HHS Public Access

Author manuscript

J Mol Biol. Author manuscript; available in PMC 2024 September 01.

Published in final edited form as:

J Mol Biol. 2023 September 01; 435(17): 168198. doi:10.1016/j.jmb.2023.168198.

Unraveling the Enzyme-Substrate Properties for APOBEC3A-mediated RNA Editing

Kyumin Kim[1], **Alan B. Shi**[1], **Kori Kelley**[1], **Xiaojiang S. Chen**[1,2,3,4,*]

[1]Molecular and Computational Biology program, Departments of Biological Sciences

[2]Genetic, Molecular and Cellular Biology Program, Keck School of Medicine

[3]Center of Excellence in NanoBiophysics

[4]Norris Comprehensive Cancer Center; University of Southern California, Los Angeles, CA 90089, USA

Abstract

The APOBEC3 family of human cytidine deaminases is involved in various cellular processes, including the innate and acquired immune system, mostly through inducing C-to-U in single-stranded DNA and/or RNA mutations. Although recent studies have examined RNA editing by APOBEC3A (A3A), its intracellular target specificity are not fully characterized. To address this gap, we performed in-depth analysis of cellular RNA editing using our recently developed sensitive cell-based fluorescence assay. Our findings demonstrate that A3A and an A3A-loop1-containing APOBEC3B (A3B) chimera are capable of RNA editing. We observed that A3A prefers to edit specific RNA substrates which are not efficiently deaminated by other APOBEC members. The editing efficiency of A3A is influenced by the RNA sequence contexts and distinct stem-loop secondary structures. Based on the identified RNA specificity features, we predicted potential A3A-editing targets in the encoding region of cellular mRNAs and discovered novel RNA transcripts that are extensively edited by A3A. Furthermore, we found a trend of increased synonymous mutations at the sites for more efficient A3A-editing, indicating evolutionary adaptation to the higher editing rate by A3A. Our results shed light on the intracellular RNA editing properties of A3A and provide insights into new RNA targets and potential impact of A3A-mediated RNA editing.

Graphical Abstract

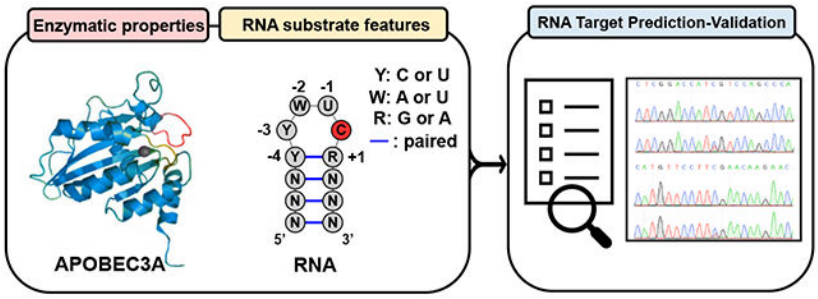
*To whom correspondence should be addressed. Tel: +1 213 740-5487; FAX: +1 213 740 4340; xiaojiac@usc.edu.

AUTHOR CONTRIBUTIONS

K. Kim and X.S.C designed the conceptual framework for this study. K. Kim, A.B.S., and K. Kelley performed and analyzed the experimental data. K. Kim wrote the manuscript draft. X.S.C. oversaw the project and revised the manuscript.

COMPETING INTEREST

The authors declare no competing interests.



Keywords

APOBEC3A; RNA editing; Substrate specificity; Cellular RNA targets

INTRODUCTION

RNA editing by deamination is a crucial event that alters a single nucleotide of a transcript, resulting in the diversification and regulation of the transcriptome [1–3]. The human RNA deaminases consist of two classes of enzymes: the APOBEC protein family and ADARs. APOBEC proteins (apolipoprotein-B (*ApoB*) mRNA editing enzyme, catalytic polypeptide-like) convert cytosine to uracil (C-to-U) in single-stranded RNA [4, 5], while ADARs (adenosine deaminases acting on RNA) modify adenosine to inosine (A-to-I) in double-stranded RNA. The A-to-I RNA editing mechanism of ADARs and the related physiological functions are relatively well studied [6–8].

The C-to-U RNA editing enzyme was first discovered when researchers found a truncated version of APOB protein, resulting from catalytic deamination (C-to-U editing) of ApoB mRNA by a novel enzyme, as a way of regulating APOB's role in lipid metabolism [9–11]. This enzyme was termed "*ApoB* mRNA editing enzyme, catalytic polypeptide 1" or APOBEC1 (A1), and the rest of the APOBEC protein members were named after this founding member [4]. The human APOBEC proteins are composed of a total of 11 members, each with a different biological role, playing various physiological functions [12–14]. Initially, only A1 was considered to have single-stranded RNA editing activity [15]. But recent papers have reported RNA editing activities mediated by A3A in specific cell types [16–18]. Moreover, (A3G)-mediated RNA editing functions have been described in HEK293T cells [19], and the role of wide spread A3G-induced RNA editing in response to mitochondrial hypoxic stress in natural killer cells [20]. More recently, APOBECs are shown to edit RNA genomes of viral pathogens [21, 22]. These findings suggest that other APOBEC enzymes, apart from A1, may engage in RNA editing to fulfill biological functions.

A3A is reported to restrict foreign DNA from pathogens such as though C-to-U deamination [23–26]. The abnormal overexpression and/or malfunction of A3A may also be associated with genomic DNA mutation and cancer [27, 28]. Furthermore, analysis of mesoscale genomic features has demonstrated that A3A-induced mutations occur in specific "ssDNA hairpin" structures that can lead to the development of oncogenic mutation hotspots [29].

The advancement of next-generation sequencing technology has allowed the identification of A3A-mediated RNA editing sites, which exhibit a similar preference for stem-loop structures as observed in A3A-mediated ssDNA deamination [17, 18, 29, 30]. However, despite the general understanding of A3A's stem-loop RNA substrates, the specific RNA sequences of such stem-loop substrates preferred by A3A have not been fully characterized.

Given the strong correlation between high expression of A3A and INF-γ related immune responses [31, 32], as well as the potential involvement of A3A in the development of cancer [27, 28, 33, 34], and A3A-related evolution and infection of RNA viruses [21, 22, 35], it is important to examine the relationship between RNA substrate features and the A3A editing efficiency. With the help of various methods developed for measuring C-to-U RNA editing activity, including traditional techniques such as Sanger sequencing, Poisoned primer extension *in vitro* [36], the latest NGS high-throughput sequencing [17], *in vivo* cytofluorimetric analysis [37, 38], cell-based subcellular localization assay [39], and a digital PCR-based RNA editing quantification [40], these advancements have enhanced the accuracy and versatility of C-to-U RNA editing measurements. This study aims to examine the enzyme-substrate features of A3A-induced RNA editing using a sensitive cell-based fluorescence assay, and to predict new RNA targets for A3A editing in human cells.

RESULTS

Confirmation of A3A RNA editing with a sensitive cell-based fluorescence assay

We used a cell-based fluorescence assay [39] to investigate A3A RNA editing, which involved the expression of an A3A and mCherry-containing editor construct and a reporter construct containing the target RNA in HEK293T cells (Figure 1(A)). The reporter construct expressed eGFP RNA followed by target RNA substrates of 48 nucleotides and an RNA encoding for a strong nuclear export signal (NES, MAPKK) as a single open reading frame. This resulted in the reporter GFP fluorescence localization primarily in the cytoplasm due to the C-terminal NES (Figure 1(A)). However, when APOBEC deaminates the target from C-to-U, the reporter mRNA formed an early stop codon prior to the NES, allowing a shift of the reporter GFP signal to the nucleus [39]. An AAV intron was added in the middle of eGFP that can be useful to identify only the spliced mature mRNA sequences by using an exon-exon junction primer for cDNA preparation [21] (Figure 1(A)), which made it possible to avoid amplifying the DNA sequence but only amplify the RNA sequences for analysis. Six target RNA candidates reported by Sharma et al. [17, 18] were initially chosen, including *ASCC2*, *EVI2B*, *ICAM3*, *PPA2*, *SDHB*, and *SETX*, for the initial test of editing because they were shown to have relatively high levels of C-to-U editing in the RNA-seq data and the sequences could create a stop codon following the C-to-U RNA editing.

The results showed significant A3A editing on all but one (*SETX*) RNA targets, as determined by a ratio between averaged cytosolic and nuclear fluorescence signals (Figure 1(B–C)). Among them, *EVI2B* and *SDHB* showed much higher level of A3A editing than the other three targets (*ASCC2*, *ICAM3*, and *PPA2*). To further verify editing of the six target RNA substrates, the reporter DNA and mRNA were extracted respectively from the cell lysates for Sanger sequencing. The sequence result confirmed that *EVI2B* and *SDHB* had the highest edited level by A3A, with an RNA editing rate of 41.4% and 37.6%,

respectively (Figure 1(D) and Figure S1(A)). *ASCC2*, *ICAM3*, and *PPA2* showed modest RNA editing activity with an RNA editing rates of 16.3%, 12.8%, and 11.5%, respectively. Finally, *SETX* showed a marginal editing rate of 3.0% with statistically insignificant compared to that of the control (1.7%) (Figure 1(D) and Figure S1(A)). The A3A editing levels estimated by the fluorescent reporter assay correlated well with results obtained by the *in vitro* RNA sequencing approach, with a correlation coefficient $R^2 = 0.9000$, suggesting that this fluorescence reporter assay is appropriate for evaluating A3A-mediated RNA editing activity (Figure S1(B)).

To investigate the relative level of A3A editing on the endogenous DNA and RNA in cells, we selected the cellular *SDHB* gene, which is known to have high endogenous expression in HEK293T cells (Figure S2(A)). We prepared both genomic SDHB DNA and mRNA from HEK293T cells for Sanger sequencing and found no detectible editing on the *SDHB* genomic DNA target site, even with the use of a uracil glycosylase inhibitor (UGI) to suppress uracil-DNA glycosylase, which is responsible for repairing DNA mismatches caused by C-to-U DNA mutations by APOBEC [41] (Figure S2(B)). However, we observed high levels of editing on the *SDHB* mRNA target C site in both the absence and presence of UGI treatments (% of U at C site: 34.0% and 32.9%, respectively) (Figure S2(B)). These finding suggests that A3A efficiently edit *SDHB* mRNA rather than its DNA at this site within cells. This may be due to the fact that, unlike mRNA, genomic DNA is mostly in a double-stranded DNA state, which means that it is relatively less likely to be edited by A3A.

