genet* **3**, e63, doi:10.1371/journal.pgen.0030063 (2007).

45 Duggal, N. K., Malik, H. S. & Emerman, M. The breadth of antiviral activity of Apobec3DE in chimpanzees has been driven by positive selection. *J Virol* **85**, 11361-11371, doi:JVI.05046-11 [pii] 10.1128/JVI.05046-11 (2011).

46 Wittkopp, C. J., Adolph, M. B., Wu, L. I., Chelico, L. & Emerman, M. A Single Nucleotide Polymorphism in Human APOBEC3C Enhances Restriction of Lentiviruses. *PLoS Pathog* **12**, e1005865, doi:10.1371/journal.ppat.1005865 (2016).

47 Matsuoka, T. *et al.* Structural basis of chimpanzee APOBEC3H dimerization stabilized by double-stranded RNA. *Nucleic Acids Res* **46**, 10368-10379, doi:10.1093/nar/gky676 (2018).

48 Feng, Y. *et al.* RNA-Mediated Dimerization of the Human Deoxycytidine Deaminase APOBEC3H Influences Enzyme Activity and Interaction with Nucleic Acids. *J Mol Biol* **430**, 4891-4907, doi:10.1016/j.jmb.2018.11.006 (2018).

49 Bohn, J. A. *et al.* APOBEC3H structure reveals an unusual mechanism of interaction with duplex RNA. *Nat Commun* **8**, 1021, doi:10.1038/s41467-017-01309-6 (2017).

50 Ito, F. *et al.* Understanding the Structure, Multimerization, Subcellular Localization and mC Selectivity of a Genomic Mutator and Anti-HIV Factor APOBEC3H. *Sci Rep* **8**, 3763, doi:10.1038/s41598-018-21955-0 (2018).

51 Cescon, D. W., Haibe-Kains, B. & Mak, T. W. APOBEC3B expression in breast cancer reflects cellular proliferation, while a deletion polymorphism is associated with immune activation. *Proc Natl Acad Sci U S A* **112**, 2841-2846, doi:10.1073/pnas.1424869112 (2015).

52 Klonowska, K. *et al.* The 30 kb deletion in the APOBEC3 cluster decreases APOBEC3A and APOBEC3B expression and creates a transcriptionally active hybrid gene but does not associate with breast cancer in the European population. *Oncotarget* **8**, 76357-76374, doi:10.18632/oncotarget.19400 (2017).

53 Liu, J. *et al.* The 29.5 kb APOBEC3B Deletion Polymorphism Is Not Associated with Clinical Outcome of Breast Cancer. *PLoS One* **11**, e0161731, doi:10.1371/journal.pone.0161731 (2016).

54 Wen, W. X. *et al.* Germline APOBEC3B deletion is associated with breast cancer risk in an Asian multi-ethnic cohort and with immune cell presentation. *Breast Cancer Res* **18**, 56, doi:10.1186/s13058-016-0717-1 (2016).

55 Buisson, R. *et al.* Passenger hotspot mutations in cancer driven by APOBEC3A and mesoscale genomic features. *Science* **364**, doi:10.1126/science.aaw2872 (2019).

56 Barroso-Sousa, R. *et al.* Prevalence and mutational determinants of high tumor mutation burden in breast cancer. *Ann Oncol* **31**, 387-394, doi:10.1016/j.annonc.2019.11.010 (2020).

57 Harris, R. S. *et al.* The DNA cytosine deaminase APOBEC3B is a molecular determinant of platinum responsiveness in clear cell ovarian cancer. *Clin Cancer Res*, doi:10.1158/1078-0432.Ccr-19-2786 (2020).

58 Ooms, M. *et al.* HIV-1 Vif adaptation to human APOBEC3H haplotypes. *Cell Host Microbe* **14**, 411-421, doi:10.1016/j.chom.2013.09.006 (2013).

