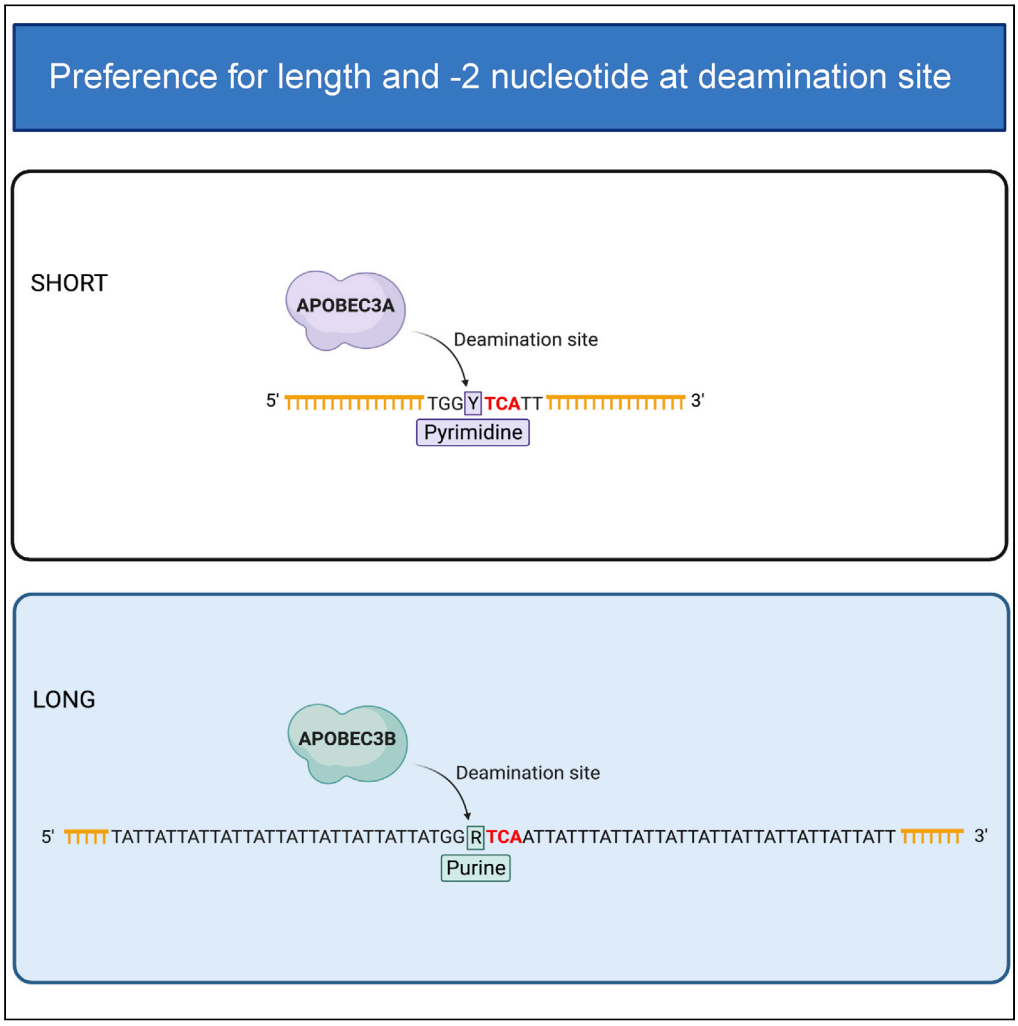
iScience

CellPress
OPEN ACCESS

Article

Exploring ABOBEC3A and APOBEC3B substrate specificity and their role in HPV positive head and neck cancer



Christina Papini, Zechen Wang, Shalley N. Kudalkar, ..., Natalia Issaeva, Wendell G. Yarbrough, Karen S. Anderson

natalia.isaeva@med.unc.edu (N.I.)
dell@med.unc.edu (W.G.Y.)
karen.anderson@yale.edu (K.S.A.)

Highlights

A3B is upregulated after 5-AzaC treatment and related to 5-AzaC sensitivity in HPV+ HNSCC

Full-length A3B prefers longer substrates and a purine at −2 site biochemically

A3B also prefers a purine at −2 site in both HPV+ and HPV− HNSCC cells

A3B signature at -2 site linked to poor patient survival in HPV+ HNSCC low smokers

Papini et al., iScience 25, 105077
October 21, 2022
https://doi.org/10.1016/j.isci.2022.105077

iScience

CellPress
OPEN ACCESS

Article

Exploring ABOBEC3A and APOBEC3B substrate specificity and their role in HPV positive head and neck cancer

Christina Papini,[1,5] Zechen Wang,[1,5] Shalley N. Kudalkar,[1] Travis Parke Schrank,[2] Su Tang,[1] Tomoaki Sasaki,[1] Cory Wu,[4] Brandon Tejada,[4] Samantha J. Ziegler,[4] Yong Xiong,[4] Natalia Issaeva,[2,3,*] Wendell G. Yarbrough,[2,3,*] and Karen S. Anderson[1,4,6,*]

SUMMARY

APOBEC3 family members are cytidine deaminases catalyzing conversion of cytidine to uracil. Many studies have established a link between APOBEC3 expression and cancer development and progression, especially APOBEC3A (A3A) and APOBEC3B (A3B). Preclinical studies with human papillomavirus positive (HPV+) head and neck squamous cell carcinoma (HNSCC) and clinical trial specimens revealed induction of A3B, but not A3A expression after demethylation. We examined the kinetic features of the cytidine deaminase activity for full length A3B and found that longer substrates and a purine at −2 position favored by A3B, whereas A3A prefers shorter substrates and an adenine or thymine at −2 position. The importance and biological significance of A3B catalytic activity rather than A3A and a preference for purine at the −2 position was also established in HPV+ HNSCCs. Our study explored factors influencing formation of A3A and A3B-related cancer mutations that are essential for understanding APOBEC3-related carcinogenesis and facilitating drug discovery.

INTRODUCTION

The Apolipoprotein B mRNA-editing enzyme-catalytic polypeptide 3 (APOBEC3, A3) family of proteins likely evolved as critical cellular defenders against pathogens, targeting foreign single stranded DNA (ssDNA) and hypermutating cytosine residues to uracil (Driscoll and Zhang, 1994; Harris et al., 2002; Navaratnam et al., 1993; Smith et al., 2012; Stenglein et al., 2010). However, the ability of these enzymes to create somatic mutations in the cellular genome has implicated them in cancer formation and maintenance (Verhalen et al., 2016) where A3-induced mutations account for up to 60% of total mutations in certain cancers (Alexandrov et al., 2013; Nik-Zainal et al., 2012; Roberts et al., 2013).

Amongst A3 family members, expression of A3A and A3B in the nucleus implicated them as the primary somatic mutagens with A3B, in particular, playing a significant role. A3A and A3B are also correlated with cancer severity and progression, suggesting that these proteins play a major role in tumorigenesis and tumor evolution (Burns et al., 2013a, 2013b; Kuong and Loeb, 2013). Determining the factors that influence cancer mutation formations by A3A and A3B is critical for further understanding of cancer pathogenesis and can inform drug discovery efforts targeting these proteins (Kelley et al., 2014; Rasmussen and Helin, 2016; Subramaniam et al., 2014; Wyatt, 2013). Consequentially, there have been many efforts to characterize ssDNA binding by A3 family members, particularly A3A and A3B (Barzak et al., 2019, 2021; Burns et al., 2013a, 2013b; Roberts et al., 2013; Shi et al., 2015, 2017b; Silvas et al., 2018; Wagner et al., 2019). For instance, A3A is thought to play a significant role in breast cancer (Burns et al., 2013a, 2013b; Taylor et al., 2013) whereas A3B has been suggested to be important in human papillomavirus expressing (HPV+) head and neck squamous cell carcinoma (HNSCC) (Cannataro et al., 2019). Our recent preclinical studies as well in our window clinical trial (NCT02178072) revealed that demethylation therapy upregulates A3B expression in HPV positive head and neck cancer patients that were treated with the drug, 5-Azacytidine (5-AzaC), an established inhibitor of DNA cytosine methylation (see below). These findings prompted the mechanistic and bioinformatic studies described here.

Analysis of mutations found in many tumor types has identified a recurring 5′-TCW-3′ motif in A3-induced mutations, where W is either a thymine or an adenine. It remains unclear whether this sequence is

[1]Department of Pharmacology, Yale University School of Medicine, 333 Cedar St., New Haven, CT 06520, USA

[2]Department of Otolaryngology/Head and Neck Surgery, University of North Carolina, Chapel Hill, NC 27599, USA

[3]Department of Pathology and Lab Medicine, Lineberger Cancer Center, University of North Carolina, Chapel Hill, NC 27599, USA

[4]Department of Molecular Biophysics and Biochemistry, Yale University, New Haven, CT 06520, USA

[5]These authors contributed equally

[6]Lead contact

*Correspondence: natalia.isaeva@med.unc.edu (N.I.), dell@med.unc.edu (W.G.Y.), karen.anderson@yale.edu (K.S.A.)

https://doi.org/10.1016/j.isci.2022.105077

determined at the kinetic level, or whether further substrate properties, such as length or neighboring residues, have effects on substrate preference (Burns et al., 2013a; Roberts et al., 2013; Wagner et al., 2019). Kinetic studies have begun to address these questions at a molecular level, reinforcing the importance of the 5′-TCA-3′ recognition motif for both A3A and the A3B catalytic domain, and demonstrating the preference of A3A for a purine in the +1 site (Barzak et al., 2019; Silvas et al., 2018). In addition, there has been significant structural analysis showing that A3 family members bind substrates in a U-shaped conformation, with the target cytidine and −1 thymine interacting with the active site (Barzak et al., 2021; Kouno et al., 2017; Shi et al., 2015, 2017b; Wagner et al., 2019). Although these studies represent significant progress in the characterization of ssDNA binding by A3A, there are many questions remaining about the enzymatic substrate preference of A3B.

Previous studies with A3B have focused primarily on the catalytic domain conserved among the A3 family, as this domain is almost entirely responsible for the catalytic activity of the protein (Barzak et al., 2019, 2021; Shi et al., 2015, 2017a, 2017b; Wagner et al., 2019). However, although A3A is comprised solely of this catalytic domain, A3B contains an additional N-terminal domain (Shandilya et al., 2014; Smith et al., 2012). This domain has little catalytic activity, and although it is thought to play a regulatory role, its function remains poorly understood (Lackey et al., 2013; Smith et al., 2012; Stenglein et al., 2008). Recent studies with A3B and A3G, a highly similar dual-domain A3 protein, have suggested that the N-terminal domains may play an important role in oligonucleotide binding and dimerization (Gorle et al., 2017; Maiti et al., 2020; Shandilya et al., 2014; Xiao et al., 2017; Yang et al., 2020). This is further supported by studies which have shown increased ssDNA binding, deaminase activity, and oligomerization in A3B catalytic domain constructs that do not truncate the Δ185-193 linker region between the two domains, and further suggests that this linker region may play an important role in these functions (Adolph et al., 2017). To further investigate the impact of the N-terminal domain on catalytic activity, here we perform a series of single enzyme turnover experiments using the full-length, wild-type A3B to examine substrate specificity around the 5′-TCA-3′ recognition motif in addition to the length of the DNA substrate. The studies are extended to a cellular level by further validating the sequence preference of A3B in a panel of six HNSCC cell lines, including both HPV-positive and HPV-negative HNSCC cells, and the clinical relevance of A3B in human tumors using bioinformatic analysis as compared with A3A.

RESULTS

Preclinical studies in HPV+ head and neck cancer cell lines and clinical trial specimens revealed upregulation of A3B after 5-AzaC treatment

We previously reported that HPV+ head and neck cancer cells are very sensitive to demethylation treatment with 5-AzaC (Biktasova et al., 2017; Hajek et al., 2020). The elevated sensitivity was partially because of restoration of tumor suppressors p53 and Rb functions (Biktasova et al., 2017), but also because of formation of replication-and transcription-dependent DNA double strand breaks that were found only in HPV+ head and neck cancer cells (Hajek et al., 2020). As shown in Figure 1, we examined a series of HPV+ and HPV− head and neck cancer cell lines after treatment with the nucleoside demethylating drug, 5-AzaC. We also examined patient tumor specimens pre- and post- 5-AzaC treatment from both HPV+ and HPV− HNSCC patients in a window clinical trial (NCT02178072) before surgery. As shown in Figure 1A for the cancer cell lines, the A3B mRNA level for HPV+ HNSCC cell line, UMSCC47, is more than 20-fold higher than that for the HPV− cancer cell line, UNC7. Treatment with 5-AzaC results in a further increase in A3B expression for the UMSCC47, whereas there is little effect of the drug in HPV− cancer cells. As shown with western blot analysis in Figure 1B, this is also the case at the level of protein expression as A3B is increased in HPV+ HNSCC line UMSCC47 on 5-AzaC treatment in a concentration-dependent manner, while completely absent in UMSCC47 CRISPR A3B knockout cells. Patient tumor samples in the clinical trial, pre and post 5-AzaC treatment were analyzed to assess relative mRNA expression of A3B. As shown in Figure 1C, in HPV+ patients, the endogenous mRNA levels of A3B were higher than HPV− tumors in 4 out of 5 HPV+ tumors and further increased on 5-AzaC treatment, whereas the A3B levels of mRNA were unchanged in the HPV− sample. A clonogenic survival assay revealed that the UMSCC47 HPV+ cell line was much more sensitive to treatment with 5-AzaC than the UMSCC47-A3B-7 CRISPR knockout where A3B was removed (Figure 1D).