A3A targeted RNA substrates are not sufficiently edited by other APOBECs

The APOBEC family of proteins in humans consists of a total of 11 members that share a conserved cytidine deaminase motif and have significantly homologous sequences, as well as a similar core structure consisting of a five beta-stranded sheet surrounded by six helices (Figure S3(A)) [12, 42] Given these similarities, it was possible that other APOBEC members might also be capable of editing the specific RNA. To examine this, RNA editing capabilities of all 11 APOBECs were examined using a reporter based on *EVI2B* in our cell-based assay, as *EVI2B* showed the highest editing levels by A3A (Figure 1).

The results indicated that only A3A exhibited significant fluorescence shifts from the cytoplasm to the nucleus, with an average intensity of 0.661 as ratio of nuclear to cytosolic fluorescence (0: cytosol - 1: whole cell), whereas all other APOBEC proteins displayed no differences compared to the control group (Figure 2(A)). This finding was further confirmed by Sanger sequencing analysis of the *EVI2B* reporter transcripts, which revealed that, except for A3A (which exhibited a C-to-U editing rate at 43.4 %), none of the other APOBEC members were able to edit the *EVI2B* site (Figure 2(B)). Western blot analysis confirmed the expression of 10 out of 11 APOBEC proteins, except for APOBEC4 [43].The differential expression levels of the APOBEC proteins observed may be attributed to the genotoxicity of some of them in HEK293T cells (Figure S3(B)). The RNA substrates employed in our cell-based fluorescent assay were specifically chosen for A3A editing, and as a result, these RNAs may not serve as preferred substrates for editing by other APOBEC proteins due to differences in sequence and/or structural conformations.

A3A performs most of its RNA editing in the cytoplasm

Overexpression A3A was shown to be localized in both nucleus and cytoplasm [44, 45], but endogenous A3A was shown to be present mostly in the cytoplasm in monocytic cells [46]. In contrast, A3B is predominantly localized in the nucleus due to its unique N-terminal domain surfaces, specifically the first 30 amino acids of A3B or loop 5/α-helix 3 of A3B [47]. We hypothesized that these subcellular localizations may affect A3A-mediated RNA editing activity, and that A3B localized in the nucleus may not perform RNA editing activity even if its C-terminal A3B-CD2 domain shares about 92% identical amino acids with A3A [12, 14]. To investigate this hypothesis, we evaluated RNA editing activity with constructs that fused a strong nuclear export signal (NES) or nuclear localization signal (NLS) at the C-terminus of either A3A or A3B, respectively.

We found that A3A-NES was capable of inducing RNA editing of the EVI2B site when predominantly localized in the cytoplasm, with editing levels comparable to A3A-WT both in the reporter assay (average ratios, WT : NES = 0.622:0.619) and the cDNA sequencing results (C-to-U editing levels, WT : NES = 42.2% : 40.1%) (Figure 3(A–C)). However, when A3A-NLS was predominantly in the nucleus, its RNA editing activity was significantly reduced, with an average ratio of 0.332 (P value = 0.011 vs. Ctrl) in the reporter assay (Figure 3(B)) and C-to-U RNA editing rates of 6.6% in the sequencing analysis (Figure 3(C)). This implies that A3A works more favorably for RNA editing in the cytoplasmic environment (Figure 3(A–C)). It may explain why the programmable target cytidine-specific RNA editor by A3A constructed with a strong 4x NLS has lower editing levels and the smallest total number of off-target editing sites[48].

Contrary to our hypothesis for A3B, we were unable to detect specific RNA editing activities by A3B when expressed either in the cytoplasm or more strongly localized only in the nucleus (Figure 3(A–D)). This suggests that the approximately 8% different residues of A3A and A3B-CD2 catalytic domain may play an important role in specific RNA editing activity, or that about 3 times lower expression levels of A3B than A3A in HEK293T cells may result in an undetectable RNA editing activity (Figure 3(E)).

A3B chimeras containing an N-terminal A3A region gain RNA editing activity

We were prompted by a fascinating question about the potential importance of specific sequences/regions unique to A3A in determining A3A-mediated RNA editing. To investigate this, we created A3B chimeras in which the C-terminal A3B-CD2 domain contains different A3A fragments in order to see if any RNA editing activity can be transferred from A3A to the A3B chimeras. Because A3B-CD2 is highly homologous to A3A sequence, only the two N-terminal regions (Region-1 and Region-2) with major sequence differences between the two proteins are switched (Figure 4(A)). Three A3B chimeras were generated, Region-1, Region-2, and Region-1&2. In chimera Region-1 (R1), α1/loop-1 residues of A3B-CD2 are replaced with those of A3A (A3B residues 192-214 to A3A residues 16-30); In chimera Region-2 (R2), the β-2 residues of A3B-CD2 are changed to those of A3A (A3B residues 227-242 to A3A residues 45-58); and chimera Region-1&2 (R1&2) combines both R1 and R2 chimeras (Figure 4(A) and Figure S4(A)). We then tested the RNA editing activity

of theses A3B chimeras on the efficient RNA substrates for A3A, *EVI2B* and *SDHB*, respectively.

Interestingly, we found that R-1 showed a gain-of-function for RNA editing on both substrates, but R-2 had no significant RNA editing (Figure 4(B–D)). However, R-1&2 showed markedly improved RNA editing levels compared to the R-1 in both the reporter assay (Figure 4(C)) and cDNA (RNA) sequencing results (Figure 4(D)), suggesting R1 exerts a critical influence on its RNA deamination activity and that R2 has synergistic effects to enhance R1's RNA editing. Unlike the results for A3A-WT shown in Fig. 1, R-1 and R-1&2 chimeras deaminated *SDHB* mRNA more strongly than *EIV2B* (Figure 4(B–D) and Figure S4(B)), indicating that the double-domain A3B R1&2 chimera may have a slightly different module in targeting specific RNA substrates compared to the single-domain A3A.

It is worth noting that our previous studies have shown that residues around their α1/loop-1 of A3A and A3B-CD2 are important in regulating overall DNA deaminase activity and mC selectivity [49, 50]. In this study, we found that the α1/loop-1 residues of A3A also played a crucial role in performing the specific A3A-mediated RNA editing activity. When the R1 region was further narrowed down, the GIGRHK (A3A residues 24-29) of A3A, which is the loop 1 region, were identified as the most critical residues (Figure 4(C)).

When an NES peptide was fused to the C-terminus of these A3B chimera constructs, we found that their RNA editing levels increased by about 5-20%, possibly a result of the shift of localization from nucleus to the cytoplasmic environment (Figure S4(C)). These results indicate that the gain of RNA editing activity by these A3B chimeras is not due to subcellular localization, nor due to protein expression levels as they all showed comparable expression level in the cells (Figure 4(D)). Interestingly, when the loop1 region of A3A is replaced with that of A3B-CD2 (replacing A3A-GIGRHK with A3B-DPLVLRRRQ), this A3A chimera behaved similarly to the inactive A3A-E72A mutant in showing no detectible specific RNA editing activity (Figure S5), further confirming the critical role of loop1 region of A3A for RNA editing activity.

Unexpectedly, the subcellular localization of R-1&2 (mCherry signals) was significantly shifted from nucleus to the cytoplasm, exhibiting more pronounced changes in subcellular localization than R-1 and R-2 alone. (Figure 4(B) and 4(E)). This finding suggests that in addition to the nuclear localization signals present in A3B-CD1 [47], A3B subcellular localization can be regulated by the CD2 R1 and R2 regions.

Structural and sequence context of RNA substrates for optimal A3A editing

Several previous studies reported that the A3A targets specific RNA substrates with certain stem-loop secondary structures and sequence specificities [17, 18, 21, 30, 51]. However, due to limitations in the design of RNA/transcripts substrates, it was difficult to thoroughly understand the detailed characteristics of these targets for the varying editing efficiency by A3A. Even for ssDNA deamination by A3A, a DNA stem-loop secondary structure was shown to be the preferred substrate, which was hypothesized as the reason for hotspot mutations in cancer mutations driven by APOBEC mutational signatures [29]. To better understand the properties of RNA substrates for A3A editing, we scrutinized the effect

of different parts of an RNA stem-loop structure on A3A editing using the cell-based fluorescence reporter assay.

We compared the editing activity on the *EVI2B* RNA after adjusting the lengths of its loop (LP) and stem (ST). The original *EVI2B* RNA target has a predicted loop length of 5 nt (LP5) with the sequence CAUCA, where the target C is underlined. We found that the change to LP4 (LP4, CAUC) from LP5 showed a dramatic increase of RNA editing activity by A3A (Figure 5(A) and Figure S6(A)). Interestingly, the editing activity of the original *EIV2B* LP5 was at intermediate levels between LP4 and LP3, possibly due to C-A+ wobble base pairing in RNA stem region (CAUCA, possible C-A+ wobble base pairing underlined) [52] (Figure 5(A)).

To investigate the effect of stem length on A3A-mediated RNA editing, we modified the *EVI2B* target RNA stem length from the original 4 bp (ST4) to a range between 3 and 7 bp (ST3-ST7). Our results showed that when the stem was shortened from ST4 to ST3, RNA editing activity was sharply decreased to only about 30% of the ST4 level, and RNA editing levels increased as the stem lengthened to ST7. These findings suggest that the longer stems may form a more stable hairpin structure, and a more stable hairpin may present itself to A3A more chance for editing (Figure 5(A)). Therefore, we conclude that an RNA hairpin structure with a longer stem up to ST7, a loop size of 4 nt (SP4), and the target C site at the 3’ end of the LP4 show higher editing efficiency by A3A. These results are consistent with the previous report [30].

We further examine the sequence context around the target C site on the loop for A3A editing. The loop sequence of the original *EVI2B* RNA target was modified with different nucleotides. Specifically, if the target C position is set to 0 in the LP5 sequence, we investigated the effect of the sequences at positions −3, −2, −1, and +1 sites (Figure 5(B–C)). Our results showed that the presence of U at the −1 position is the most crucial sequence context for A3A targeting (Figure 5(B–C) and Figure S6(B)), which is not surprising as A3A is known to target TC motif (target C underlined) in ssDNA [25, 53]. A recent study reported that A3A prefers DNA substrate YTCA motif with the −2 position Y being pyrimidine, and thus the −2 position favors C/T [27]. In our RNA editing system, however, the −2 position shows a strong preference for A at −2 position, with the order of preference of A > U > G > C (Figure 5(B–C) and Figure S6(B)). Although it showed that −3 position prefers a pyrimidine (Y) and the +1 position prefers a purine (R) (Figure 5(B)), certain sequences at the −3 and +1 positions at both ends of the LP5 allow base-pair formation to increase the RNA stem length by one bp (i.e. ST4 to ST5) and reduce the loop size by 2 nt (i.e. LP5 to LP3), making it difficult to evaluate the effect of the sequence context at these two positions in the original *EVI2B* RNA substrate (Figure 5(B–C) and Figure S6(B)).