59 Refsland, E. W. *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* **10**, e1004761, doi:10.1371/journal.pgen.1004761 (2014).

60 Adolph, M. B. *et al.* Cytidine deaminase efficiency of the lentiviral viral restriction factor APOBEC3C correlates with dimerization. *Nucleic Acids Res* **45**, 3378-3394, doi:10.1093/nar/gkx066 (2017).

61 Duggal, N. K. & Emerman, M. Evolutionary conflicts between viruses and restriction factors shape immunity. *Nat Rev Immunol* **12**, 687-695, doi:10.1038/nri3295 (2012).

62 Law, E. K. *et al.* The DNA cytosine deaminase APOBEC3B promotes tamoxifen resistance in ER-positive breast cancer. *Sci Adv* **2**, e1601737, doi:10.1126/sciadv.1601737 (2016).

63 Cescon, D. W. & Haibe-Kains, B. DNA replication stress: a source of APOBEC3B expression in breast cancer. *Genome Biol* **17**, 202, doi:10.1186/s13059-016-1069-y (2016).

64 Green, A. M. *et al.* APOBEC3A damages the cellular genome during DNA replication. *Cell Cycle* **15**, 998-1008, doi:10.1080/15384101.2016.1152426 (2016).

bioRxiv preprint doi: https://doi.org/10.1101/2020.02.28.970509; this version posted March 2, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

65 Maier, J. A. *et al.* ff14SB: Improving the Accuracy of Protein Side Chain and Backbone Parameters from ff99SB. *Journal of Chemical Theory and Computation* **11**, 3696--3713, doi:10.1021/acs.jctc.5b00255 (2015).

66 Zgarbov. Refinement of the Sugar-Phosphate Backbone Torsion Beta for AMBER Force Fields Improves the Description of Z- and B-DNA. *Journal of Chemical Theory and Computation* **11**, 5723--5736, doi:10.1021/acs.jctc.5b00716 (2015).

67 Yildirim, I., Stern, H. A., Kennedy, S. D., Tubbs, J. D. & Turner, D. H. Reparameterization of RNA χ torsion parameters for the AMBER force field and comparison to NMR spectra for cytidine and uridine. *Journal of Chemical Theory and Computation* **6**, 1520--1531, doi:10.1021/ct900604a (2010).

68 Jorgensen, W. L., Chandrasekhar, J., Madura, J. D., Impey, R. W. & Klein, M. L. Comparison of simple potential functions for simulating liquid water. *The Journal of Chemical Physics* **79**, 926--935, doi:10.1063/1.445869 (1983).

69 Götz, A. W. *et al.* Routine Microsecond Molecular Dynamics Simulations with AMBER on GPUs. 1. Generalized Born. doi:10.1021/ct200909j (2012).

70 Salomon-Ferrer, R., Götz, A. W., Poole, D., Grand, S. L. & Walker, R. C. Routine Microsecond Molecular Dynamics Simulations with AMBER on GPUs. 2. Explicit Solvent Particle Mesh Ewald. doi:10.1021/ct400314y (2013).

71 Eastman, P. *et al.* OpenMM 7: Rapid development of high performance algorithms for molecular dynamics. *PLoS Computational Biology* **13**, e1005659, doi:10.1371/journal.pcbi.1005659 (2017).

72 Ito, F. *et al.* Understanding the Structure, Multimerization, Subcellular Localization and mC Selectivity of a Genomic Mutator and Anti-HIV Factor APOBEC3H. *Scientific Reports* **8**, 3763, doi:10.1038/s41598-018-21955-0 (2018).

73 Bateman, A. *et al.* UniProt: the universal protein knowledgebase. *Nucleic Acids Research* **45**, D158--D169, doi:10.1093/nar/gkw1099 (2017).

74 Hix, M. A. & Cisneros, G. A. Computational Investigation of APOBEC3H Substrate Orientation and Selectivity. doi:10.26434/chemrxiv.11916615.v2 (2020).