Effect of Oligonucleotide Length on A3B cytidine deaminase activity

To have a better understanding of the function and the regulation of A3B, we first sought to determine the kinetic properties of A3B at a molecular level. To test the effect of ssDNA oligonucleotide length on A3B

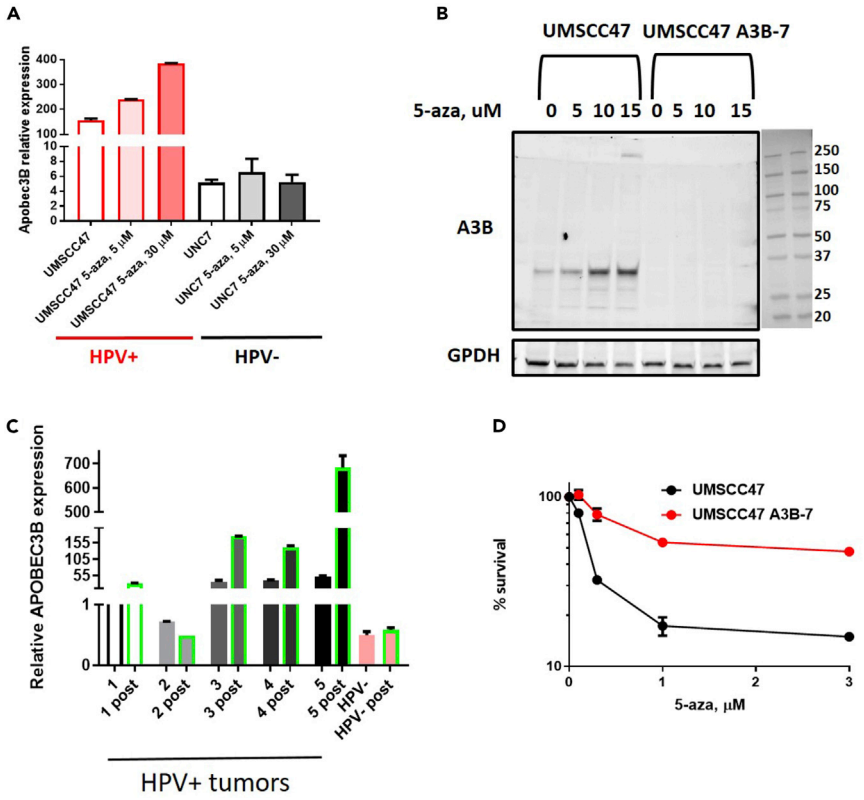


Figure 1. Preclinical and clinical trial specimens showing 5-AzaC upregulates A3B in HPV+ HNSCC

(A) A3B is upregulated in HPV+, but not in HPV− head and neck cancer cells after 5-AzaC treatment. Relative to GPDH mRNA, levels of A3B in cells after 5-AzaC treatment.

(B) A3B protein levels in HPV+ UMSCC47 and UMSCC47 A3B CRISPR cells (UMSCC47 A3B-7) after 5-AzaC treatment, GPDH shown as loading control.

(C) Relative mRNA levels of A3B in tumors before and after 5 or 7 days of 5-AzaC treatment from HNSCC patients enrolled in a window clinical trial. 1–5 represent tumors from five different HPV HNSCC patients; "post" indicates a tumor sample after 5-AzaC treatment.

(D) Clonogenic survival after 5-AzaC treatment of HPV+ UMSCC47 and APOBEC3B CRISPR (A3B-7) cells. Data are represented as mean ± SD.

cytidine deaminase activity, we designed and radiolabeled four oligonucleotides with the lengths of 10, 25, 43, and 70 nucleotides, respectively. Each of these constructs contained a centered 5′-TCA-3′ recognition motif derived from previous studies, with each substrate having an identical sequence context flanking the central recognition motif (Burns et al., 2013a). A series of single enzyme turnover experiments were carried out with purified full-length A3B (A3B-FL) protein (1 μM) in excess of oligonucleotide substrate (40 nM) using a previously described uracil DNA glycosylase (UDG)-dependent deaminase assay on these substrates (Ziegler et al., 2018) (Figure 2A, A3B-FL). Although the 43mer and 70mer underwent complete product formation by A3B within three hours, the 10mer and 25mer substrate reactions did not reach completion and instead plateaued at 46% and 60% product formation, respectively. The observed rate constants (k_{obs}) for each of these reactions were determined by fitting the data to a single exponential curve (Table 1). The k_{obs} increases with the length of the substrate oligonucleotide and appeared to reach a maximum rate with the 70mer substrate, indicating a preference for longer ssDNA by A3B. A graphical comparison of the rates is shown in Figure 2B, (A3B-FL).

Effect of Oligonucleotide Length on A3A Cytosine Deaminase Activity

We next examined whether a similar substrate length preference could be observed for the single-domain A3A. The same set of oligonucleotide substrates were assayed with purified A3A protein under similar single enzyme turnover conditions. Notably, the only substrate to not undergo complete product formation within 3 h was the longer 70mer, instead reaching a plateau at 51% product formation (Figure 2A, A3A).

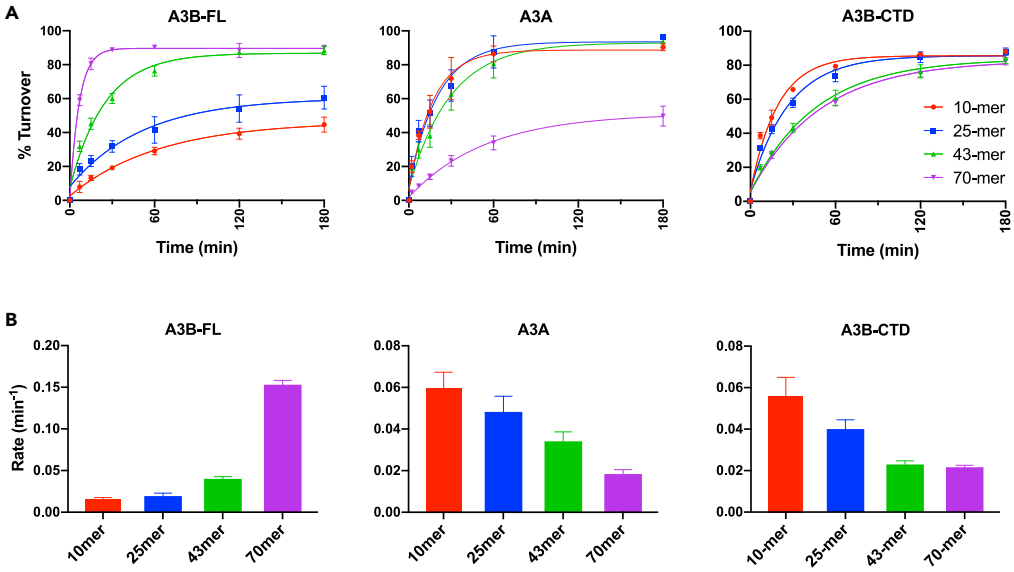


Figure 2. Effect of Oligonucleotide Length on A3B-FL, A3A, and A3B-CTD Cytosine Deaminase Activity

(A)Time course single-turnover kinetics of A3B-FL, A3A, and A3B-CTD against ssDNA substrates of varying lengths: 10mer, 25mer, 43mer, and 70mer.

(B) Graphical comparison of the deamination rates of the substrates for A3B-FL, A3A, and A3B-CTD. Data are represented as mean ± SD.

Moreover, as the length of the substrate decreased, k_{obs} increased, showing a clear preference of A3A for shorter ssDNA substrates (Table 1). This trend is opposite that observed with A3B, where longer ssDNA substrates were preferred (Figure 2B, A3A). This observation provides interesting context for previous experiments of A3A and A3B deamination activity and makes drawing direct comparisons difficult (Barzak et al., 2019; Silvas et al., 2018).

We propose that the difference in the ssDNA length preference between A3A and A3B is because of the extra non-catalytic N-terminal domain of A3B, which may give the enzyme additional affinity to the distal oligo nucleotides outside the recognition motif. To address this hypothesis, the same set of substrates were assayed against a truncated A3B construct which only contains the C-terminal catalytic domain (A3B-CTD) more closely approximately the length and domain structure of A3A. As shown in Figure 2A, A3B-CTD, unlike A3B-FL, favors shorter ssDNA, which is akin to A3A. A comparison of the rates is illustrated in Figure 2B, A3B-CTD and Table 1. This result suggests that the role of the N-terminal domain of A3B is to enhance affinity to longer ssDNA and improve overall catalytic efficiency *in vivo*.

Effect of ssDNA primary sequence on A3B cytidine deaminase activity

Although previous studies have proposed a 5′-TCW-3′ recognition motif, systematic kinetic analysis has been lacking so far to determine the full sequence preference within the active region of the ssDNA substrate and the neighboring residues (Burns et al., 2013a). We queried effects of the primary sequence by systematically surveying the −2, −1, +1, and +2 positions flanking the central cytidine on a set of radiolabeled 43mers. A series of single enzyme turnover experiments was performed using A3B-FL (1 μM) and each of the oligonucleotide 43mer substrates (40 nM) using the UDG-dependent deaminase assay. The reaction time courses for the various substrates are shown (Figures 3A–3D).

Kinetic analysis of the time courses demonstrated that A3B exhibits the strongest sequence preference at the −1 position (Figure 3B). The presence of a thymidine at the −1 site is heavily favored, showing a k_{obs} six-fold greater than that of cytidine as the next most favored nucleotide at the −1 position showing a preference for pyrimidine at this position (Table 2). Although the reaction with a thymidine went to completion within one hour, in the case of a cytidine, the reaction was 64% complete at 3 h. Oligonucleotide substrates with a purine at the −1 position underwent below 15% product formation and appeared to plateau over the

Table 1. Deamination rates of A3B-FL, A3A and A3B-CTD on substrates of varying lengths

Substrate	A3B-FL k_{obs} (min^{-1})	A3A k_{obs} (min^{-1})	A3B-CTD k_{obs} (min^{-1})
10mer	0.016 ± 0.002	0.059 ± 0.008	0.056 ± 0.009
25mer	0.019 ± 0.003	0.048 ± 0.008	0.040 ± 0.004
43mer	0.040 ± 0.003	0.034 ± 0.004	0.023 ± 0.002
70mer	0.153 ± 0.005	0.018 ± 0.002	0.022 ± 0.001

course of 3 h. Of interest, although cytosine, as a pyrimidine, is more favored than purines, the reaction with this construct similarly did not reach completion and instead reached a plateau at 64% product formation.

These experiments further demonstrate a preference for a purine at the −2 position, with a two-fold increase in k_{obs} for 43mer oligonucleotide substrates containing a purine at the −2 position over those containing a pyrimidine at the −2 position (Figure 3A). Although not as drastic a preference as that seen at the −1 position, this is a marked difference in kinetic activity.

Interestingly, there was very little difference in the determined k_{obs} values for the constructs with varying nucleotides at the +1 position (Figure 3C). Although there is a statistically significant preference for a purine at the +1 position with A3B, with thymidine being the least preferred, this preference does not affect the net amount of product formation and the reaction going to completion over the time course. This finding stands in contrast with studies that have previously identified APOBEC3 signatures consisting of a 5′-TCW-3′ motif, where thymidine would be preferred at the +1 position (Burns et al., 2013a; Roberts et al., 2013). However, this preference for a purine at the +1 position has similarly been observed for A3A, which also disfavors thymidine at this location (Silvas et al., 2018).

There appeared to be no preference for a specific base at the +2 position, with no functional impact on product formation by A3B (Figure 3D). This is in agreement with previous structural studies where no direct interactions have been observed between the +2 nucleotide and the A3B catalytic domain (Shi et al., 2017b).