To identify the most preferred RNA substrate for A3A editing, the sequences on the modified *EIV2B* RNA hairpin stem (ST5) and loop (LP4) were systematically altered and tested for A3A editing efficiency. For the stem base-pair, we find that the sequence of the first stem base-pair adjoint to the loop matters. A CG pair showed the highest A3A editing efficiency than other base pairs, with the order of preference of CG > UA > GC > AU (Figure 5(D)). For the RNA loop, the −3 position no longer can form a base pair with +1 in a

4 nt loop (LP4), and our findings indicate that the nt preference at −3 for A3A RNA editing was in the order of C > U > G > A. The −2 position nt preference was shown to be A > U > G > C (Figure 5(D)).

Based on the above findings on preferred substrate features for A3A editing, we made a few hairpin RNAs containing LP4 and ST5 containing various combination of the 4 nt loop sequences and compare the editing efficiencies on these RNAs by A3A. The results showed a clear ranking of editing efficiency of these RNA substrates (Figure 5(D–E) and Figure S6(C)). The preferred loop sequence was CAUC > CUUC > UAUC > UUUC (target C sites are underlined). Additionally, the first stem base pair CG exhibits approximately 30% higher than the UA pair (Figure 5(E)). The ranking of the six stem-loop RNA sequences as preferred substrates for A3A editing is shown in Figure 5(F).

Prediction and validation of potential RNA targets for A3A editing

We attempted to predict potential RNA targets for A3A editing based on the RNA sequence and structural features identified as the favored substrates for A3A editing (Figure 5(E–F)). By searching the RNA database by inputting the 12 nt RNA sequences in the six RNA groups (group 1-6) depicted in Fig. 5F, we identified a total of 1162 sites in the coding region of human Refseq RNA using the NCBI Blast tool (Figure 6(A), Table S1, and See Methods). Further analysis of these 1162 RNA target sites revealed 175 of these sites overlap with previously reported sites [17, 18, 51], and 987 of these sites are not identified before (Figure 6(B), Figure S7, and Table S1).

To validate the prediction of A3A editing at the predicted sites of cellular RNAs, we selected a total of 30 cellular RNA transcripts with predicted stop-gain or non-synonymous codon changes, including 22 novel RNAs and 8 previously reported RNAs, for examining the endogenous RNA editing levels by sequencing analysis (Figure 6(C) and Table S2). Our results confirmed that all 30 selected sites displayed A3A RNA editing (Figure 6(C) and Figure S8). Consistent with our prediction, the analysis of the results showed that the average level of A3A-induced RNA editing on the tetra-loop (LP4) sites follows the order: CAUC > CUUC > UAUC (target C underlined)) (Table S2). Furthermore, the overall RNA editing efficiency was higher in the presence of CG pairing in the first stem base-pair than the UA pairing (Table S2). Additionally, the increased stability of the hairpin structure, which arises from longer stem lengths, has a significant impact on the elevated level of RNA editing activity. (Table S2 and Figure S9(A)). Therefore, multiple factors, including sequence context and RNA secondary structure, are involved in determining the A3A-mediated RNA editing activity.

Based on the known target RNA sequences for A3A as shown in the 6 RNA groups (Figure 5(F) and Figure 6(A)), we evaluated the synonymous, nonsynonymous, or stop codon changes following C-to-U editing by A3A in the human mRNAs that contain the sequences of the RNA groups 1-6. Interestingly, we found that the highest proportion of codon mutations were synonymous (approximately 80% of the codon mutations) for the sites containing the sequence group exhibiting the highest A3A editing activity (Group-1). The percentage of synonymous codon mutations of these mRNAs generally decreased for less efficiently edited sites (from Group-2 to Group-6) after A3A editing (Figure 6(D)).

On the other hand, the proportion of nonsynonymous plus the stop codon mutations of the mRNA increased for the mRNA sequences from Group-1 to Group-6 with lower A3A editing-efficiency (Figure 6(D)). There was an obvious positive correlation between the ratio of synonymous codon changes and the observed relative editing efficiency among these sequence contexts (Figure 6(E)).

DISCUSSION

DNA mutation by APOBEC deamination, if not repaired, is permanent and changes the entire corresponding RNA transcript, which can lead to various genetic diseases, whereas RNA editing by APOBEC is a transient event that only affects the partially edited RNA transcript. A better understanding of RNA editing by A3A and other APOBECs is necessary for understanding their biological functions and for developing strategies for programmable targeted RNA editing for potential therapy [48, 54]. In this report, we characterized the detailed features of RNA substrates for A3A editing and used that information to predict and validate new cellular RNA targets for A3A.

It has been reported that RNA editing activity by A1 plus its cofactors is generally suppressed in the cytoplasm and occurs within the nucleus in a temporal and spatial framework after pre-mRNA splicing and before mRNA nuclear export [55–57]. We found that changing A3A subcellular localization by fusion of A3A with a nuclear localization signal (NLS) led to a significant decrease in RNA editing activity (Figure 3). This result suggests that the primary subcellular location for A3A-mediated RNA editing is the cytoplasm, which is contrary to that of A1 plus its cofactors. However, despite the close sequence similarities between A3A and A3B-CD2 (92% identity) [12, 14], re-localizing A3B from nucleus to cytoplasm did not show any RNA editing activity on the *EVI2B* and *SDHB* RNAs that are known substrates for A3A (Figure 3).

Nonetheless, an A3B chimera with its CD2 domain DPLVLRRQ peptide replaced with GIGRHK peptide from A3A showed editing on both A3A substrates in cytosolic and nuclear localized versions of the A3B chimeras (Figure 4(A) and Figure S4). These findings suggest that the small sequence difference between A3A and A3B-CD2 determines RNA editing on specific RNA targets, and that other features of A3B, such as A3B-CD1, may influence its ability to edit RNAs in both the cytosol and nucleus of the same RNA targets, contrary to A3A and A1 in that RNA editing appears to occur primarily in the cytoplasm or nucleus, respectively. These results also imply that A3B may also possess intrinsic RNA editing activity on other substrates, which is consistent with a prior report of potential association of breast cancer with RNA editing by A3B [58] and a recent BioRxiv report describing RNA editing events by overexpression of human A3B in a mouse model [59].

Previous studies showed that A3A-mediated RNA editing prefers a stem-loop structure with the target C at the 3’-end of the loop [17, 18, 21, 30, 51]. We performed comprehensive analysis of the effect of the stem-loop structures and sequence context on A3A-induced RNA editing using our cell based fluorescent assay system. Besides confirming previous findings, we found that the first stem base-pair type and particular sequence of the loop also had significant impact on A3A editing efficiency (Figure 5). The investigation on optimal

single-stranded DNA sequences for interaction with A3A [27] and the latest NMR structures of the A3A-DNA complex [60] have underscored the importance of T at −1 and A at +1 for positioning the target C at 0. Considering the specific characteristics of our RNA sequences, which include U at −1 and a preference for purine (G/A) at +1, it is plausible to propose that the interaction between A3A and RNA may resemble the A3A-ssDNA interaction within the catalytic active region. However, for a more precise analysis, it is crucial to elucidate the A3A-RNA structure using appropriate RNA sequences.

Based on the findings on the impact on A3A editing by the sequence context of the RNA stem-loop, we were able to predict potential cellular RNA substrates (Figure 5(F)), by entering query sequences showing optimal RNA editing activity to search for mRNAs in the NCBI blast tool. Our sequence-based prediction results have been validated from two angles. First, a good fraction of the predicted RNA targets for A3A overlaps with previously reported targets [17, 18, 51], with 987 targets that were not identified before. Second, sanger sequencing of 30 randomly selected predicted mRNA candidates extracted from cells over-expressing A3A all showed A3A-editing at the expected sites (Figure 6(C) and Table S2). These results demonstrated that our prediction system is useful for identifying A3A's RNA targets that have not been discovered due to technical limitations of RNA sequencing and/or limited expression of transcriptomes in certain cell types. However, as the prediction system may deviate from the expected RNA secondary structure, especially when RNA molecules dynamically interact and fold with their own or other RNAs *in vivo* [61], it is imperative to establish an additional system that considers the overall dynamics of global RNA folding to achieve a more accurate prediction system. Further studies are needed to understand the biological consequences of the RNA editing by A3A.

The analysis of potential mRNA target sites for A3A editing showed that sequences for higher A3A-editing groups were associated with a higher frequency of synonymous codon changes by the C-to-U mutation (Figure 6(D–E)). Despite a high percentage of synonymous codon changes (702 sites, 60.4%) following C-to-U mutation in these high A3A-editing groups, there are still many sites of non-synonymous (294 sites, 25.3%) or stop-gain (166 sites, 14.3%) codon changes by A3A editing (Figure 6(D)). These findings suggest that, while A3A specific RNA editing is used for certain biological functions, cells have evolved a mechanism to minimize the potentially negative impact of too much mutation on the mRNA carrying the sequences that can be efficiently edited by A3A. We performed gene ontology analysis on a set of 460 predicted sites that exhibited codon changes to analyze their potential biological functions or cellular pathways (Figure S9(B)). This analysis enables the investigation of potential indirect functions of A3A through the A3A-mediated RNA editing, such as protein de-ubiquitination or vesicular cargo loading (Figure S9(B)). Furthermore, we utilized the NCBI ClinVar tool to analyze whether our predicted sites are reported as mutations in patients with diseases. Out of the total 59 identified sites, most of them exhibited likely benign or potentially clinical variations. However, three genes (ASPM, ATM, and SDHB) were found to have pathogenic significance associated with these mutations (Table S3). Despite showing synonymous codon changes and having a likely benign variation of clinical significance, DDOST mRNA displayed the highest C-to-U RNA editing in cancer patients with high endogenous A3A expression (Table S3) [51]. Our A3A RNA editing prediction system successfully identified the DDOST mRNA site within

Group 1, characterized by the highest editing rate (Table S1). These findings underscore the potential of our prediction system in enhancing the understanding of RNA editing sites associated with A3A-related diseases, including cancer.