75 Bohn, J. A. *et al.* APOBEC3H structure reveals an unusual mechanism of interaction with duplex RNA. *Nature Communications* **8**, 1021, doi:10.1038/s41467-017-01309-6 (2017).

76 Pettersen, E. F. *et al.* UCSF Chimera--a visualization system for exploratory research and analysis. *J Comput Chem* **25**, 1605-1612, doi:10.1002/jcc.20084 (2004).

77 Feng, Y. *et al.* Natural polymorphisms and Oligomerization of Human APOBEC3H contribute to single-stranded DNA scanning ability. *Journal of Biological Chemistry* **290**, 27188--27203, doi:10.1074/jbc.M115.666065 (2015).

78 Case, D. A. *et al.* AMBER 2018, University of California, San Francisco. (2018).

79 Schafmeister, C. E. A. F., Ross, W. S. & Romanovski, V. Leap . University of California, San Francisco;. (1995).

80 Anandakrishnan, R., Aguilar, B. & Onufriev, A. V. H++ 3.0: automating pK prediction and the preparation of biomolecular structures for atomistic molecular modeling and simulations. *Nucleic Acids Research* **40**, W537-W541, doi:10.1093/nar/gks375 (2012).

81 Essmann, U. *et al.* A smooth particle mesh Ewald method. doi:1.470117 (1998).

82 Izaguirre, J. A., Sweet, C. R. & Pande, V. S. Multiscale dynamics of macromolecules using normal mode Langevin. *Pac Symp Biocomput*, 240-251 (2010).

83 Graham, S. E., Syeda, F. & Cisneros, G. A. Computational Prediction of Residues Involved in Fidelity Checking for DNA Synthesis in DNA Polymerase I. doi:10.1021/bi201856m (2012).

84 Elias, A. A. & Cisneros, G. A. in *Biomolecular Modelling and Simulations* Vol. 96 (ed Tatyana Karabencheva-Christova) 39-75 (Academic Press, 2014).

85 Dewage, S. W. & Cisneros, G. A. Computational Analysis of Ammonia Transfer Along Two Intramolecular Tunnels in Staphylococcus aureus Glutamine-Dependent Amidotransferase (GatCAB). doi:10.1021/jp5123568 (2015).

86 Case, D. A. *et al.* AMBER 2018. (2018).

87 Roe, D. R. & Cheatham, T. E. PTRAJ and CPPTRAJ: Software for Processing and Analysis of Molecular

bioRxiv preprint doi: https://doi.org/10.1101/2020.02.28.970509; this version posted March 2, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

Dynamics Trajectory Data. *Journal of chemical theory and computation* **9**, 3084--3095, doi:10.1021/ct400341p (2013).

88 Chelico, L., Prochnow, C., Erie, D. A., Chen, X. S. & Goodman, M. F. Structural model for deoxycytidine deamination mechanisms of the HIV-1 inactivation enzyme APOBEC3G. *J Biol Chem* **285**, 16195-16205, doi:10.1074/jbc.M110.107987 (2010).

89 Creighton, S., Bloom, L. B. & Goodman, M. F. Gel fidelity assay measuring nucleotide misinsertion, exonucleolytic proofreading, and lesion bypass efficiencies. *Methods Enzymol* **262**, 232-256 (1995).

90 Chelico, L., Pham, P., Calabrese, P. & Goodman, M. F. APOBEC3G DNA deaminase acts processively 3' --> 5' on single-stranded DNA. *Nat Struct Mol Biol* **13**, 392-399, doi:10.1038/nsmb1086 (2006).

91 Refsland, E. W. *et al.* Quantitative profiling of the full APOBEC3 mRNA repertoire in lymphocytes and tissues: implications for HIV-1 restriction. *Nucleic Acids Res* **38**, 4274-4284, doi:gkq174 [pii] 10.1093/nar/gkq174 (2010).