Effect of ssDNA primary sequence on A3A and A3B-CTD cytidine deaminase activity at the −2 position

Although the observed sequence preferences for A3B at the −1, +1, and +2 positions match those previously reported for A3A (Silvas et al., 2018), in contrast the observed preference for A3B-FL at the −2 position led us to further consider this sequence preference for A3A and A3B-CTD. Using the same oligonucleotides as for our A3B-FL assays, we performed a series of single enzyme turnover experiments with A3A and A3B-CTD using the UDG-dependent deaminase assay to test the substrate sequence preference at the −2 position (Figure 4, Table 3) (Silvas et al., 2018). As illustrated in Figure 4A, for A3A, we observed a slight preference for adenine and thymidine over cytidine with guanidine being the most disfavored. Notably, this is in direct contrast to the previously reported preference for A3A, where a pyrimidine is favored at the −2 position (Silvas et al., 2018).

We further investigated the impact of the N-terminal domain of A3B on sequence preference at the −2 position by examining the A3B-CTD. As shown in Figure 4B, similar to the A3B-FL, the A3B-CTD construct also favors a purine at the −2 position with the k_{obs} rank of G>A>C/T, indicating that this unique preference by A3B is independent of the N-terminal domain.

Endogenous cytidine deaminase activity in HNSCC cells is largely attributable to A3B rather than A3A and also prefers purines over pyrimidines at the −2 position

There has been accumulating evidence showing that aberrant expression of A3A and A3B can be a marked contributor to mutagenesis for many cancer types (Balaji et al., 2021). As shown by our initial findings in HPV+ HNSCC patients and cancer cell lines described in Figure 1 as well as previous reports, the expression level of A3B is significantly higher in HPV+ HNSCC than in HPV− HNSCC, suggesting that A3B may be the dominant APOBEC family member driving mutations in this tumor type (Cannataro et al., 2019; Zapatka et al., 2020). To further explore the role that A3B plays in cytidine deaminase activity in HNSCC cell lines, we examined a panel of both HPV+ and HPV− cancer cells. These studies were designed to explore the

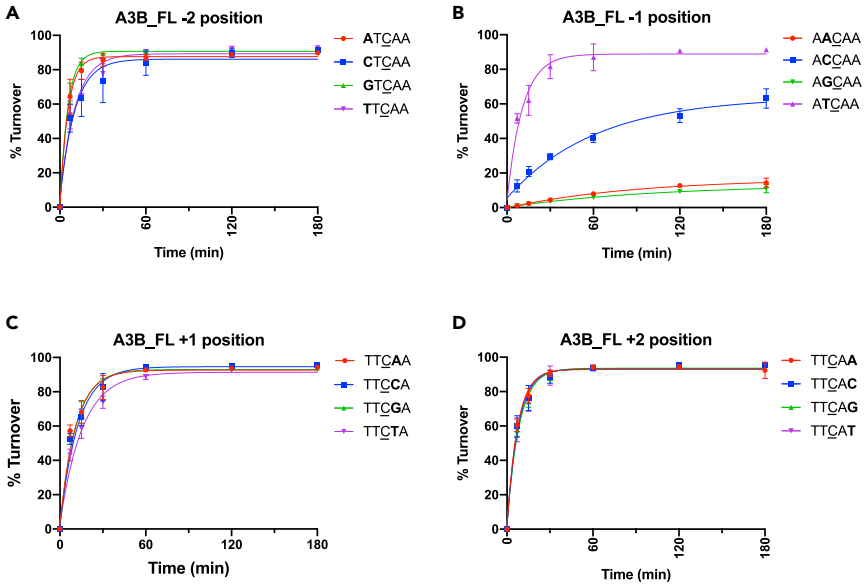


Figure 3. Effect of ssDNA Primary Sequence on A3B-FL Cytidine Deaminase Activity

(A) Effect of the identity of the nucleotide in the −2 position on the rate of A3B cytidine deaminase activity.

(B) Effect of the identity of the nucleotide in the −1 position on the rate of A3B cytidine deaminase activity. Constructs with a pyrimidine in the −1 position are deaminated more completely than those with a purine in the −1 position. Constructs with a thymine in the −1 position undergo complete product formation.

(C) Effect of the identity of the nucleotide in the +1 position on the rate of A3B deaminase activity.

(D) Effect of the identity of the nucleotide in the +2 position on the rate of A3B deaminase activity. Data are represented as mean ± SD.

primary sequence preference at the −2 position using 5′-FAM labeled purine (guanine) and pyridimine (thymine) 43mer substrates analogous to those used in the biochemical experiments described above. In addition, we wanted to investigate the relative contributions of cytidine deaminase activity for A3B versus A3A using siRNA or CRISPR to knockdown A3B expression.

Our results are described in Figure 5. We further confirmed the primary sequence preference at −2 position for A3B in a panel of six HNSCC cell lines, including both HPV+ (UMSCC47, UDSCC2, and YSCC94) and HPV− (SCC35, UNC521, and WSCC283) cells. We extracted the whole cell lysates and evaluated the cytidine deaminase activity using UDG-assay as described above with the 5′-FAM labeled DNA 43mers. In accordance with the kinetic results, DNA substrates with a guanine at −2 position had faster turnover of the 43mer substrate to the 30mer product than the corresponding thymine in 5 out of 6 cell lines (Figure 5A). Of note, the turnover of SCC35 cell lysate is below the detection level, compared to other cell lysates (Figure 5A). This is in line with the observation that A3B expression level is significantly lower in HPV− HNSCC, especially SCC35 (Figure 5B).

Using TCGA data for HNSCC, RNA expression of A3B transcripts in HNSCC tumors (Figure 5C) was correlated with substrate specific mutations for A3A and A3B and data displayed by HPV status (Chan et al., 2015)(also see STAR Methods). HPV+ HNSCC had higher expression of A3B, but no difference was noted in the number of tumors with A3A vs A3B mutations dominating. To further explore contributions of A3B versus other APOBEC family members such as A3A to the cytidine deaminase activity in HNSCC, we knocked out or knocked down A3B in all six HNSCC cell lines by CRIPSR (UMSCC47) or siRNA (UDSCC2, YSCC94, SCC35, UNC521, and WSCC283), respectively (Figure 5D). A3B was successfully knocked down or knocked out, whereas the expression level of A3A was not affected (Figure 5D). Compared to control CRISPR or siRNA cells, loss of A3B drastically decreased the turnover of DNA substrates regardless of guanine or a thymine at −2 position to below the detection level. (Figures 5E and 5F). These findings further verify that A3B prefer a purine over a pyrimidine at −2 position and A3B is the major source of cytidine deaminase activity in the HNSCC cell lines we examined.

Table 2. Deamination rates of A3B on substrates varying the identity of the nucleotide in the −2, −1, +1, and +2 positions

Substrate	k_{obs} (min^{-1})
−2 position	
ATCAA	0.180 ± 0.012
CTCAA	0.096 ± 0.012
GTCAA	0.182 ± 0.007
TTCAA	0.094 ± 0.008
−1 position	
AACAA	0.011 ± 0.002
ACCAA	0.016 ± 0.002
AGCAA	0.010 ± 0.002
ATCAA	0.093 ± 0.008
+1 position	
TTCAA	0.091 ± 0.007
TTCCA	0.078 ± 0.005
TTCGA	0.091 ± 0.005
TTCTA	0.063 ± 0.003
+2 position	
TTCAA	0.137 ± 0.006
TTCAC	0.130 ± 0.008
TTCAG	0.118 ± 0.005
TTCAT	0.125 ± 0.008

Contribution of A3A and A3B to mutations and correlation with survival in HNSCC

Following the biochemical analysis, which determined the effects of primary sequence at −2 position on A3A and A3B cytidine deaminase activity, we sought to understand whether this sequence preference at −2 position contributes to patterns of mutagenesis for the tumors found in cancer patients by examining HNSCC tumors profiled by The Cancer Genome Atlas (TCGA). In total, 180 HPV-negative and 39 HPV+ tumors with both RNA expression data and genomic variant calls were identified, and potential APOBEC related (Xt[c]a > t/g) mutations were included in our analysis.

Consistent with our kinetic data which demonstrates guanine is the least favored at −2 position for A3A, and A3B favoring A and G at the −2 position; we classified HNSCC tumors to one of two classes based on the dominance of the A3B (A/Gt[c]a > t/g) vs. A3A (T/Ct[c]a > t/g) mutational signature in that tumor, see STAR Methods. 47% of tumors displayed the A3B signature, and this proportion was similar in HPV+ and HPV− HNSCC. Of interest, A3A signature tumors harbored more APOBEC related mutations (Figure 6A). To determine biological significance and potential translational importance of expanded APOBEC signatures that can delineate A3A from A3B mutagenesis, we examined progression-free survival of HNSCC patients using TCGA data (Figures 6B and 6C). Of interest, an A3B mutational signature has no prognostic significance in HPV− HNSCC (Figure 6B), but was strongly associated with poor patient outcomes in HPV+ HNSCC (Figure 6C). To examine the relative value of the A3B/A3A mutational signatures as compared to other related metrics, we also examined survival in HPV+ HNSCC stratifying by median expression of A3A and A3B, both of which failed to demonstrate a significant relationship. Similarly, stratifying HPV+ HNSCC tumors by the total number of (Xt[c]a > t/g) mutations was not significantly related to progression-free survival.

DISCUSSION

Our preclinical and clinical trial results in head and neck cancer pointing to a role for A3B in HPV+ HNSCC toxicity to demethylation served as a catalyst for the current work to further investigate substrate specificity. This study is the one of the first to explore of the impact of various substrate properties on A3B catalytic activity versus A3A using a full-length, wild-type A3B and the truncated C-terminal catalytic

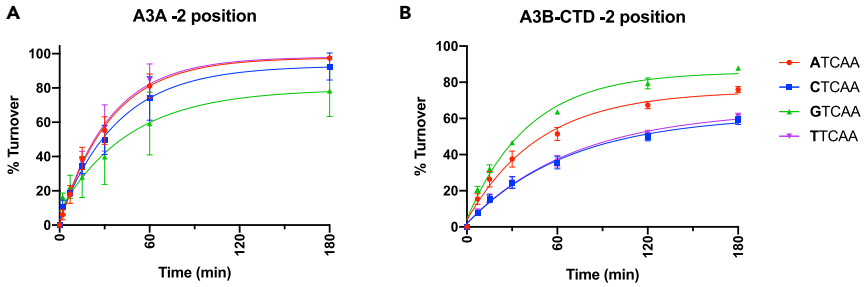


Figure 4. Effect of the identity of the nucleoside at the -2 position on A3A and A3B-CTD cytidine deaminase activity

(A) Effect of the identity of the nucleotide in the −2 position on the rate of A3A cytidine deaminase activity.

(B) Effect of the identity of the nucleotide in the −2 position on the rate of A3B-CTD cytidine deaminase activity. Data are represented as mean ± SD.

domain A3B-CTD constructs in a series of single enzyme turnover experiments. Our kinetic studies are in accord with earlier studies that were among the first to purify full-length A3B and demonstrate much higher catalytic efficiency as compared with A3B-CTD (Adolph et al., 2017). The current study provides unique insight into the selection of A3B substrates on a kinetic level, further informing the mechanisms of A3B-induced mutagenesis. Longer ssDNA substrates were deaminated more completely and at a faster rate than shorter substrates by A3B as compared with A3A or the truncated A3B-CTD. This would suggest that longer strands of exposed single stranded DNA are required by A3B for more efficient turnover, and therefore that *in vivo* substrates are more likely to occur in areas of the genome that contain extensive, more frequently exposed regions of ssDNA. These kinetic results are consistent with previous findings that A3B mutations are more prevalent on the lagging strand, where long stretches of ssDNA are exposed during the replication process (Verhalen et al., 2016). The study by Adolph et al., (2017) confirmed that A3B could move over double-stranded DNA to complete processive deaminations by intersegmental transfer.

As previous structural studies of A3B have only shown the 5′-TCA-3′ region forming strong interactions with the catalytic domain active site, the underlying cause for this preference, is not immediately clear (Shi et al., 2017b). However, more recent studies have shown that a tryptophan residue, Trp127 in the N-terminal domain plays an important role in ssDNA binding (Xiao et al., 2017). Trp127 is conserved among the dual-domain A3 family members, and has been shown to play an important role in A3G dimerization. The A3G dimer interface has been suggested to form part of an ssDNA binding channel that stretches between the catalytic C-terminal and non-catalytic N-terminal domains (Maiti et al., 2020; Shandilya et al., 2014; Yang et al., 2020). Of interest, mutations preventing dimerization showed a significant decrease in nucleic acid association, suggesting that although the presence of the N-terminal domain allows for dimerization and the formation of this binding channel, the N-terminal domain does not itself improve nucleic acid association (Gorle et al., 2017; Maiti et al., 2020). The high degree of sequence similarity and dimerization ability shared by A3B and A3G leads us to propose there may be a similar channel formed by dimerized full-length A3B that assists in binding to extended regions of ssDNA and increasing overall catalytic activity. Previous studies have shown that another APOBEC dual family member, A3F, also prefers longer DNA substrates (Adolph et al., 2017; Schutsky et al., 2018). In contrast, A3A more readily deaminates shorter ssDNA substrates, which is consistent with the structural information showing that this enzyme contains only one domain and exists primarily as a monomer (Maiti et al., 2020; Shandilya et al., 2014; Yang et al., 2020).