MATERIALS AND METHODS

Transfection

The experimental methods utilized for this study involved culturing HEK 293T cells in DMEM media supplemented with 10% FBS, streptomycin (100μg/mL), and penicillin (100U/mL) and incubated at 37°C with 5% CO2. To achieve transfection, 8-well glass slides (CellVis) were first coated with 0.1 mg/mL poly-D-lysine (Sigma). The cells were then diluted to a concentration of 250,000 cells/mL to ensure a clean monolayer for visualization during experimentation. Subsequently, 250 μL of the diluted cell suspension was added to each well. After 24 hours of initial adherence and growth, the cells were transfected using Lipofectamine 3000 Transfection Reagent (Thermo Fisher) and a mixture of APOBEC editor (500ng) and substrate reporter vector (50ng) in 25μL volumes, following the manufacturer's recommended instructions. The plasmids used in this study have been provided in Table S4. The mixture was allowed to sit at room temperature for 30 minutes before 15 μL was added dropwise to each well. Expression was then allowed to occur for 48 hours.

Confocal fluorescence microscopy

For live cell microscopy in these experiments, an inverted Zeiss LSM-700 confocal microscope with a 40x water-immersion objective was utilized to provide optimal visualization for analysis and maximum cell countability per image. Prior to imaging, cells were washed with phosphate-buffered saline (PBS) and stained with a 5 μg/mL solution of Hoechst 33342 nuclear stain in PBS for 15 minutes. The cells were then rinsed twice with PBS and stored in imaging buffer (140 mM NaCl, 2.5 mM KCl, 1.8 mM $CaCl_2$, 1.0 mM $MgCl_2$, 20 mM Hepes 7.4, 5 mM glucose). Imaging was conducted at a higher laser intensity (15-20%) and lower gain (500-600 units) to enhance signal-to-noise ratio. Excitation wavelengths for Hoechst 33342, eGFP, and mCherry were 405, 488, and 555 nm, respectively, and the emission band-pass filters were set to 400-480 nm, 490-555 nm, and 555-700 nm, respectively. Approximately 3-5 images were captured for each well, and a total of 30 cells were quantified. All image analysis was performed using the LSM Toolbox plugin incorporated within the FIJI distribution of ImageJ2 Software. [62].

RNA/DNA extraction for Sanger Sequencing

To extract RNA, the cells were harvested and Trizol reagent (Thermo Fisher) was added directly to the wells to lyse the cells. The manufacturer's protocol for Trizol was followed, with the addition of 50 μL of chloroform and the extraction of approximately 40 μL of the aqueous phase. Total RNA was then reverse transcribed into single-stranded cDNA using ProtoScript II reverse transcriptase (NEB) and a specific primer designed to anneal downstream of substrate reporter segments. The reaction was conducted in a 20 μL volume, which contained 1 μg of total RNA, 100 μM of reverse primer, 10 mM dNTP, 0.1 M DTT, 8 U RNase inhibitor, and 0.2 μL of ProtoScript II reverse transcriptase (NEB), at 42°C for

1 hour. The cDNA was amplified with Phusion® High-Fidelity DNA Polymerase (NEB) for 30 cycles (98°C: 2.5 min - (98°C: 20s, 71.7°C: 20s, 72°C: 30s)×30 - 72°C: 5 min) using a forward primer that anneals to the junction region where the AAV intron was spliced out. To extract DNA, harvested cells were treated with approximately 100 μL of QuickExtract™ DNA Extraction Solution (Lucigen), vortexed, and incubated at 65 °C for 15 minutes, followed by 2 minutes at 98°C and cooled to 4°C. 1 μL of the extracted DNA was used in a PCR reaction with Phusion® High-Fidelity DNA Polymerase (NEB) and a forward primer that targets a region containing the AAV intron. The PCR reaction was performed using the same cycle as the cDNA amplification. The final PCR products were purified using a spin column PCR cleanup kit (Thermo Fisher) and submitted to Genwiz for Sanger sequencing.

Western blot and antibodies

To assess the expression levels of FLAG-tagged APOBECs in cells, Western blot analysis was performed. Cells were lysed using 1x RIPA buffer (Sigma) to prepare lysates. The Western blot analysis was carried out from two independent transfections, and α-Tubulin was used as an internal loading control. Primary antibodies used in this analysis included anti-FLAG M2 mAb (F3165, Sigma) diluted 1:3,000 and anti-α-tubulin mAb from mouse (GT114, GeneTex) diluted 1:5,000. A secondary antibody, Cy3-labelled goat-anti-mouse mAb (PA43009, GE Healthcare) was subsequently used. The Cy3 signals were detected and visualized using the Typhoon RGB Biomolecular Imager (GE Healthcare).

Prediction of potential target RNA candidates for A3A editing

A Python script was utilized to generate stem-loop sequence queries using the arrangements package, as detailed in Figure S10. Briefly, a sequence corresponding to N x N x N was determined for each group, and the complementary sequence was automatically assigned to obtain the opposite stem (e.g., A-U, C-G, G-C, and U-A), resulting in 64 queries per group (4 x 4 x 4 = 64). A total of 384 queries (64 x 6 groups = 384) were generated for the six groups. These queries were utilized to search for Refseq RNA sites with a 100% match on NCBI Blastn (https://blast.ncbi.nlm.nih.gov/Blast.cgi). The parameters used were as follows: “Blastn -db refseq_rna_db/refseq_rna -query stem_size_4.fa -strand ‘plus’ -evalue 1000-word_size 7 -taxids 9606 -perc_identity 100 -qcov_hsp_perc 100 -outfmt ‘6 delim = qaccver stitle saccver pident length sstart send’ -out stem_size_4 -num_threads 4”. To obtain the data presented in Table S1 (or the results in Figure 6(A)), we extracted the coding sequence (CDS) features of mRNA from the search results and subsequently removed any overlapping sites.

Analysis tools

In this experiment, all graphs were generated using GraphPad Prism 9, and statistical significance was evaluated using a two-tailed Student’s t-test implemented in GraphPad Prism 9. To predict the RNA secondary structures for the local RNA region, we employed the RNAstructure software [63, 64]. Gene ontology analysis was conducted on http://geneontology.org/. Analysis of mutation-related diseases was performed by inputting all predicted sites into NCBI ClinVar (https://www.ncbi.nlm.nih.gov/clinvar/) for verification (Table S3).

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

ACKNOWLEDGEMENT

We acknowledge the support of the Translational Imaging Center at the University of Southern California for providing the fluorescent microscope. We thank Zifan Zhu and Fengzhu Sun (Department of Quantitative and Computational Biology, University of Southern California) for helpful advice designing the prediction system in this study. This work is supported by the NIH grant R01 AI150524 to X.S.C.

REFERENCES

[1]. Baysal BE, Sharma S, Hashemikhabir S, Janga SC. RNA Editing in Pathogenesis of Cancer. Cancer Res. 2017;77:3733–9. [PubMed: 28667076]

[2]. Christofi T, Zaravinos A. RNA editing in the forefront of epitranscriptomics and human health. J Transl Med. 2019;17:319. [PubMed: 31547885]

[3]. Gagnidze K, Rayon-Estrada V, Harroch S, Bulloch K, Papavasiliou FN. A New Chapter in Genetic Medicine: RNA Editing and its Role in Disease Pathogenesis. Trends Mol Med. 2018;24:294–303. [PubMed: 29483039]

[4]. Lerner T, Papavasiliou FN, Pecori R. RNA Editors, Cofactors, and mRNA Targets: An Overview of the C-to-U RNA Editing Machinery and Its Implication in Human Disease. Genes (Basel). 2018;10. [PubMed: 30586904]

[5]. Pecori R, Di Giorgio S, Paulo Lorenzo J, Nina Papavasiliou F. Functions and consequences of AID/APOBEC-mediated DNA and RNA deamination. Nat Rev Genet. 2022;23:505–18. [PubMed: 35256818]

[6]. Ramaswami G, Lin W, Piskol R, Tan MH, Davis C, Li JB. Accurate identification of human Alu and non-Alu RNA editing sites. Nat Methods. 2012;9:579–81. [PubMed: 22484847]

[7]. Nishikura K. Functions and regulation of RNA editing by ADAR deaminases. Annu Rev Biochem. 2010;79:321–49. [PubMed: 20192758]

[8]. Pecori R, Papavasiliou NF. It takes two (and some distance) to tango: how ADARs join to edit RNA. Nat Struct Mol Biol. 2020;27:308–10. [PubMed: 32221506]

[9]. Chen SH, Habib G, Yang CY, Gu ZW, Lee BR, Weng SA, et al. Apolipoprotein B-48 is the product of a messenger RNA with an organ-specific in-frame stop codon. Science. 1987;238:363–6. [PubMed: 3659919]

[10]. Driscoll DM, Wynne JK, Wallis SC, Scott J. An in vitro system for the editing of apolipoprotein B mRNA. Cell. 1989;58:519–25. [PubMed: 2758465]

[11]. Teng B, Burant CF, Davidson NO. Molecular cloning of an apolipoprotein B messenger RNA editing protein. Science. 1993;260:1816–9. [PubMed: 8511591]

[12]. Conticello SG. The AID/APOBEC family of nucleic acid mutators. Genome Biol. 2008;9:229.

[13]. Salter JD, Bennett RP, Smith HC. The APOBEC Protein Family: United by Structure, Divergent in Function. Trends Biochem Sci. 2016;41:578–94. [PubMed: 27283515]

[14]. Yang B, Li X, Lei L, Chen J. APOBEC: From mutator to editor. J Genet Genomics. 2017;44:423–37. [PubMed: 28964683]

[15]. Prohaska KM, Bennett RP, Salter JD, Smith HC. The multifaceted roles of RNA binding in APOBEC cytidine deaminase functions. Wiley Interdiscip Rev RNA. 2014;5:493–508. [PubMed: 24664896]

[16]. Niavarani A, Currie E, Reyal Y, Anjos-Afonso F, Horswell S, Griessinger E, et al. APOBEC3A is implicated in a novel class of G-to-A mRNA editing in WT1 transcripts. PLoS One. 2015;10:e0120089. [PubMed: 25807502]

[17]. Sharma S, Patnaik SK, Taggart RT, Kannisto ED, Enriquez SM, Gollnick P, et al. APOBEC3A cytidine deaminase induces RNA editing in monocytes and macrophages. Nat Commun. 2015;6:6881. [PubMed: 25898173]

[18]. Sharma S, Patnaik SK, Kemer Z, Baysal BE. Transient overexpression of exogenous APOBEC3A causes C-to-U RNA editing of thousands of genes. RNA Biol. 2017;14:603–10. [PubMed: 27149507]

[19]. Sharma S, Patnaik SK, Taggart RT, Baysal BE. The double-domain cytidine deaminase APOBEC3G is a cellular site-specific RNA editing enzyme. Sci Rep. 2016;6:39100. [PubMed: 27974822]

[20]. Sharma S, Wang J, Alqassim E, Portwood S, Cortes Gomez E, Maguire O, et al. Mitochondrial hypoxic stress induces widespread RNA editing by APOBEC3G in natural killer cells. Genome Biol. 2019;20:37. [PubMed: 30791937]

[21]. Kim K, Calabrese P, Wang S, Qin C, Rao Y, Feng P, et al. The roles of APOBEC-mediated RNA editing in SARS-CoV-2 mutations, replication and fitness. Sci Rep. 2022;12:14972. [PubMed: 36100631]

[22]. Nakata Y, Ode H, Kubota M, Kasahara T, Matsuoka K, Sugimoto A, et al. Cellular APOBEC3A deaminase drives mutations in the SARS-CoV-2 genome. Nucleic Acids Res. 2023;51:783–95. [PubMed: 36610792]

[23]. Bogerd HP, Wiegand HL, Doehle BP, Lueders KK, Cullen BR. APOBEC3A and APOBEC3B are potent inhibitors of LTR-retrotransposon function in human cells. Nucleic Acids Res. 2006;34:89–95. [PubMed: 16407327]

[24]. Chen H, Lilley CE, Yu Q, Lee DV, Chou J, Narvaiza I, et al. APOBEC3A is a potent inhibitor of adeno-associated virus and retrotransposons. Curr Biol. 2006;16:480–5. [PubMed: 16527742]

[25]. Thielen BK, McNevin JP, McElrath MJ, Hunt BV, Klein KC, Lingappa JR. Innate immune signaling induces high levels of TC-specific deaminase activity in primary monocyte-derived cells through expression of APOBEC3A isoforms. J Biol Chem. 2010;285:27753–66. [PubMed: 20615867]

[26]. Schutsky EK, Nabel CS, Davis AKF, DeNizio JE, Kohli RM. APOBEC3A efficiently deaminates methylated, but not TET-oxidized, cytosine bases in DNA. Nucleic Acids Res. 2017;45:7655–65. [PubMed: 28472485]

[27]. Chan K, Roberts SA, Klimczak LJ, Sterling JF, Saini N, Malc EP, et al. An APOBEC3A hypermutation signature is distinguishable from the signature of background mutagenesis by APOBEC3B in human cancers. Nat Genet. 2015;47:1067–72. [PubMed: 26258849]

[28]. Cortez LM, Brown AL, Dennis MA, Collins CD, Brown AJ, Mitchell D, et al. APOBEC3A is a prominent cytidine deaminase in breast cancer. PLoS Genet. 2019;15:e1008545. [PubMed: 31841499]

[29]. Buisson R, Langenbucher A, Bowen D, Kwan EE, Benes CH, Zou L, et al. Passenger hotspot mutations in cancer driven by APOBEC3A and mesoscale genomic features. Science. 2019;364. [PubMed: 31624212]

[30]. Sharma S, Baysal BE. Stem-loop structure preference for site-specific RNA editing by APOBEC3A and APOBEC3G. PeerJ. 2017;5:e4136. [PubMed: 29230368]

[31]. Wang Y, Schmitt K, Guo K, Santiago ML, Stephens EB. Role of the single deaminase domain APOBEC3A in virus restriction, retrotransposition, DNA damage and cancer. J Gen Virol. 2016;97:1–17. [PubMed: 26489798]

[32]. Olson ME, Harris RS, Harki DA. APOBEC Enzymes as Targets for Virus and Cancer Therapy. Cell Chem Biol. 2018;25:36–49. [PubMed: 29153851]

[33]. Swanton C, McGranahan N, Starrett GJ, Harris RS. APOBEC Enzymes: Mutagenic Fuel for Cancer Evolution and Heterogeneity. Cancer Discov. 2015;5:704–12. [PubMed: 26091828]

[34]. Gohler S, Da Silva Filho MI, Johansson R, Enquist-Olsson K, Henriksson R, Hemminki K, et al. Impact of functional germline variants and a deletion polymorphism in APOBEC3A and APOBEC3B on breast cancer risk and survival in a Swedish study population. J Cancer Res Clin Oncol. 2016;142:273–6. [PubMed: 26320772]

[35]. Schmitt K, Guo K, Algaier M, Ruiz A, Cheng F, Qiu J, et al. Differential virus restriction patterns of rhesus macaque and human APOBEC3A: implications for lentivirus evolution. Virology. 2011;419:24–42. [PubMed: 21868050]

[36]. Smith HC. Measuring editing activity and identifying cytidine-to-uridine mRNA editing factors in cells and biochemical isolates. Methods Enzymol. 2007;424:389–416. [PubMed: 17662851]

[37]. Chieca M, Torrini S, Conticello SG. Live-Cell Quantification of APOBEC1-Mediated RNA Editing: A Comparison of RNA Editing Assays. Methods Mol Biol. 2021;2181:69–81. [PubMed: 32729075]

[38]. Severi F, Conticello SG. Flow-cytometric visualization of C>U mRNA editing reveals the dynamics of the process in live cells. RNA Biol. 2015;12:389–97. [PubMed: 25806564]

[39]. Wolfe AD, Arnold DB, Chen XS. Comparison of RNA Editing Activity of APOBEC1-A1CF and APOBEC1-RBM47 Complexes Reconstituted in HEK293T Cells. J Mol Biol. 2019;431:1506–17. [PubMed: 30844405]

[40]. Oh S, Buisson R. A digital PCR-based protocol to detect and quantify RNA editing events at hotspots. STAR Protoc. 2022;3:101148. [PubMed: 35284835]

[41]. Acharya N, Kumar P, Varshney U. Complexes of the uracil-DNA glycosylase inhibitor protein, Ugi, with Mycobacterium smegmatis and Mycobacterium tuberculosis uracil-DNA glycosylases. Microbiology. 2003;149:1647–58. [PubMed: 12855717]

[42]. Prochnow C, Bransteitter R, Klein MG, Goodman MF, Chen XS. The APOBEC-2 crystal structure and functional implications for the deaminase AID. Nature. 2007;445:447–51. [PubMed: 17187054]

[43]. Marino D, Perkovic M, Hain A, Jaguva Vasudevan AA, Hofmann H, Hanschmann KM, et al. APOBEC4 Enhances the Replication of HIV-1. PLoS One. 2016;11:e0155422. [PubMed: 27249646]

[44]. Muckenfuss H, Hamdorf M, Held U, Perkovic M, Lower J, Cichutek K, et al. APOBEC3 proteins inhibit human LINE-1 retrotransposition. J Biol Chem. 2006;281:22161–72. [PubMed: 16735504]

[45]. Lackey L, Law EK, Brown WL, Harris RS. Subcellular localization of the APOBEC3 proteins during mitosis and implications for genomic DNA deamination. Cell Cycle. 2013;12:762–72. [PubMed: 23388464]

[46]. Land AM, Law EK, Carpenter MA, Lackey L, Brown WL, Harris RS. Endogenous APOBEC3A DNA cytosine deaminase is cytoplasmic and nongenotoxic. J Biol Chem. 2013;288:17253–60. [PubMed: 23640892]

[47]. Salamango DJ, McCann JL, Demir O, Brown WL, Amaro RE, Harris RS. APOBEC3B Nuclear Localization Requires Two Distinct N-Terminal Domain Surfaces. J Mol Biol. 2018;430:2695–708. [PubMed: 29787764]

[48]. Huang X, Lv J, Li Y, Mao S, Li Z, Jing Z, et al. Programmable C-to-U RNA editing using the human APOBEC3A deaminase. EMBO J. 2020;39:e104741. [PubMed: 33058229]

[49]. Fu Y, Ito F, Zhang G, Fernandez B, Yang H, Chen XS. DNA cytosine and methylcytosine deamination by APOBEC3B: enhancing methylcytosine deamination by engineering APOBEC3B. Biochem J. 2015;471:25–35. [PubMed: 26195824]

[50]. Ito F, Fu Y, Kao SA, Yang H, Chen XS. Family-Wide Comparative Analysis of Cytidine and Methylcytidine Deamination by Eleven Human APOBEC Proteins. J Mol Biol. 2017;429:1787–99. [PubMed: 28479091]

[51]. Jalili P, Bowen D, Langenbucher A, Park S, Aguirre K, Corcoran RB, et al. Quantification of ongoing APOBEC3A activity in tumor cells by monitoring RNA editing at hotspots. Nat Commun. 2020;11:2971. [PubMed: 32532990]

[52]. Garg A, Heinemann U. A novel form of RNA double helix based on G.U and C.A(+) wobble base pairing. RNA. 2018;24:209–18. [PubMed: 29122970]

[53]. Love RP, Xu H, Chelico L. Biochemical analysis of hypermutation by the deoxycytidine deaminase APOBEC3A. J Biol Chem. 2012;287:30812–22. [PubMed: 22822074]

[54]. Abudayyeh OO, Gootenberg JS, Franklin B, Koob J, Kellner MJ, Ladha A, et al. A cytosine deaminase for programmable single-base RNA editing. Science. 2019;365:382–6. [PubMed: 31296651]

[55]. Lau PP, Xiong WJ, Zhu HJ, Chen SH, Chan L. Apolipoprotein B mRNA editing is an intranuclear event that occurs posttranscriptionally coincident with splicing and polyadenylation. J Biol Chem. 1991;266:20550–4. [PubMed: 1939106]

[56]. Sowden M, Hamm JK, Spinelli S, Smith HC. Determinants involved in regulating the proportion of edited apolipoprotein B RNAs. RNA. 1996;2:274–88. [PubMed: 8608451]

[57]. Sowden MP, Smith HC. Commitment of apolipoprotein B RNA to the splicing pathway regulates cytidine-to-uridine editing-site utilization. Biochem J. 2001;359:697–705. [PubMed: 11672445]

[58]. Asaoka M, Ishikawa T, Takabe K, Patnaik SK. APOBEC3-Mediated RNA Editing in Breast Cancer is Associated with Heightened Immune Activity and Improved Survival. Int J Mol Sci. 2019;20.

[59]. Alonso de la Vega A, Temiz NA, Tasakis R, Somogyi K, Reuveni E, Ben-David U, et al. Acute expression of human APOBEC3 B in mice causes lethality associated with RNA editing. 2022:2022.06.01.494353.