With regard to substrate sequence preference for A3B, it seems like there is no clear preference for a specific base at the +2 position, which is in line with previous structural studies that showed no direct interactions between the +2 nucleotide and the A3B catalytic domain (Shi et al., 2017b). In addition, although the proposed ssDNA binding channel for dual-domain A3 proteins would be in contact with the 3′ region of the DNA substrate, the highly positive electrostatic surface potential of this channel would result primarily in interactions with the ssDNA phosphate backbone (Maiti et al., 2020; Yang et al., 2020). Together, these results suggest that any further binding of the 3′ ssDNA is not done in a sequence specific manner and would not impact substrate preference.

Table 3. Deamination rates of A3A and A3B-CTD on substrates varying the identity of the nucleotide in the −2 position

Substrate	A3A k_{obs} (min^{-1})	A3B-CTD k_{obs} (min^{-1})
ATCAA	0.030 ± 0.003	0.021 ± 0.004
CTCAA	0.026 ± 0.003	0.015 ± 0.003
GTCAA	0.022 ± 0.006	0.024 ± 0.003
TTCAA	0.030 ± 0.003	0.014 ± 0.003

Although bioinformatics studies have identified a prevalent 5′-TCW-3′ motif in APOBEC3-induced mutations, our data shows that A3B has little difference in observed rate constant for substrates varying the identity at the +1 position, with a slight preference for a purine (Burns et al., 2013a; Roberts et al., 2013; Wagner et al., 2019). In fact, these results identify thymine as the least preferred residue at the +1 site, similar to the observed sequence preference of A3A, but contrary to the identified 5′-TCW-3′ motif (Silvas et al., 2018). This discrepancy between kinetic preference and observed mutations would suggest that there are additional regulatory mechanisms of A3B deamination *in vivo* or other APOBEC3 family members that contribute in determining the 5′-TCW-3′ motif, instead of A3B. Indeed, studies have shown that although A3A prefers deamination substrates containing hairpin secondary structure and disfavors the long exposed ssDNA substrates found in the transcription bubble, this preference is exacerbated by the addition of whole cell lysates (Brown et al., 2021). Recent studies indicate that "passenger mutational hotspots" with A3A are because of enzyme substrate preference rather than effects on tumors (Buisson et al., 2019). Taken together, these results clearly demonstrate that substrate preference of A3A is further driven by interactions with other ssDNA binding proteins within the cell, and it is likely that similar factors play a role in cellular A3B substrate preference as well.

Varying the identity of the nucleotide in the −1 position has confirmed the necessity of a thymine in the −1 position for efficient catalysis of A3B. Substrates with a purine in the −1 position underwent minimal deamination, and even the presence of a −1 cytidine, another pyrimidine, resulted in significantly decreased activity. The significant difference in product formation and deamination activity between thymine and cytosine suggest that the carbonyl groups on the −1 thymine may form stabilizing interactions near the protein active site, promoting efficient activity. Indeed, previous structural studies have shown the −1 thymine forming key interactions with a binding pocket next to the active site, further supporting the necessity of a thymine in the −1 site for efficient turnover (Shi et al., 2017b).

The current study revealed some interesting differences in substrate specificity at the −2 position between A3B and A3A. The kinetic results demonstrated a two-fold increase in observed rate constant of A3B when the nucleotide substrate contains a purine rather than pyrimidine at the −2 position. This was also the case for the catalytic A3B-CTD confirming that the N-terminal domain does not play a role in defining sequence specificity. These results are also consistent with previous NMR studies examining the −2 position (Liu et al., 2018b). On the other hand, for A3A, adenine and thymine are slightly favored over cytosine whereas guanine is the most disfavored. Other studies examining the sequence preference for A3A, suggest a pyrimidine is preferred in the −2 position, particularly cytosine (Silvas et al., 2018; Ziegler et al., 2018). Although A3B-related mutational signature was not prognostic in HPV- negative HNSCC (Figure 6B), it indicated poor patient survival in HPV+ HNSCC (Figure 6C), emphasizing its biological significance in HPV-associated cancer.

A number of modeling studies have examined differences in various APOBEC cytidine deaminases with the structures available (Maiti et al., 2021; Ng and Fraternali, 2020; Salter and Smith, 2018). We sought to understand the −2 position preference structurally. According to the crystal structure of A3A bound to oligonucleotide (PDB code: 5SWW, (Shi et al., 2017a)), it is clear that the preference of the −2 position is dictated by the Loop 1, one of the three loops that consist the enzyme active site (Kouno et al., 2017; Shi et al., 2017b). However, unlike the 0 and −1 nucleotide binding pockets which are well defined in the substrate binding groove, the −2 nucleotide binding pocket is located on the protein surface and is very shallow (Shi et al., 2017a). This is in agreement with the observed weaker selectivity for the −2 position compared with 0 and −1 positions. Previous NMR studies with A3B-CTD have noted the importance of Loop 1 (Byeon et al., 2016). Notably, although A3B-CTD and A3A share ~90% sequence identity, the former contains a longer Loop 1 with a unique ^{210}R-R-R^{212} motif that presumably governs −2 position preference. So far, the crystal

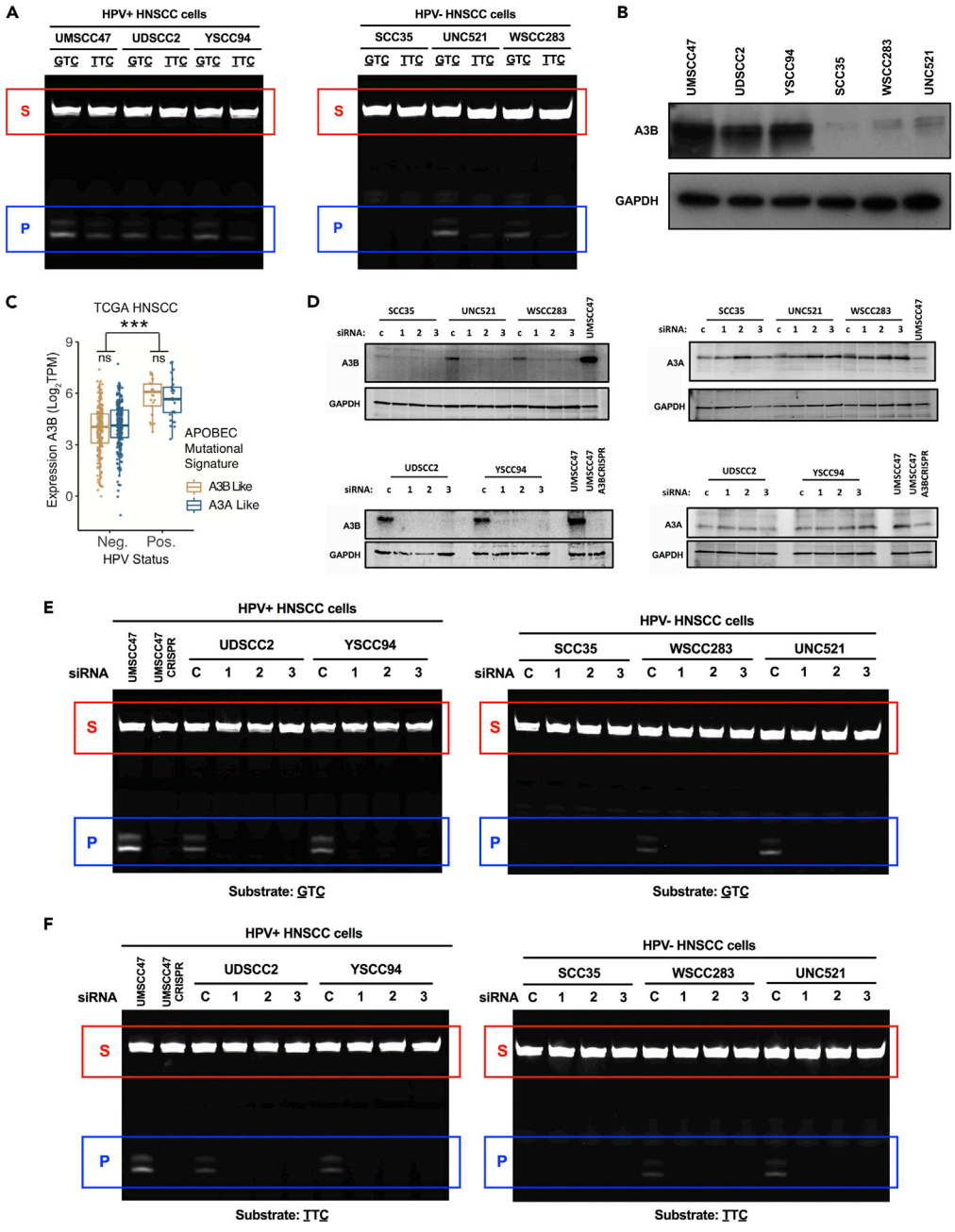


Figure 5. Endogenous A3B from HNSCC cell lines confirms preference for purine at -2 position

(A) Whole cell lysates from five out of six HNSCC cell lines prefer a guanine at −2 position over a thymine, except for SCC35 cell lysate, whose cytidine deaminase activity is undetectable. Nucleotides at −2 position and the cytidine that is deaminated are indicated by underscore. Red box designates substrate (S) and blue box product (P).

(B) Western blot showing the expression level of A3B in all six HNSCC cell lines. A3B expression in SCC35 cells is below the detection limit, and HPV-positive cells (UMSCC47, UDSCC2, and YSCC94) have higher A3B expression compared to HPV-negative cells (SCC35, UNC521, and WSCC283).

(C) Correlation of APOBEC3B expression with mutational profiles in HNSCC tumors. Tumors were classified by HPV status and substrate specificity for A3A and A3B. Xt[c]a > t/g polymorphisms were considered APOBEC related. Mutational signatures from individual tumors with >=50% A/Gt[c]a > t/g mutations were considered A3B like; and conversely tumors with >50% T/Ct[c]a > t/g designated A3A like, see also STAR Methods. Wilcoxon rank sum test. ns – not significant. *** p-value < 5*10^-3. TPM – transcripts per million.

Figure 5. *Continued*
(D) Western blot showing A3B expression in A3B CRISPR and A3B siRNA clones. A3A expression is not significantly altered. C = siRNA control.
(E) A3B knockout and knockdown decrease the cytidine deaminase activity to below the detection level with the DNA substrate having a guanine at −2 position. DNA substrates (S) and products (P) are indicated in red and blue boxes, respectively. C = siRNA control.
(F) A3B knockout and knockdown also decrease the turnover of the DNA substrate with a thymine at −2 position to below the detection level. DNA substrates (S) and products (P) are indicated in red and blue boxes, respectively. C, siRNA control.

structure of A3B-CTD with an intact Loop 1 has only been solved in the apo form (5CQK) where the active site is in closed conformation not ready for substrate binding (Shi et al., 2015). The NMR solution structure of apo A3B-CTD (PDB code: 2NBQ) shows that Loop 1 is highly flexible and adopts multiple conformations (Byeon et al., 2016; Shi et al., 2017b). It is not immediately clear how Loop 1 will move on substrate binding and how the consecutive arginine residues can determine −2 selectivity. Unfortunately, the only A3B-CTD structures available have a modified Loop 1 similar to A3A precluding definitive assessments (Shi et al., 2017a). Therefore, the co-crystal structure of A3B maintaining the native Loop 1 with substrate bound is highly desired to fully understand the structural basis of selectivity at this position.

Our studies also examined cytidine deaminase catalytic activity at the cellular level in HPV+ and HPV− HNSCC cells using two of the substrates used in biochemical studies probing the −2 position containing a purine (guanine) or pyrimidine (thymine). The panel of cell lines examined included ones in which A3B was either knocked out with CRISPR or knocked down using siRNA to establish the relative catalytic contributions of A3B versus A3A. In the cell lines where A3B expression was reduced or knocked out there was no catalytic turnover observed confirming the importance of A3B. Notably, the guanine substrate was converted to product much more rapidly than the thymine substrate in agreement with the findings at the biochemical level reflecting the sequence preference for A3B. Taken together, these results affirm the catalytic role for A3B rather than A3A in HNSCC. The finding that more HNSCC tumors had a majority of mutations with A3A sequence preference combined with finding that loss of A3B markedly decreased APOBEC catalytic activity in HNSCC, suggest that although A3B is maintained in these tumors, A3A or other APOBEC activity may episodically contribute to mutations (Granadillo Rodriguez et al., 2020). Earlier studies in yeast offer *in vivo* context for our kinetic results. In those studies, the mutational signatures in yeast that might arise from deamination with A3B versus A3A to define sequence preferences were examined (Chan et al., 2015). Their *in vivo* findings in yeast mirrored our *in vitro* biochemical studies and cellular studies showing a preference for purine at the −2 position for A3B and a pyrimidine for A3A (Chan et al., 2015). Additional studies with A3A in *E. coli* also found preference for pyrimidine (Ziegler et al., 2019).