[60]. Liu Y, Lan W, Wang C, Cao C. Two different kinds of interaction modes of deaminase APOBEC3A with single-stranded DNA in solution detected by nuclear magnetic resonance. Protein Sci. 2022;31:443–53. [PubMed: 34792260]

[61]. Schroeder SJ. Challenges and approaches to predicting RNA with multiple functional structures. RNA. 2018;24:1615–24. [PubMed: 30143552]

[62]. Schneider CA, Rasband WS, Eliceiri KW. NIH Image to ImageJ: 25 years of image analysis. Nat Methods. 2012;9:671–5. [PubMed: 22930834]

[63]. Mathews DH, Disney MD, Childs JL, Schroeder SJ, Zuker M, Turner DH. Incorporating chemical modification constraints into a dynamic programming algorithm for prediction of RNA secondary structure. Proc Natl Acad Sci U S A. 2004;101:7287–92. [PubMed: 15123812]

[64]. Reuter JS, Mathews DH. RNAstructure: software for RNA secondary structure prediction and analysis. BMC Bioinformatics. 2010;11:129. [PubMed: 20230624]

Highlights

- APOBEC3A exhibits unique and specific RNA editing activity distinct from the other APOBEC family members, and this activity primarily takes place in the cytoplasm.
- APOBEC3B chimeras containing APOBEC3A-loop1 region are capable of RNA editing activity on APOBEC3A's RNA substrates.
- Specific RNA sequence contexts in stem-loop structures significantly affect APOBEC3A-mediated RNA editing efficiency, and this knowledge can be used to predict and validate novel RNA targets.
- The observed trend of increased synonymous mutations at sites with higher APOBEC3A-induced RNA editing efficiency suggests the potential impact on evolutionary adaptation.

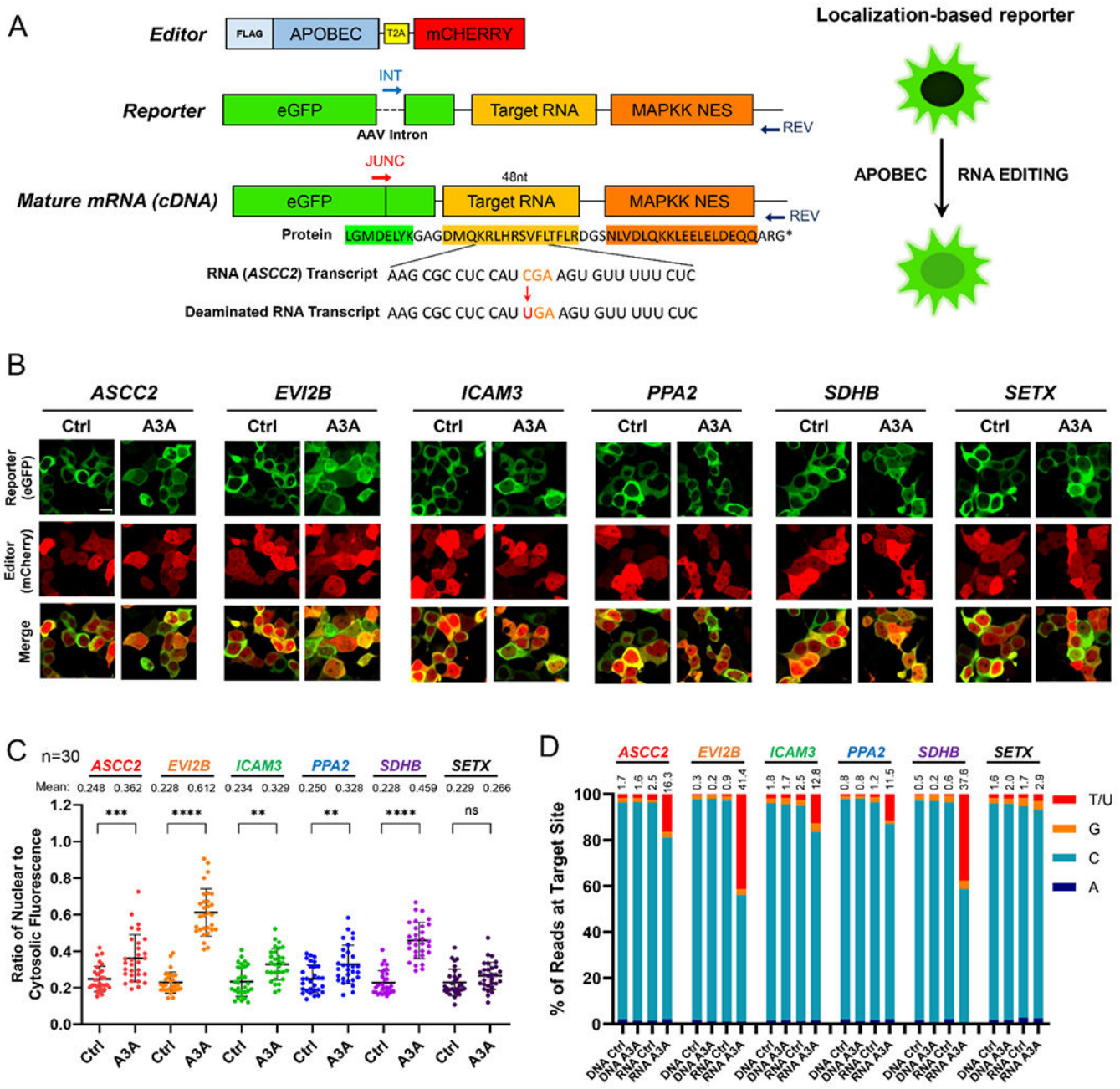


Figure 1. A sensitive cell-based fluorescence analysis and sequencing method confirms A3A-mediated specific RNA editing.

(A) Design of specific RNA editing by APOBEC proteins using a sensitive cell-based fluorescence assay. The editing vector contains APOBEC protein (A3A in this case) and mCherry protein for transfection validation at the C-terminus following the self-cleaved T2A peptides. The reporter vector is labeled with eGFP containing 48 nt target RNA, and the expressed reporter protein is localized in the cytoplasm by a strong MAPKK NES signal. When the target RNA is generated with an early stop codon due to C-to-U deamination by the editor (A3A), the reporter eGFP can be localized throughout the cell, including the nucleus (right panel). **(B)** Confocal microscopy images show eGFP and

mCherry signals by co-transfection of the editor vector (A3A in this case) and each reporter vector. (Scale bar: 10μm). **(C)** Quantification of eGFP nuclear/cytoplasmic fluorescence ratio change indicates A3A-mediated target C-to-U editing levels. Statistical significance was calculated by unpaired two-tailed student's t-test with P-values represented as: P > 0.05 = not signifcant; ns, ** P < 0.01, *** P < 0.001, and **** P < 0.0001. **(D)** Sanger sequencing data of the nucleotides % ratio at the target C site analyzed from DNA and RNA, respectively. The nucleotide percentage was quantified by the area of each nucleotide in the sequencing chromatogram.

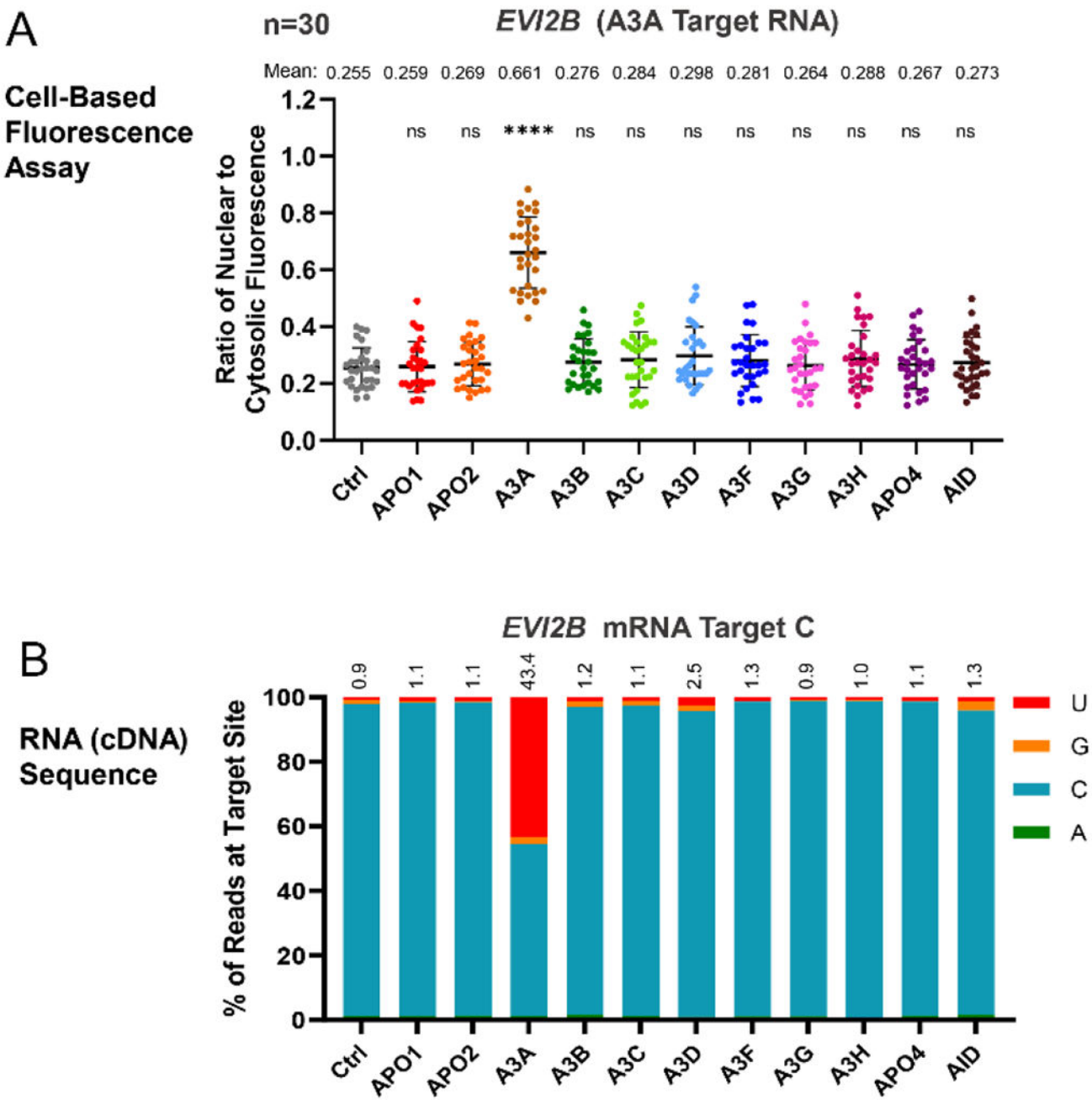


Figure 2. C-to-U RNA Editing on *EVI2B* RNA substrate by all APOBEC family members.