Taken together, these findings show clear similarities and differences in substrate preferences between A3A and A3B, and suggest mechanisms involving both the active site binding pocket and the role of the N-terminal domain to explain the observed differences at a molecular level. A3B expression is increased in cells on infection with HPV and increased expression is maintained in HPV+ HNSCC (Cannataro et al., 2019; Zapatka et al., 2020). Consistent with these finding, our studies suggest that A3B contributes the majority of ongoing cytidine deaminase activity in this tumor subtype. An unusual dependence of HPV+ cancers on APOBEC activity, which may explain their preserved expression, has been noted (Swanton et al., 2015), but the basis of this connection is not well understood and requires further exploration. The current study provides critical insight into the underlying molecular mechanisms of APOBEC3-driven mutations and their role in HPV+ HNSCC and may reveal possible biomarkers to guide design of clinical trial to develop more effective therapeutic regimens.

Limitations of the study

In this study, we demonstrated that A3B prefers longer ssDNA substrates and a purine at −2 position using recombinant enzyme and endogenously expressed in cell lysates. Bioinformatic analysis further established the correlation between with predominate A3B mutation profile and poor patient overall survival in HPV+ patients with low or moderate smoking. However, the underlying molecular mechanism of the preference at −2 position for A3B still requires further exploration. Furthermore, it is still not clear whether A3B substrate preference at −2 position can determine which genes are more susceptible to A3B cytidine deaminase activity, especially for those oncogenes that are frequently mutated in HPV-positive HNSCC. We exercise caution in trying to draw direct parallels between the kinetic features of A3B observed in

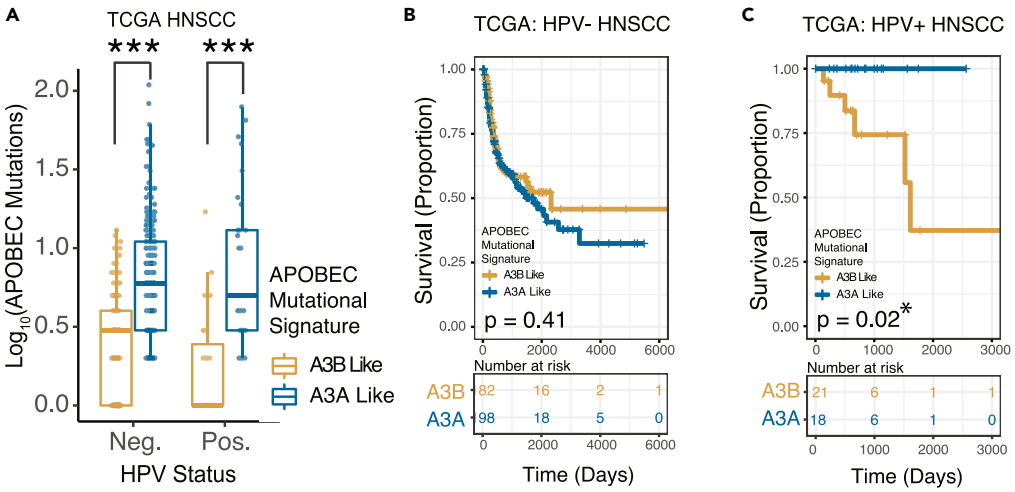


Figure 6. Analysis of A3A and A3B mutational signatures in TCGA HNSCC

Tumors were classified by HPV status and substrate specificity for A3A and A3B. Xt[c]a > t/g polymorphisms were considered APOBEC related. Mutational signatures from individual tumors with >=50% A/Gt[c]a > t/g mutations were considered A3B like; and conversely tumors with >50% T/Ct[c]a > t/g designated A3A like, see also STAR Methods.

(A) APOBEC mutations per HNSCC tumor in the TCGA, stratified by HPV status and A3A vs. A3B mutational signature. Log10(n APOBEC mutations + 1) is plotted for clarity. Wilcoxon rank sum test. *** p-value < 5*10^-3.

(B and C) Progression-free Survival of HPV-negative (B) and HPV-positive (C) HNSCC by APOBEC mutation profile. HPV+ HNSCC included oropharyngeal cancers in patients with up to 30 pack year tobacco smoke exposure, see STAR Methods.

biochemical and cellular assays and previous bioinformatic analysis of APOBEC mutational signatures. For instance, bioinformatics studies identified 5′-TCW-3′ motif as APOBEC3-induced mutations, whereas our data shows that A3B has little difference in k_{obs} for various nucleotides at the +1 position, with a slight preference for a purine. Therefore, we hypothesize that other than the sequence preference of A3B itself, A3B activities can also be regulated through other mechanisms in cells and *in vivo*. This is also likely the case for A3A especially in light of more recent findings showing the preference of hairpin substrates, which requires further study. With more complex factors influencing mutational signatures and regulation of APOBEC proteins at a cellular level and in patient tumors, a full understanding of the underlying mechanisms requires further investigation.

STAR★METHODS

Detailed methods are provided in the online version of this paper and include the following:

- KEY RESOURCES TABLE
- RESOURCE AVAILABILITY
  - Lead contact
  - Materials availability
  - Data and code availability
- EXPERIMENTAL MODEL AND SUBJECT DETAILS
  - Human samples
  - Cell culture
  - Microbe strains
- METHOD DETAILS
  - A3B expression and purification
  - A3B-CTD expression and purification
  - A3A expression and purification
  - Design and preparation of oligonucleotides
  - Biochemical cytidine deaminase activity assay
  - Imaging and kinetic data analysis
  - A3B knockout/knockdown and cell lysate preparation
  - Immunoblotting
  - Clonogenic survival assay

- RT-qPCR
- Cytidine deaminase activity assay with cell lysates
- RNA expression data
- Mutational data
- Cohort selection and inclusion criteria
- Correlation analysis
- Survival analysis
- APOBEC signature analysis

- QUANTIFICATION AND STATISTICAL ANALYSIS
- ADDITIONAL RESOURCES

SUPPLEMENTAL INFORMATION

Supplemental information can be found online at https://doi.org/10.1016/j.isci.2022.105077.

ACKNOWLEDGMENT

This study is funded by NIH/NIDCR R01 to K.S.A., W.G.Y., and N.I., NIH/NIDCR P50DE030707 to K.S.A. and W.G.Y, NIH/NIAIDR 37AI116313 to Y.X. and China Scholarship Council–Yale World Scholars Fellowship to Z.W.

AUTHOR CONTRIBUTIONS

Conceptualization, K.S.A., C.P., Z.W., S.K., N.I., W.G. Y., and T.S.; Methodology, C.P., Z.W., S.K., T.S., T.P.S., C.W., S.T., N.I., W.G. Y., and B.T.; Validation, C.*P. and* Z.W.; Formal Analysis, C.P., Z.W., S.K., S.T., N.I., and T.P.S.; Investigation, C.P., Z.W., S.K., and T.P.S.; Resources, S.Z. and Y.X.; Writing – Original Draft, C.P., Z.W., S.K., and T.P.S.; Writing – Review and Editing, K.S.A., N.I., W.G.Y., C.P., Z.W., S.T., and T.P.S; Y.X. Supervision, K.S.A., W.G.Y., and N.I.; Project administration, K.S.A., W.G.Y., and N.I.; Funding acquisition, K.S.A., W.G.Y., and N.I.

DECLARATION OF INTERESTS

The authors declare no competing interests.

INCLUSION AND DIVERSITY

We support inclusive, diverse, and equitable conduct of research.

Received: February 7, 2022
Revised: August 5, 2022
Accepted: August 31, 2022
Published: October 21, 2022

REFERENCES

Adolph, M.B., Love, R.P., Feng, Y., and Chelico, L. (2017). Enzyme cycling contributes to efficient induction of genome mutagenesis by the cytidine deaminase APOBEC3B. Nucleic Acids Res. *45*, 11925–11940.

Alexandrov, L.B., Nik-Zainal, S., Wedge, D.C., Aparicio, S.A.J.R., Behjati, S., Biankin, A.V., Bignell, G.R., Bolli, N., Borg, A., Børresen-Dale, A.L., et al. (2013). Signatures of mutational processes in human cancer. Nature *500*, 415–421.

Balaji, H., Demers, I., Wuerdemann, N., Schrijnder, J., Kremer, B., Klussmann, J.P., Huebbers, C.U., and Speel, E.J.M. (2021). Causes and consequences of HPV integration in head and neck squamous cell carcinomas: state of the art. Cancers *13*, 4089.

Barzak, F.M., Harjes, S., Kvach, M.V., Kurup, H.M., Jameson, G.B., Filichev, V.V., and Harjes, E. (2019). Selective inhibition of APOBEC3 enzymes by single-stranded DNAs containing 2′-deoxyzebularine. Org. Biomol. Chem. *17*, 9435–9441.

Barzak, F.M., Ryan, T.M., Kvach, M.V., Kurup, H.M., Aihara, H., Harris, R.S., Filichev, V.V., Harjes, E., and Jameson, G.B. (2021). Small-angle X-ray scattering models of APOBEC3B catalytic domain in a complex with a single-stranded DNA inhibitor. Viruses *13*, 290.

Biktasova, A., Hajek, M., Sewell, A., Gary, C., Bellinger, G., Deshpande, H.A., Bhatia, A., Burtness, B., Judson, B., Mehra, S., et al. (2017). Demethylation therapy as a targeted treatment for human papillomavirus-associated head and neck cancer. Clin. Cancer Res. *23*, 7276–7287.

Brown, A.L., Collins, C.D., Thompson, S., Coxon, M., Mertz, T.M., and Roberts, S.A. (2021). Single-stranded DNA binding proteins influence APOBEC3A substrate preference. Sci. Rep. *11*, 21008.

Buisson, R., Langenbucher, A., Bowen, D., Kwan, E.E., Benes, C.H., Zou, L., and Lawrence, M.S. (2019). Passenger hotspot mutations in cancer driven by APOBEC3A and mesoscale genomic features. Science *364*, eaaw2872.

Burns, M.B., Lackey, L., Carpenter, M.A., Rathore, A., Land, A.M., Leonard, B., Refsland, E.W., Kotandeniya, D., Tretyakova, N., Nikas, J.B., et al. (2013a). APOBEC3B is an enzymatic source of mutation in breast cancer. Nature *494*, 366–370.

Burns, M.B., Temiz, N.A., and Harris, R.S. (2013b). Evidence for APOBEC3B mutagenesis in multiple human cancers. Nat. Genet. *45*, 977–983.

Byeon, I.J.L., Byeon, C.H., Wu, T., Mitra, M., Singer, D., Levin, J.G., and Gronenborn, A.M. (2016). Nuclear magnetic resonance structure of the APOBEC3B catalytic domain: structural basis for substrate binding and DNA deaminase activity. Biochemistry *55*, 2944–2959.

Cannataro, V.L., Gaffney, S.G., Sasaki, T., Issaeva, N., Grewal, N.K.S., Grandis, J.R., Yarbrough, W.G., Burtness, B., Anderson, K.S., and Townsend, J.P. (2019). APOBEC-induced mutations and their cancer effect size in head and neck squamous cell carcinoma. Oncogene *38*, 3475–3487.

Chan, K., Roberts, S.A., Klimczak, L.J., Sterling, J.F., Saini, N., Malc, E.P., Kim, J., Kwiatkowski, D.J., Fargo, D.C., Mieczkowski, P.A., et al. (2015). An APOBEC3A hypermutation signature is distinguishable from the signature of background mutagenesis by APOBEC3B in human cancers. Nat. Genet. *47*, 1067–1072.

Deng, M., Bragelmann, J., Kryukov, I., Saraiva-Agostinho, N., and Perner, S. (2017). FirebrowseR: an R Client to the Broad Institute's Firehose Pipeline. Database (Oxford) *2017*, baw160.

Driscoll, D.M., and Zhang, Q. (1994). Expression and characterization of p27, the catalytic subunit of the apolipoprotein B mRNA editing enzyme. J. Biol. Chem. *269*, 19843–19847.

Ellrott, K., Bailey, M.H., Saksena, G., Covington, K.R., Kandoth, C., Stewart, C., Hess, J., Ma, S., Chiotti, K.E., McLellan, M., et al. (2018). Scalable open science approach for mutation calling of tumor exomes using multiple genomic pipelines. Cell Syst. *6*, 271–281.e7.