(A) Quantification of EVI2B RNA editing by all other APOBEC proteins using the eGFP nuclear/cytosolic ratio change assay. Data are represented as mean values ± SD indicated by dots (n=30 cells). Statistical significance was calculated between control (Ctrl) and each APOBEC protein by unpaired two-tailed student's t-test with P-values represented as: P > 0.05 = not significant; ns, and **** P < 0.0001. **(B)** Quantification of the nucleotide distribution at *EVI2B* target C site analyzed by RNA (cDNA) sequencing. Only A3A, but not other APOBEC proteins, edits the *EVI2B* target C site.

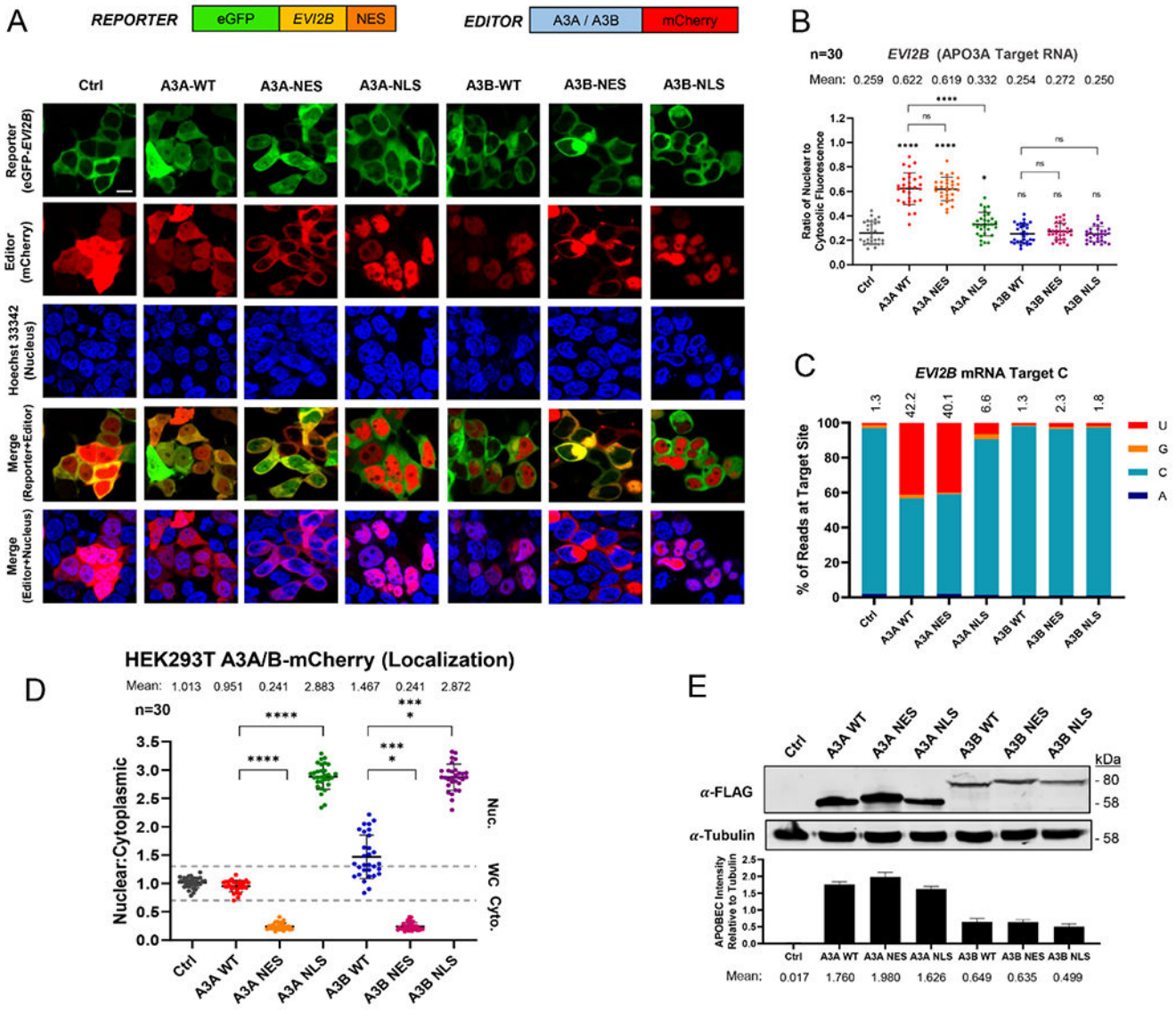


Figure 3. A3A-mediated RNA editing primarily occurs in the cytoplasm.

(A) Representative fluorescence images of HEK293T cells co-transfected with the editor vectors (A3A-WT, A3A-NES, A3A-NLS, A3B-WT, A3B-NES, and A3B-NLS) labeled with mCherry and the EVI2B reporter vector labeled with eGFP. The reporter protein localization indicates RNA editing activity, and the nuclear periphery is stained with Hoechst 33342 (scale bar: 10 μm). The reporter and editor vectors are described in the cartoons above the image panel. **(B)** Quantification of the subcellular localization of the eGFP reporter (Nuclear/Cytosolic ratio). Statistical significance was indicated without a scale for comparison with Ctrl, and comparison with each wild type was indicated with scales. **(C)** Nucleotide sequencing distribution on the target *EVI2B* C site to analyze C-to-U editing levels. **(D)** Quantification of subcellular localization of each mCherry-labeled editor. **(E)** Western blot showing ectopic transient expression of A3A/B in HEK293T cells. Relative expression intensities were quantified as mean values ± SD in duplicate independent

experiments (n = 2) and normalized to Tubulin endogenous loading control. Statistical significances in (B) and (D) were calculated by unpaired two-tailed student's t-test with P-values represented as: P > 0.05 = not significant; ns, * 0.01 < P < 0.05, and **** P < 0.0001.

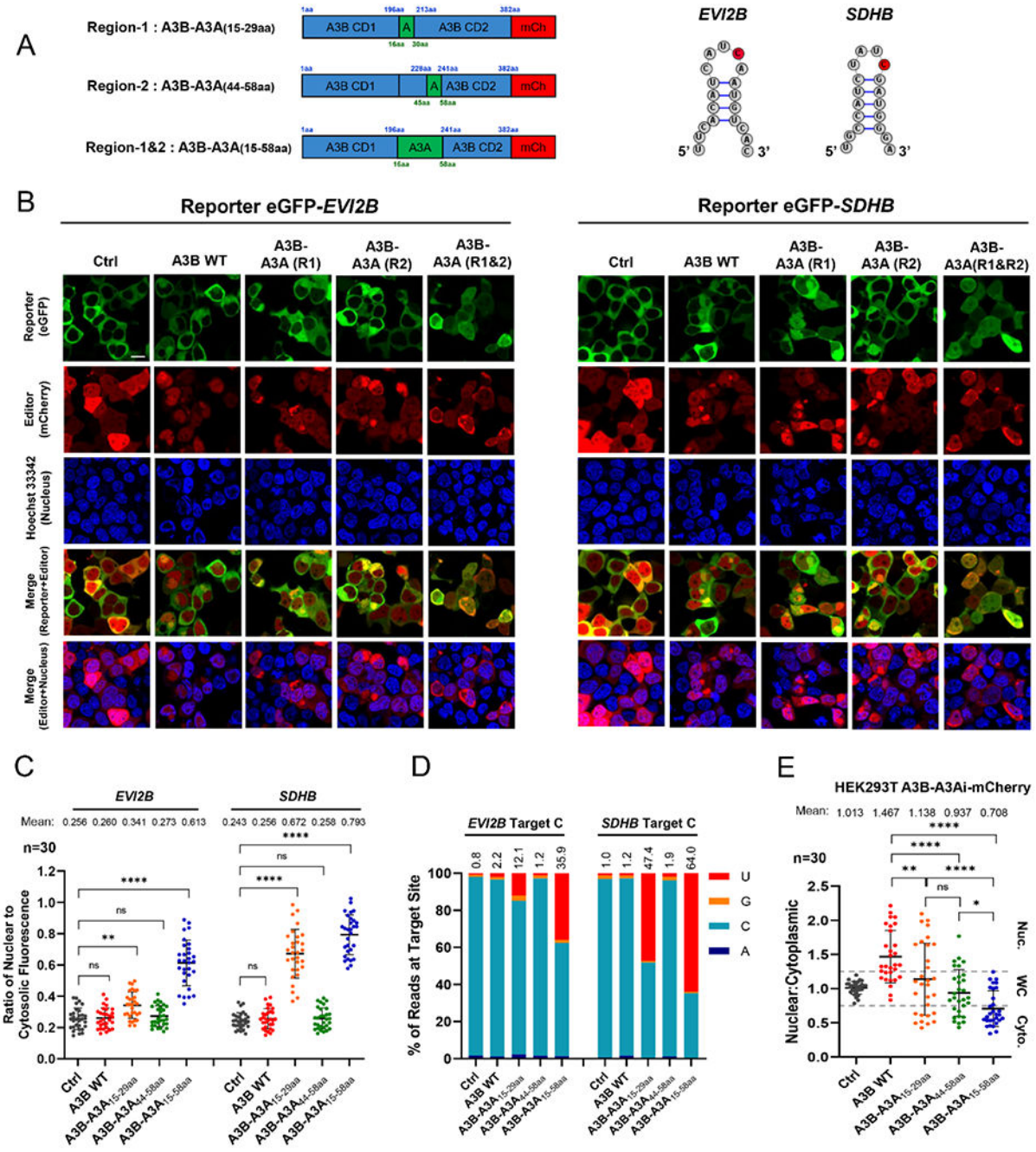


Figure 4. Specific RNA editing activity of A3B chimeras containing short A3A N-terminal peptide sequences.

(A) Schematic descriptions of A3B chimera editor vectors (right panel, A3B-A3A-R1/-R2/-R1&2) and RNA secondary structures of the reporter substrates (left panel, *EVI2B* and *SDHB*, marked red as target C site). **(B)** Cell fluorescence images showing the subcellular localization of eGFP indicating RNA editing activity of the EVI2B and SDHB reporters and each editor's localization labeled with mCherry. Nuclear periphery is stained with Hoechst 33342 to distinguish nuclear and cytoplasmic regions (scale bar: 10 μm). **(C)** Quantification

of the subcellular localization of the eGFP reporter (Nuclear/Cytosolic ratio). Data are represented as mean values ± SD indicated by dots (n=30 cells). **(D)** Nucleotide sequencing distributions on the target C site on *EVI2B* and *SDHB* to measure C-to-U editing levels. **(E)** Quantification of subcellular localization of each mCherry-labeled editor (A3B-WT, A3B-A3A-R1/-R2/-R1&2). Data are represented as mean values ± SD indicated by dots (n=30 cells). Statistical significances in (C) and (E) were calculated by unpaired two-tailed student's t-test with P-values represented as: $P > 0.05$ = not significant; ns, * $0.01 < P < 0.05$, ** $P < 0.01$, *** $P < 0.001$, and **** $P < 0.0001$.