Goldman, M.J., Craft, B., Hastie, M., Repečka, K., McDade, F., Kamath, A., Banerjee, A., Luo, Y., Rogers, D., Brooks, A.N., et al. (2020). Visualizing and interpreting cancer genomics data via the Xena platform. Nat. Biotechnol. *38*, 675–678.

Gorle, S., Pan, Y., Sun, Z., Shlyakhtenko, L.S., Harris, R.S., Lyubchenko, Y.L., and Vuković, L. (2017). Computational model and dynamics of monomeric full-length APOBEC3G. ACS Cent. Sci. *3*, 1180–1188.

Granadillo Rodríguez, M., Flath, B., and Chelico, L. (2020). The interesting relationship between APOBEC3 deoxycytidine deaminases and cancer: a long road ahead. Open Biol. *10*, 200188.

Hajek, M., Biktasova, A., Sewell, A., Gary, C., Cantalupo, P., Anderson, K.S., Yarbrough, W.G., and Issaeva, N. (2020). Global genome demethylation causes transcription-associated DNA double strand breaks in HPV-associated head and neck cancer cells. Cancers *13*, E21.

Harris, R.S., Petersen-Mahrt, S.K., and Neuberger, M.S. (2002). RNA editing enzyme APOBEC1 and some of its homologs can act as DNA mutators. Mol. Cell *10*, 1247–1253.

Kati, W.M., Johnson, K.A., Jerva, L.F., and Anderson, K.S. (1992). Mechanism and fidelity of HIV reverse transcriptase. J. Biol. Chem. *267*, 25988–25997.

Kelley, M.R., Logsdon, D., and Fishel, M.L. (2014). Targeting DNA repair pathways for cancer treatment: what's new? Future Oncol. *10*, 1215–1237.

Kouno, T., Silvas, T.V., Hilbert, B.J., Shandilya, S.M.D., Bohn, M.F., Kelch, B.A., Royer, W.E., Somasundaran, M., Kurt Yilmaz, N., Matsuo, H., and Schiffer, C.A. (2017). Crystal structure of APOBEC3A bound to single-stranded DNA reveals structural basis for cytidine deamination and specificity. Nat. Commun. *8*, 15024.

Kuong, K.J., and Loeb, L.A. (2013). APOBEC3B mutagenesis in cancer. Nat. Genet. *45*, 964–965.

Lackey, L., Law, E.K., Brown, W.L., and Harris, R.S. (2013). Subcellular localization of the APOBEC3 proteins during mitosis and implications for genomic DNA deamination. Cell Cycle *12*, 762–772.

Liu, J., Lichtenberg, T., Hoadley, K.A., Poisson, L.M., Lazar, A.J., Cherniack, A.D., Kovatich, A.J., Benz, C.C., Levine, D.A., Lee, A.V., et al. (2018a). An integrated TCGA pan-cancer clinical data resource to drive high-quality survival outcome analytics. Cell *173*, 400–416.e11.

Liu, M., Mallinger, A., Tortorici, M., Newbatt, Y., Richards, M., Mirza, A., van Montfort, R.L.M., Burke, R., Blagg, J., and Kaserer, T. (2018b). Evaluation of APOBEC3B recognition motifs by NMR reveals preferred substrates. ACS Chem. Biol. *13*, 2427–2432.

Maiti, A., Hou, S., Schiffer, C.A., and Matsuo, H. (2021). Interactions of APOBEC3s with DNA and RNA. Curr. Opin. Struct. Biol. *67*, 195–204.

Maiti, A., Myint, W., Delviks-Frankenberry, K.A., Hou, S., Kanai, T., Balachandran, V., Sierra Rodriguez, C., Tripathi, R., Kurt Yilmaz, N., Pathak, V.K., et al. (2020). Crystal structure of a soluble APOBEC3G variant suggests ssDNA to bind in a channel that extends between the two domains. J. Mol. Biol. *432*, 6042–6060.

Mislak, A.C., and Anderson, K.S. (2016). Insights into the molecular mechanism of polymerization and nucleoside reverse transcriptase inhibitor incorporation by human PrimPol. Antimicrob. Agents Chemother. *60*, 561–569.

Navaratnam, N., Morrison, J.R., Bhattacharya, S., Patel, D., Funahashi, T., Giannoni, F., Teng, B.B., Davidson, N.O., and Scott, J. (1993). The p27 catalytic subunit of the apolipoprotein B mRNA editing enzyme is a cytidine deaminase. J. Biol. Chem. *268*, 20709–20712.

Ng, J.C.F., and Fraternali, F. (2020). Understanding the structural details of APOBEC3-DNA interactions using graph-based representations. Curr. Res. Struct. Biol. *2*, 130–143.

Nik-Zainal, S., Alexandrov, L.B., Wedge, D.C., Van Loo, P., Greenman, C.D., Raine, K., Jones, D., Hinton, J., Marshall, J., Stebbings, L.A., et al. (2012). Mutational processes molding the genomes of 21 breast cancers. Cell *149*, 979–993.

Rasmussen, K.D., and Helin, K. (2016). Role of TET enzymes in DNA methylation, development, and cancer. Genes Dev. *30*, 733–750.

Roberts, S.A., Lawrence, M.S., Klimczak, L.J., Grimm, S.A., Fargo, D., Stojanov, P., Kiezun, A., Kryukov, G.V., Carter, S.L., Saksena, G., et al. (2013). An APOBEC cytidine deaminase mutagenesis pattern is widespread in human cancers. Nat. Genet. *45*, 970–976.

Salter, J.D., and Smith, H.C. (2018). Modeling the embrace of a mutator: APOBEC selection of nucleic acid ligands. Trends Biochem. Sci. *43*, 606–622.

Schutsky, E.K., DeNizio, J.E., Hu, P., Liu, M.Y., Nabel, C.S., Fabyanic, E.B., Hwang, Y., Bushman, F.D., Wu, H., and Kohli, R.M. (2018). Nondestructive, base-resolution sequencing of 5-hydroxymethylcytosine using a DNA deaminase. Nat. Biotechnol. *36*, 1083–1090.

Shandilya, S.M.D., Bohn, M.F., and Schiffer, C.A. (2014). A computational analysis of the structural determinants of APOBEC3's catalytic activity and vulnerability to HIV-1 Vif. Virology *471–473*, 105–116.

Shi, K., Carpenter, M.A., Banerjee, S., Shaban, N.M., Kurahashi, K., Salamango, D.J., McCann, J.L., Starrett, G.J., Duffy, J.V., Demir, Ö., et al. (2017a). Structural basis for targeted DNA cytosine deamination and mutagenesis by APOBEC3A and APOBEC3B. Nat. Struct. Mol. Biol. *24*, 131–139.

Shi, K., Carpenter, M.A., Kurahashi, K., Harris, R.S., and Aihara, H. (2015). Crystal structure of the DNA deaminase APOBEC3B catalytic domain. J. Biol. Chem. *290*, 28120–28130.

Shi, K., Demir, Ö., Carpenter, M.A., Wagner, J., Kurahashi, K., Harris, R.S., Amaro, R.E., and Aihara, H. (2017b). Conformational switch regulates the DNA cytosine deaminase activity of human APOBEC3B. Sci. Rep. *7*, 17415.

Silvas, T.V., Hou, S., Myint, W., Nalivaika, E., Somasundaran, M., Kelch, B.A., Matsuo, H., Kurt Yilmaz, N., and Schiffer, C.A. (2018). Substrate sequence selectivity of APOBEC3A implicates intra-DNA interactions. Sci. Rep. *8*, 7511.

Smith, H.C., Bennett, R.P., Kizilyer, A., McDougall, W.M., and Prohaska, K.M. (2012). Functions and regulation of the APOBEC family of proteins. Semin. Cell Dev. Biol. *23*, 258–268.

Stenglein, M.D., Burns, M.B., Li, M., Lengyel, J., and Harris, R.S. (2010). APOBEC3 proteins mediate the clearance of foreign DNA from human cells. Nat. Struct. Mol. Biol. *17*, 222–229.

Stenglein, M.D., Matsuo, H., and Harris, R.S. (2008). Two regions within the amino-terminal half of APOBEC3G cooperate to determine cytoplasmic localization. J. Virol. *82*, 9591–9599.

Subramaniam, D., Thombre, R., Dhar, A., and Anant, S. (2014). DNA methyltransferases: a novel target for prevention and therapy. Front. Oncol. *4*, 80.

Swanton, C., McGranahan, N., Starrett, G.J., and Harris, R.S. (2015). APOBEC enzymes: mutagenic fuel for cancer evolution and heterogeneity. Cancer Discov. *5*, 704–712.

Taylor, B.J., Nik-Zainal, S., Wu, Y.L., Stebbings, L.A., Raine, K., Campbell, P.J., Rada, C., Stratton, M.R., and Neuberger, M.S. (2013). DNA deaminases induce break-associated mutation showers with implication of APOBEC3B and 3A in breast cancer kataegis. Elife *2*, e00534.

Verhalen, B., Starrett, G.J., Harris, R.S., and Jiang, M. (2016). Functional upregulation of the DNA cytosine deaminase APOBEC3B by polyomaviruses. J. Virol. *90*, 6379–6386.

Wagner, J.R., Demir, Ö., Carpenter, M.A., Aihara, H., Harki, D.A., Harris, R.S., and Amaro, R.E. (2019). Determinants of oligonucleotide selectivity of APOBEC3B. J. Chem. Inf. Model. *59*, 2264–2273.

Wyatt, M.D. (2013). Advances in understanding the coupling of DNA base modifying enzymes to processes involving base excision repair. Adv. Cancer Res. *119*, 63–106.

Xiao, X., Yang, H., Arutiunian, V., Fang, Y., Besse, G., Morimoto, C., Zirkle, B., and Chen, X.S. (2017). Structural determinants of APOBEC3B non-catalytic domain for molecular assembly and catalytic regulation. Nucleic Acids Res. *45*, 7540.

Yang, H., Ito, F., Wolfe, A.D., Li, S., Mohammadzadeh, N., Love, R.P., Yan, M., Zirkle, B., Gaba, A., Chelico, L., and Chen, X.S. (2020). Understanding the structural basis of HIV-1 restriction by the full length double-domain APOBEC3G. Nat. Commun. *11*, 632.

Zapatka, M., Borozan, I., Brewer, D.S., Iskar, M., Grundhoff, A., Alawi, M., Desai, N., Sültmann, H., Moch, H., et al.; PCAWG Pathogens (2020). The landscape of viral associations in human cancers. Nat. Genet. *52*, 320–330.

Ziegler, S.J., Hu, Y., Devarkar, S.C., and Xiong, Y. (2019). APOBEC3A loop 1 is a determinant for single-stranded DNA binding and deamination. Biochemistry *58*, 3838–3847.

Ziegler, S.J., Liu, C., Landau, M., Buzovetsky, O., Desimmie, B.A., Zhao, Q., Sasaki, T., Burdick, R.C., Pathak, V.K., Anderson, K.S., and Xiong, Y. (2018). Insights into DNA substrate selection by APOBEC3G from structural, biochemical, and functional studies. PLoS One *13*, e0195048.