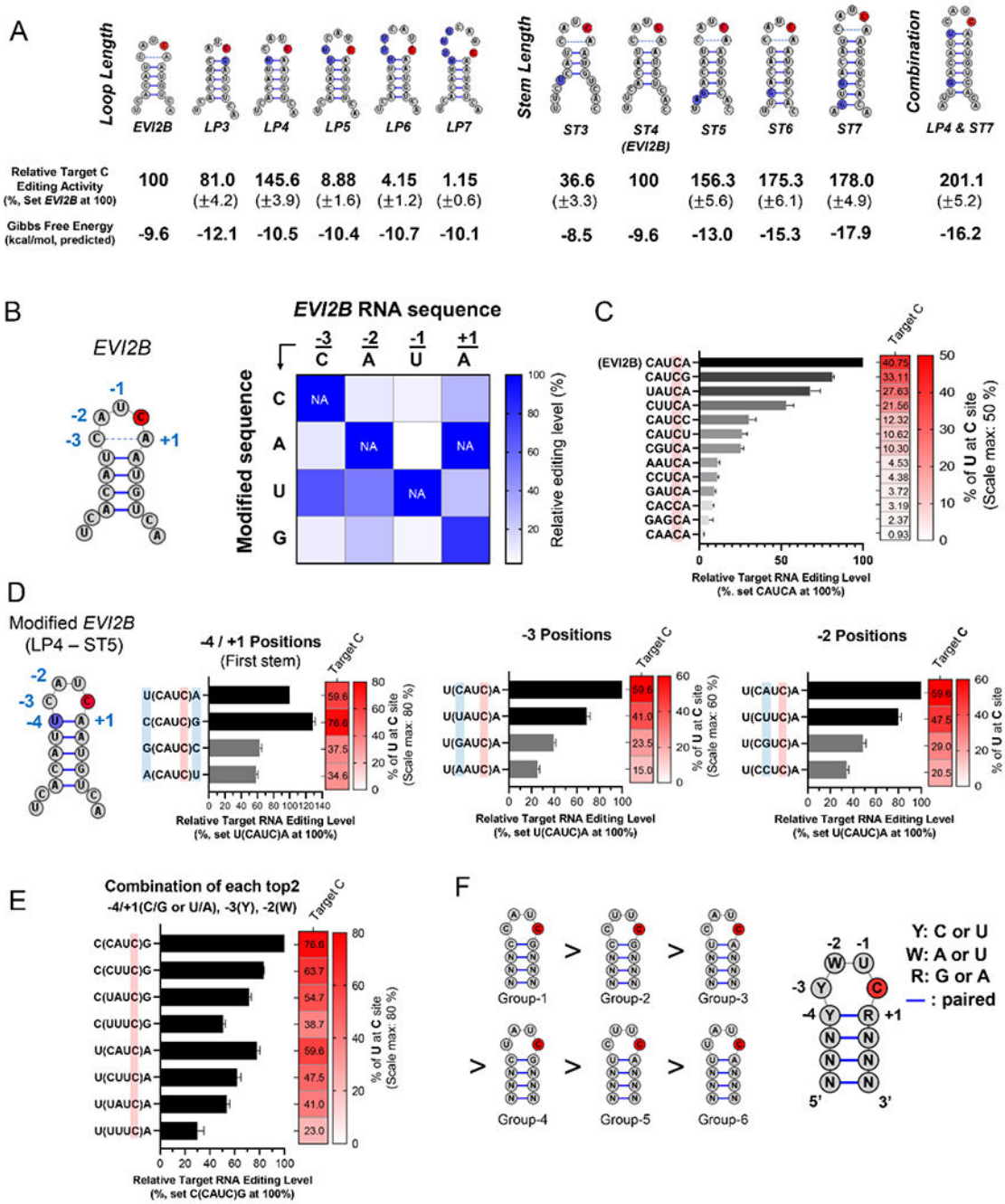


Figure 5. Structural and sequence features for optimal A3A-mediated RNA editing events.

(A) Relative A3A-mediated RNA editing activity with the C-to-U editing level of *EIV2B* set to 100% according to change in loop length and stem length, respectively. The predicted secondary structure is shown as schematic diagrams in the upper panel, with target C shown in red and the modified sequence in blue, respectively. The predicted Gibbs free energy was calculated when forming the secondary structure of the 48nt substrates and indicated in kcal/mol value. **(B)** Relative editing levels shown as heatmaps according to sequence modifications around target C (−4, −3, −2, −1, and +1) of EVI2B substrate. Predicted RNA

secondary structures are indicated on the left, lines indicate RNA double-helix and dotted line indicates possible C-A+ wobble base pairing. **(C)** Quantification of RNA editing level by A3A due to sequence variations surrounding target C region of *EVI2B*. The editing levels were represented by the ratio of U distribution at the C site quantified by Sanger sequencing results. **(D)** Measurement of RNA editing in response to sequence variations around target C in the modified tetraloop and stem 5 of *EVI2B* and **(E)** by combining the sequences of the top two preferred targets at each position. The relative editing efficiencies were represented in bar graphs, and the quantified editing levels of target C were indicated by heat maps. The sequence changes were represented by blue shading, and target C was depicted in red shading. **(F)** Groups and cartoon depictions of preferred sequences at each position surrounding target C.

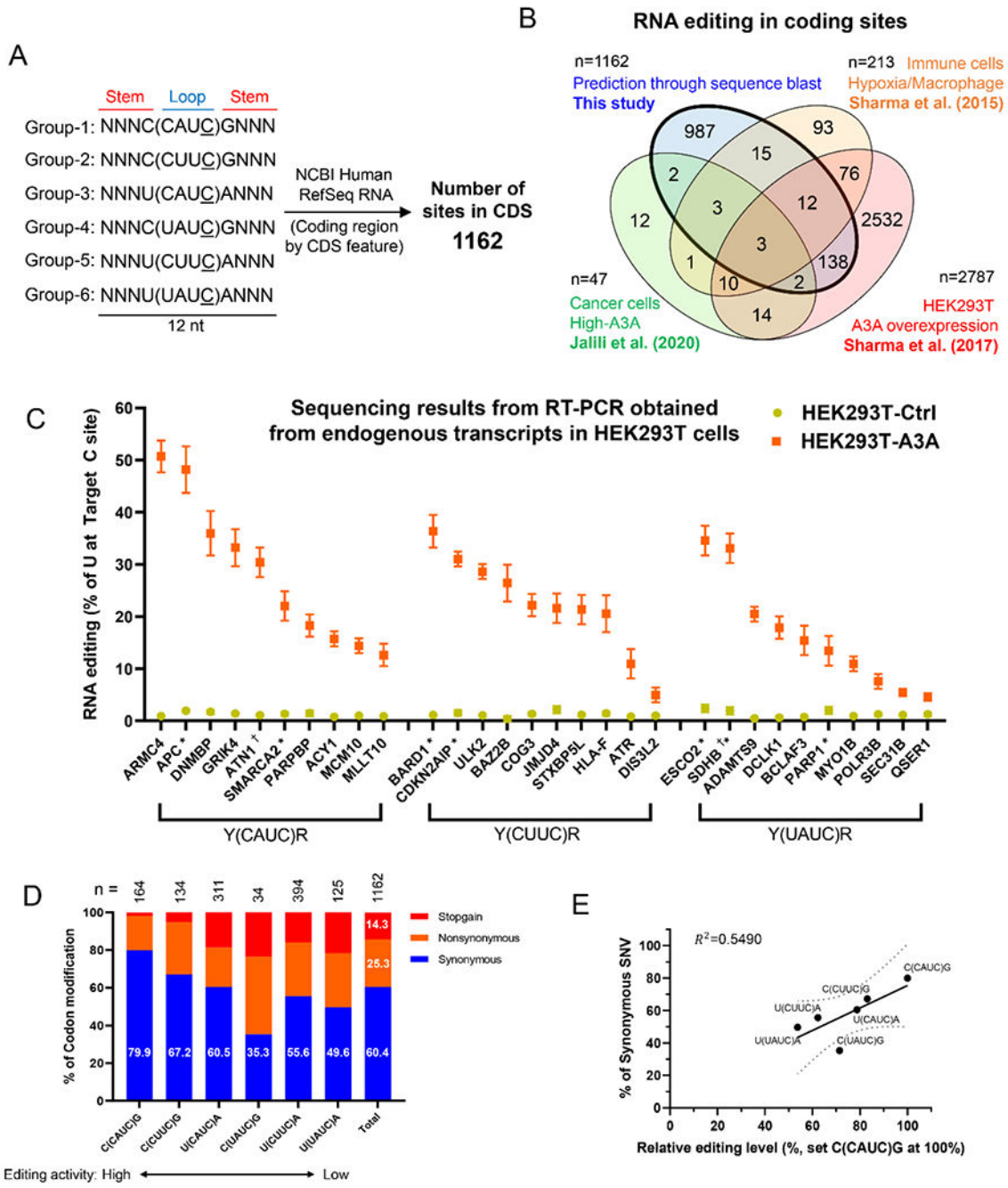


Figure 6. Prediction and validation of novel RNA editing sites on cellular mRNA.

(A) Identification of A3A-mediated RNA editing candidates in human transcripts corresponding to six optimized sequence groups. Inputting query sequences corresponding to each sequence group using the NCBI Blast tool to predict potential editing sites by A3A in coding regions. **(B)** Venn diagram showing the comparison between previous studies and the candidate group of this study [17, 18, 51]. The regions corresponding to each study are indicated in different colors. **(C)** Quantification of A3A-mediated RNA editing efficiency of 30 predicted sites in endogenous transcripts from HEK 293T cells. Overlap sites with

the previous studies indicated in gene names from Sharma *et al.* (2015) Nature comms.† [17] and Sharma *et al.* (2017) RNA Biology* [18]. **(D)** The rate of codon changes at sites corresponding to each sequence group by C-to-U editing. The number of sites retrieved in each group is indicated above the bar graph, with synonymous in blue, nonsynonymous in orange, and stopgain in red. **(E)** Correlation of the ratio of synonymous codon variants in each sequence group to the relative RNA editing level measured in Figure 5(E).