STAR★METHODS

KEY RESOURCES TABLE

REAGENT or RESOURCE	SOURCE	IDENTIFIER
Antibodies		
Rabbit polyclonal APOBEC3A antibody	Thermo Fisher Scientific	Cat#PA5-78800; RRID:AB_2745916
Rabbit monoclonal APOBEC3B antibody (clone E9A2G)	Cell Signaling Technology	Cat# 41494, RRID:AB_2799203
Mouse monoclonal GAPDH antibody (clone D4C6R)	Cell Signaling Technology	Cat# 97166, RRID:AB_2756824
Goat anti-Mouse IgG (H+L) Secondary Antibody, DyLight™ 550	Thermo Fisher Scientific	Cat# 84540, RRID:AB_10942171
Goat anti-Rabbit IgG (H+L) Secondary Antibody, DyLight™ 650	Thermo Fisher Scientific	Cat# 84546, RRID:AB_10943244
Bacterial and virus strains		
OverExpress C43(DE3) SOLOs chemically competent *E. coli* cells	BioSearch Technologies	Cat#60446-1
One Shot BL21 Star (DE3) chemically competent *E. coli* cells	Invitrogen	Cat#C601003
BL21-Gold (DE3) Competent Cells	Agilent	Cat#230132
Biological samples		
HNSCC patient tumor specimen	Yale University	window clinical trial (NCT02178072)
Chemicals, peptides, and recombinant proteins		
Uracil-DNA Glycosylase	New England BioLabs	Cat#M0280S
Sequel™NE, Part A, Concentrate	AmericanBio	Cat#AB-13121
Dulbecco's modified Eagle's medium	Genesee Scientific	Cat#25-501N
DMEM/F12	ThermoFisher	Cat#11320033
Fetal bovine serum, 100% US origin, heat inactivated	Genesee Scientific	Cat#25-514H
Non-essential amino acid solution, 100×, sterile	Genesee Scientific	Cat#25–536
L-glutamine solution, 100×, 200mM solution, sterile	Genesee Scientific	Cat#25–509
Penicillin-Streptomycin, 100× solution, sterile	Genesee Scientific	Cat#25–512
Hydrocortisone	Sigma	Cat#H6909
Puromycin	InvivoGen	Cat#ant-pr-1
Lipofectamine RNAiMAX Transfection Reagent	Invitrogen	Cat#13778150
cOmplete protease inhibitor cocktail EDTA-free tablets	Sigma Aldrich	Cat#11836170001
[γ-32P] adenosine triphosphate	Perkin Elmer	Cat #NEG502Z250UC
Uracil-DNA Glycosylase	New England BioLabs	Cat#M0280S
T4 Polynucleotide Kinase	New England BioLabs	Cat#M0201S
5-aza-cytidine	Sigma	Cat#A2385
Critical commercial assays		
MycoAlert Mycoplasma Detection Kit	Lonza	Cat#LT07-218
Subcellular Protein Fractionation Kit for Cultured Cells	Thermo Scientific	Cat#78840
Total RNA Extraction & Purification kit	New England BioLabs	Cat#T2010S
QIAGEN RNeasy Mini Kit	QIAGEN	Cat#74104
iScript™ cDNA Synthesis Kit	Bio-Rad	Cat#1708890
PowerUp™ SYBR™ Green Master Mix	Applied Biosystems	Cat#A25741

(*Continued on next page*)

Continued

REAGENT or RESOURCE	SOURCE	IDENTIFIER
Experimental models: Cell lines		
Human: UMSCC47 cell line	University of Michigan	N/A
Human: YSCC94 cell line	Yale University	N/A
Human: UNC521 cell line	UNC at Chapel Hill	N/A
Human: WSCC283 cell line	UNC at Chapel Hill	N/A
Human: SCC35 cell line	University of Michigan	N/A
Human: UDSCC2 cell line	University of Düsseldorf	N/A
Oligonucleotides		
Oligonucleotides for biochemical assays, see Table S1.	Integrated DNA Technologies	N/A
5′ 6-FAM labeled DNA 43mer for cell lysate APOBEC activities, see Table S2.	Integrated DNA Technologies	N/A
APOBEC3B DsiRNAs, see Table S3.	Integrated DNA Technologies	N/A
RT-qPCR primers, see Table S4.	Integrated DNA Technologies	N/A
Recombinant DNA		
APOBEC3B CRISPR/Cas9 KO Plasmid (h)	Santa Cruz Biotechnology	Cat# sc-401700
APOBEC3B HDR Plasmid (h)	Santa Cruz Biotechnology	Cat# sc-401700-HDR
Plasmid: MBP-A3B	Donated by Xiaojiang Chen; Xiao et al. (2017)	N/A
Plasmid: MBP-A3A	Schutsky et al. (2018)	Addgene plasmid #109231
Plasmid: A3B-CTD	Donated by Angela M. Gronenborn; Byeon et al. (2016)	N/A
Software and algorithms		
Quantity One 1-D analysis software	BioRad	https://www.bio-rad.com/en-us/product/quantity-one-1-d-analysis-software?ID=1de9eb3a-1eb5-4edb-82d2-68b91bf360fb
Prism	GraphPad	https://www.graphpad.com/scientific-software/prism/
Original code	This paper	https://github.com/TravisParkeSchrank/OPSCC_APOBEC
FirebrowseR: an R client to the Broad Institute's Firehose Pipeline	https://rdrr.io/github/mariodeng/FirebrowseR/f/vignettes/FirebrowseR.Rmd	https://doi.org/10.1038/s41587-020-0546-8
R survival package v3.2–7	https://cran.r-project.org/web/packages/survival/index.html	
R survminer package v0.4.8	https://cran.r-project.org/web/packages/survminer/index.html	
R version v4.1.2 "Bird Hippie"	R Foundation for Statistical Computing	
R maftools package v2.12.0	https://bioconductor.org/packages/release/bioc/html/maftools.html	

RESOURCE AVAILABILITY

Lead contact

Further information and requests for resources and reagents should be directed to and will be fulfilled by the lead contact, Karen S. Anderson (karen.anderson@yale.edu).

Materials availability

This study did not generate new unique reagents.

Data and code availability

This article analyzes existing, publicly available data. These accession numbers for the datasets are listed in the key resources table. All other data reported in this article will be shared by the lead contacton request. All original code has been deposited at GitHub and is publicly available as of the date of publication. DOIs are listed in the key resources table. Any additional information required to reanalyze the data reported in this article is available from the lead contact on request.

EXPERIMENTAL MODEL AND SUBJECT DETAILS

Human samples

Tumor specimens derived from research-consented patients with HNSCC enrolled in a window 5-azaC trial (NCT02178072) were used. The clinical trial of pre-operative 5-azacytidine administration was approved by the Yale Human Investigation Committee, HIC#1404013771. All study participants provided written informed consent. Patient information, including age, developmental stage, and gender, is not available because all specimen were provided in a de-identified manner.

Cell culture

UMSCC47, UDSCC2, YSCC94, UNC521, and WSCC283 cells were cultured in Dulbecco's modified Eagle's medium (DMEM) (Genesee #25–501N) supplemented with 10% FBS (Genesee # 25–514H), 1% penicillin-streptomycin (Genesee #25–512), 1% non-essential amino acids (Genesee #25–536), and 1% L-glutamine (Genesee #25–509). SCC35 cells were cultured in DMEM/F12 medium (Genesee Scientific) supplemented with 10% FBS, 1% penicillin-streptomycin, 1% L-glutamine, and 0.4 μg/mL hydrocortisone (Sigma). All cells were cultured in a humidified atmosphere of 5% CO_2 at 37°C. All cells were tested mycoplasma negative by MycoAlert Mycoplasma Detection Kit (Lonza).

UMSCC47 and SCC35 cells were established from male patients. Sex for all other cells was unspecified because they were derived from de-identified patients. Cell authentication was performed on UMSCC47 and SCC35 cells using GenePrint 10 System from Promega (#B9510), and cell authentication for other cell lines has not been available yet.

Microbe strains

C43(DE3) *E. coli* cells (BioSearch Technologies) were grown overnight on LB plates containing ampicillin and 1% glucose. Colonies were grown in LB broth containing ampicillin and 1% glucose to an OD of 0.4. The cells were induced with 1 mM IPTG and grown overnight at 16°C for protein expression.

BL21 Star DE3 *E. coli* cells (Invitrogen) were grown overnight on LB plates containing kanamycin. Colonies were grown in LB broth containing kanamycin to an OD of 0.6. The cells were induced with 1 mM IPTG and grown overnight at 16°C for protein expression.

BL21-Gold (DE3) Competent Cells (Agilent) were grown in LB broth containing 100 mg/L ampicillin to an OD600 of 0.8 and was then transferred to 16°C followed by protein expression induction overnight with 0.4 mM IPTG.

METHOD DETAILS

A3B expression and purification

pMAL c5x plasmid containing the sequence of N-terminal MBP-tagged A3B was generously donated by the Xiaojiang Chen laboratory at the University of Southern California. This plasmid was transformed into C43(DE3) *E. coli* cells (BioSearch Technologies) that were grown overnight on LB plates containing ampicillin and 1% glucose. Colonies were grown in LB broth containing ampicillin and 1% glucose to an OD of 0.4. The cells were induced with 1 mM IPTG and grown overnight at 16°C. Cells were harvested and resuspended in Lysis Buffer (50 mM Tris pH 8.0, 500 mM NaCl, 2 mM DTT, cOmplete protease inhibitor tablet (Sigma Aldrich), 0.2% Triton X-100). Cells were lysed using a cell disruptor. The lysate was centrifuged at 16,000 RPM and 4°C for 1 h. The supernatant was then run over an Amylose column at 4°C for 3 h. The column was washed with 6 column volumes of Wash Buffer (50 mM Tris pH 8.0, 1 M NaCl, 2 mM DTT). The protein was eluted with 6 column volumes of Elution Buffer (50 mM Tris pH 8.0, 500 mM NaCl, 2 mM DTT, 30 mM maltose). Fractions containing A3B protein were concentrated and buffer exchanged into Storage Buffer (50 mM Tris, 200 mM NaCl, 10% glycerol, 2 mM DTT). The MBP tag is located at the N-terminus

whereas the cytidine deaminase catalytic activity is located at the C-terminus. The presence of the MBP significantly improves the solubility of A3B to allow kinetic analysis. Similar deamination kinetics are observed upon tag removal however there is a 2-fold decrease in catalytic activity therefore the studies were performed with the MPB-tagged A3B (see Figure S1).

A3B-CTD expression and purification

The pET21b plasmid expressing the A3B-CTD (residues 187–382) construct was a courtesy of the Gronenborn Lab from University of Pittsburg School of Medicine. The A3B-CTD expression and purification procedure was adapted from published protocols with modifications (Byeon et al., 2016). Briefly, the plasmid was transformed into BL21-Gold (DE3) Competent Cells (Agilent). The bacteria cells were grown in LB broth containing 100 mg/L ampicillin to an OD600 of 0.8 and was then transferred to 16°C followed by protein expression induction overnight with 0.4 mM IPTG. The cells were pelleted and resuspended in Buffer A (50 mM Tris-HCl pH 7.5, 100 mM NaCl, 30 mM imidazole) and lysed using a cell disruptor. The supernatant was clarified by centrifugation at 16,000 RPM for 45 min and was applied to Ni-sepharose 6 fast flow resin (Cytiva). The resin was washed with 20 column volumes of Buffer A. The protein was eluted with Buffer B (50 mM Tris-HCl pH 7.5, 100 mM NaCl, 300 mM imidazole). The eluate was concentrated and further purified with a Superdex 75 10/300 GL size exclusion column (Cytiva) pre-equilibrated with buffer containing 20 mM Tris-HCl pH 7.5, 100 mM NaCl. Peak fractions containing A3B-CTD were pooled and concentrated to ~4 mg/mL for the assay.

A3A expression and purification

A3A protein for cytidine deaminase activity assays was expressed using pET41 plasmid containing the sequence of N-terminal MBP-tagged and C-terminal His-tagged A3A from Rahul Kohli (Addgene plasmid #109231)(Schutsky et al., 2018). The plasmid was transformed into BL21 Star DE3 *E. coli* cells (Invitrogen) that were grown overnight on LB plates containing kanamycin. Colonies were grown in LB broth containing kanamycin to an OD of 0.6. The cells were induced with 1 mM IPTG and grown overnight at 16°C. Cells were then harvested and purified using the same method as described above for A3B.

Design and preparation of oligonucleotides

Oligonucleotides with varying lengths were used to test the impact of the length of the oligonucleotide on the cytidine deaminase activity of the protein. Each oligonucleotide contained a centered TCA motif. Oligonucleotides were ordered from Integrated DNA Technologies, PAGE purified for maximal purity.

Length of Oligonucleotide	
10mer	5′ - TGGATCAATT– 3′
25mer	5′ - ATTATTATGGATCAATTATTTATTA– 3′
43mer	5′ - ATTATTATTATTATTATGGATCAATTATTTATTATTATTATTA – 3′
70me	r5′ – TATTATTATTATTATTATTATTATTATTATGGATCAATTATTTATTATTATTATTATTATTATTATTATT – 3′

Oligonucleotides with varying sequences were used to test the impact of the identity of varying nucleotides at the −2, −1, +1, and +2 positions on the cytidine deaminase activity of the protein. Oligonucleotides were ordered from Integrated DNA Technologies, PAGE purified for maximal purity. 5′-FAM labeled 43mer DNA oligonucleotides with variation at −2 position were used for determining deaminase activity in whole cell lysates and purified by HPLC.

Base Sequence	
XXCXX	5′ – ATT ATT ATT ATT CAA ATG GAT TTA TTT ATT TAT TTA TTT ATT T – 3′

Variation at −2 Position	
ATCAA	5′ – ATT ATT ATT AAT CAA ATG GAT TTA TTT ATT TAT TTA TTT ATT T – 3′
CTCAA	5′ – ATT ATT ATT ACT CAA ATG GAT TTA TTT ATT TAT TTA TTT ATT T – 3′
GTCAA	5′ – ATT ATT ATT AGT CAA ATG GAT TTA TTT ATT TAT TTA TTT ATT T – 3′
TTCAA	5′ – ATT ATT ATT ATT CAA ATG GAT TTA TTT ATT TAT TTA TTT ATT T – 3′

Variation at −1 Position	
AACAA	5′ – ATT ATT ATT AAC AAA ATG GAT TTA TTT ATT TAT TTA TTT ATT T – 3′
ACCAA	5′ – ATT ATT ATT ACC AAA ATG GAT TTA TTT ATT TAT TTA TTT ATT T – 3′
AGCAA	5′ – ATT ATT ATT AGC AAA ATG GAT TTA TTT ATT TAT TTA TTT ATT T – 3′
ATCAA	5′ – ATT ATT ATT ATC AAA ATG GAT TTA TTT ATT TAT TTA TTT ATT T – 3′

Variation at +1 Position	
TTCAA	5′ – ATT ATT ATT ATT CAA ATG GAT TTA TTT ATT TAT TTA TTT ATT T – 3′
TTCCA	5′ – ATT ATT ATT ATT CCA ATG GAT TTA TTT ATT TAT TTA TTT ATT T – 3′
TTCGA	5′ – ATT ATT ATT ATT CGA ATG GAT TTA TTT ATT TAT TTA TTT ATT T – 3′
TTCTA	5′ – ATT ATT ATT ATT CTA ATG GAT TTA TTT ATT TAT TTA TTT ATT T – 3′

Variation at +2 Position	
TTCAA	5′ – ATT ATT ATT ATT CAA ATG GAT TTA TTT ATT TAT TTA TTT ATT T – 3′
TTCAC	5′ – ATT ATT ATT ATT CAC ATG GAT TTA TTT ATT TAT TTA TTT ATT T – 3′
TTCAG	5′ – ATT ATT ATT ATT CAG ATG GAT TTA TTT ATT TAT TTA TTT ATT T – 3′
TTCAT	5′ – ATT ATT ATT ATT CAT ATG GAT TTA TTT ATT TAT TTA TTT ATT T – 3′

Oligonucleotides for testing length and sequence identity preference were end-labeled with [γ-^{32}P] adenosine triphosphate (PerkinElmer, Waltham, MA) with T4 polynucleotide kinase (New England Biolabs, Ipswich, MA) as previously described (Kati et al., 1992; Mislak and Anderson, 2016). The resulting oligonucleotides were purified on a Bio-Spin 6 column (Bio-Rad Laboratories, Hercules, CA) and equilibrated in 30 mM Tris pH 8.0 and 100 mM NaCl buffer.

Biochemical cytidine deaminase activity assay

Stocks of the radiolabeled oligonucleotides were prepared at 40,000 counts per minute (cpm) per μL. A series of single enzyme turnover experiments were used to assess catalytic activity. Reaction mixtures were prepared with 40 nM of a radiolabeled oligonucleotide and 1 μM of enzyme in APOBEC Assay Buffer (50 mM Tris pH 8.0, 100 mM NaCl, 1 mM DTT, 0.1% Triton X-100). The reactions were incubated, and aliquots were taken at 7 min, 15 min, 30 min, 1 h, 2 h, and 3 h. At each of these time points, the samples were quenched by adding 0.5 M EDTA to a final concentration of 50 mM. Samples were then incubated for 5 min at 95°C. Pre-quenched control reactions were prepared by pre-incubating 1 μm enzyme in APOBEC Assay Buffer and a final concentration of 50 mM EDTA for 5 min at 95°C, and then adding radiolabeled oligonucleotide to a final concentration of 40 nM. After all reactions were completed, samples were cooled and 5 units of Uracil-DNA Glycosylase (New England Biolabs, Ipswich, MA) were added. Samples were incubated at 37°C for one hour. NaOH was added to a final concentration of 100mM and samples were incubated at 37°C for 30 min. An equal volume denaturing PAGE Dye (20 mM EDTA, 0.3% bromophenol blue, 0.3% xylene cyanol, 95% formamide) was added and samples were incubated for 5 min at 95°C. Samples were then run on a 20% denaturing 8 M urea PAGE gel (Sequel™NE, americanBIO) at 200V for 1.5 h. Gels were exposed to an imaging phosphor screen (Bio-Rad Laboratories, Hercules, CA) for six hours. All experiments were carried out in three sets of functional duplicates, for six total data points for each time point.

Imaging and kinetic data analysis

Screens were imaged using a Bio-Rad phosphorimager (Bio-Rad Laboratories, Hercules, CA). Densitometry was then performed for each band using the Quantity One 1-D analysis software (Bio-Rad Laboratories, Hercules, CA), with product formation at 0 h subtracted from all values to eliminate background density. The intensities of substrate and product oligonucleotide at individual timepoints were expressed as a ratio value to give the percent product formed. These percentages were fit to a single exponential curve, $Percent\ product = A(1\text{-}e^{kobs*time})$, in which A is the maximum percent product, to calculate observed single turnover rate constants (k_{obs}).

A3B knockout/knockdown and cell lysate preparation

For UMSCC47 cells, co-transfection of A3B CRISPR/Cas KO (Santa Cruz # sc-401700) and A3B HDR (Santa Cruz # sc-401700-HDR) plasmids were used per manufacture's protocol to establish A3B knockout UMSCC47 cells. A3B KO clones were selected with 2 μg/mL puromycin (InvivoGen ant-pr-1).

For UDSCC2, YSCC94, SCC35, UNC521, and WSCC283 cells, A3B is knocked down by siRNAs. DsiRNAs are purchased from Integrated DNA Technologies (IDT) and the sequence of DsiRNAs is listed below. Reverse transfection was performed to knock down A3B using Lipofectamine RNAiMAX Transfection Reagent (Thermo Fisher) according to manufacture instructions.

DsiRNA.APOBEC3B.1: 5′-ACCUACGAUGAGUUUGAGUACUGCT-3′

5′-AGCAGUGCUCAAACUCAUCGUAGGUCA-3′

DsiRNA.APOBEC3B.2: 5′-UUUCUAUGCAACGAGGCUAAGAATC-3′

5′-GAUUCUUAGCCUCGUUGCAUAGAAAGC-3′

DsiRNA.APOBEC3B.3: 5′-CUAUGCAACGAGGCUAAGAAUCUTC-3′

5′-GAAGAUUCUUAGCCUCGUUGCAUAGAA-3′

A3B knockdown was confirmed by western blot analysis.

To make whole cell lysate, cells were incubated with nuclear buffer from Thermo Fisher Subcellular Protein Fractionation Kit for Cultured Cells (#78840) for 15 min, sonicated and incubated for another 15min; lysates were then centrifuged at 4°C and 16,000 RPM for 15 min to remove insoluble material.

Immunoblotting

Identical amounts of cell lysate (5–20ug) was mixed with 2× loading Laemmli buffer (Biorad) supplemented with DTT (Sigma) and incubated for 10 min at 95°C. Proteins were separated in 4–20% Tris-glycine polyacrylamide gels (Mini-PROTEAN; Bio-Rad) and electrophoretically transferred onto polyvinylidene fluoride membranes. Membranes were blocked with 3% BSA in PBS and incubated with primary antibodies against APOBEC3A (PA5-78800 from Thermo Fisher) and/or APOBEC3B (E9A2G, Cell Signaling), as well as GAPDH (D4C6R, Cell Signaling). Secondary antibodies were goat anti-mouse IgG (H + L) DyLight 550 or goat anti-rabbit IgG (H + L) DyLight 650 (Thermo Fisher). After sequential washes in TBST buffer, fluorescent bands were visualized using a ChemiDoc Bio-Rad imager.

Clonogenic survival assay

Cells were plated into six well plates at a density of 1000 cells/well. The next day, the cells were treated with increasing concentrations of 5-AzaC. Colonies were fixed and stained with methylene blue in methanol (4 g/L) 10 days later. Colonies consisting of more than 30 cells were counted.

RT-qPCR

Total RNA was extracted from cells or tumor tissue by the Total RNA Extraction & Purification kit (NEB) or by Qiagen RNA extraction kit and cDNA was synthesized using iScript cDNA Synthesis Kit (Bio-Rad) according to the manufacturer's instructions. Quantitative real-time reverse transcription (RT-qPCR) was done using PowerUp™ SYBR™ Green Master Mix (Applied Biosystems) and primer pairs indicated below on the QuantStudio Real-Time PCR system (Thermo Fisher). Each RT-qPCR reaction was done in duplicate at least, and the ΔΔCt method was used to analyze the data; relative expression (% from GPDH) is presented.

APOBEC3A Forward: 5′ – ATGGCATTGGAAGGCATAAG – 3′

APOBEC3A Reverse: 5′ – CAAAGAAGGAACCAGGTCCA – 3′

APOBEC3B Forward: 5′ – TTTGCATACTGCTGGGAAAA – 3′

APOBEC3B Reverse: 5′ – GCTCCACCTCATAGCACAAG – 3′

Cytidine deaminase activity assay with cell lysates

Reaction mixtures were prepared with 1μM of 5′-FAM labeled DNA 43mer oligonucleotides and 10 μg of cell lysate in APOBEC Assay Buffer. The reactions were incubated at 37°C overnight. Samples were then subjected to UDG-dependent cytidine deaminase activity assays as described above. Samples were then separated by 20% denaturing 8 M urea-PAGE gel and monitored using a ChemiDoc Bio-Rad imager (Bio-Rad Laboratories, Hercules, CA).

RNA expression data

Publicly available Kalisto transcript per million (TPM) data for the TCGA pan-cancer data set was downloaded through the UCSC Xena portal. Transcript identifiers ENST00000333467.3, ENST00000407298.7, and ENST00000402182.7 APOBEC3B (A3B) were identified according to the Ensembl database. Transcript identifiers ENST00000402255.5, ENST00000249116.6 and ENST00000618553.1 for APOBEC3A (A3A) were similarly identified. Gene expression of A3A and A3B were considered to the sum of all related transcripts in TPM. Data were $\log_2(1 + \text{TPM})$ transformed before downstream analyses.

Mutational data

TCGA pan-cancer mutational data was downloaded via the Xena UCSC website. Specifically, TCGA Unified Ensemble "MC3" mutation calls were used (Ellrott et al., 2018). Data was accessed in MAF format which included SNP nucleotide contexts. Various Xt[c]a > t/g polymorphism were summarized by frequency per tumor using in house scripts.

Cohort selection and inclusion criteria

RNA assigned HPV status from the Broad Firehose portal TCGA metadata were used to assign HPV status; HPV positive tumors were included if smoking exposure was less than or equal to 30 pack-years and anatomic subsites from the oropharynx, tonsil, and base of tongue; as well as nearby hypopharynx and oral tongue. Tumors from sites more distal to the anatomic oropharynx (e.g., larynx, alveolar ridge, maxilla) were excluded from the HPV+ HNSCC cohort. A total of 39 patients met these criteria. All HPV negative cases were included as HPV− HNSCC for a total of 180 patients (Deng et al., 2017; Goldman et al., 2020).

Correlation analysis

Correlation analysis was performed for various Xt[c]a > t/g polymorphisms and gene level expression of A3A and A3B were both data types were available. Spearman rank correlation coefficients were calculated with the R corr function. The Spearman CI R package was used to calculate 95% confidence intervals for all correlation coefficients, using default settings. The R corrplot and wesanderson packages were used to visualize the correlation data.

Survival analysis

Clinical data, specifically progression-free interval (PFI), were extracted from Liu et al. across the full cohort (n = 61) (Liu et al., 2018a). We note that the values for PFI from Liu et al. were very similar or identical (but included four more cases) when compared to recurrence-free survival (RFS) data available from Broad Firehose Portal. (Deng et al., 2017) Therefore, we analyzed PFI to increase the number of available cases. Survival statistics were generated with the R survival package (v3.2–7) and visualized with the R survminer package (0.4.8). p-values represent log-rank test.

APOBEC signature analysis

Analysis was performed for various Xt[c]a > t/g polymorphisms and gene level expression of A3A and A3B where both data types were available. Similar to Chan et al., and in agreement with the presented kinetic data, we considered tumors with 50% or more A/Gt[c]a > t/g as displaying an A3B signature; and tumors with >50% T/Ct[c]a > t/g as displaying an A3A mutational signature. The R ggplot2 and wesanderson packages were used to visualize the correlation data (Chan et al., 2015).

QUANTIFICATION AND STATISTICAL ANALYSIS

Statistical analysis was performed by GraphPad Prism 8.1.0. All data are presented as mean ± SD from at least three replicates, unless otherwise indicated. Statistical significance was calculated by student's t-test for the comparison between two groups or one-way ANOVA for the comparison between more than two groups. In all analyses, $p < 0.05$ is considered significant. And p-value is annotated as follows: *, $p < 0.05$; **, $p \leq 0.01$; ***, $p \leq 0.001$; ****, $p \leq 0.001$.

ADDITIONAL RESOURCES

The clinical registry number is NCT02178072 for the clinical trial of pre-operative 5-azacytidine administration in HNSCC patients. Any additional information of this clinical trial can be found through the link: https://clinicaltrials.gov/ct2/show/NCT02178072.