Title

RNA molecular recording with an engineered RNA deaminase

Journal

Nature Methods, 20(12)

ISSN

1548-7091

Authors

Lin, Yizhu
Kwok, Samentha
Hein, Abigail E
et al.

Publication Date

2023-12-01

DOI

10.1038/s41592-023-02046-z

Peer reviewed

HHS Public Access

Author manuscript

Nat Methods. Author manuscript; available in PMC 2024 October 23.

Published in final edited form as:

Nat Methods. 2023 December ; 20(12): 1887–1899. doi:10.1038/s41592-023-02046-z.

RNA molecular recording with an engineered RNA deaminase

Yizhu Lin[1], **Samentha Kwok**[1], **Abigail E. Hein**[2], **Bao Quoc Thai**[1,¶], **Yewande Alabi**[3], **Megan S. Ostrowski**[4], **Ke Wu**[4], **Stephen N. Floor**[1,5,†]

[1]Department of Cell and Tissue Biology, University of California, San Francisco, San Francisco, California, USA, 94143

[2]Tetrad Graduate Program, University of California, San Francisco, San Francisco, California, USA, 94143

[3]Biomedical Sciences Graduate Program, University of California, San Francisco, San Francisco, California, USA, 94143

[4]Gladstone Institute for Data Science and Biotechnology, San Francisco, California, USA, 94158

[5]Helen Diller Family Comprehensive Cancer Center, University of California, San Francisco, San Francisco, California, USA, 94143

Abstract

RNA deaminases are powerful tools for base editing and as RNA molecular recorders. However, the enzymes used in currently available RNA molecular recorders such as TRIBE or DART/STAMP have limitations due to RNA structure and sequence dependence. We designed a platform for directed evolution of RNA molecular recorders. We engineered an RNA A-to-I deaminse (rABE) that has high activity, low bias, and low background. Using rABE, we present REMORA (RNA Encoded Molecular Recording in Adenosines), wherein deamination by rABE writes a molecular record of RNA-protein interactions. By combining rABE with the C-to-U deaminase APOBEC1 and long-read RNA sequencing, we measured binding by two RNA binding proteins on single mRNAs. Orthogonal RNA molecular recording of PUM1 and PUM2 reveals that PUM1 competes with PUM2 for a subset of sites in cells. We further identify transcript-isoform specific RNA–protein interactions driven by isoform changes distal to the binding site. The genetically encodable RNA deaminase rABE enables single-molecule identification of RNA-protein interactions with cell type specificity.

[†]Correspondence: stephen@floorlab.org.
[¶]Current Address: MSTP Program, University of Arizona, Tuscon, Arizona, USA, 85724

Author Contributions:
Conceptualization, Y.L. and S.N.F.; Investigation, Y.L., S.K. B.Q.T., S.N.F., and Y.A.; Writing – Original Draft, Y.L. and S.N.F.; Writing – Review & Editing, Y.L. and S.N.F.; Methodology, Y.L. and S.N.F.; Resources, M.S.O. and K.W.; Formal analysis, Y.L.; Software, Y.L. and A.E.H.; Funding Acquisition, S.N.F.; Supervision, S.N.F.

Code availability
Analysis code is available at https://github.com/linyz/remora.

Competing Interests:
None.

Keywords

directed evolution; RNA deaminase; RNA-binding proteins; TRIBE; STAMP; DART

Introduction

Regulation of gene expression is critical for proper cellular function. RNA-binding proteins (RBP) play important roles in the regulation of gene expression[1]. Measuring the sites where RBPs bind RNAs is necessary to understand how they regulate RNA biology. Since genes often generate multiple RNA products, it is also important to measure RBP sites with transcript isoform and single-molecule resolution.

Several high-throughput methods exist to identify RBP binding sites. RNA bind-n-seq, RNA-MaP, and related methods enable identification of RBP sequence preferences using purified proteins[2,3]. In cells, RNA immunoprecipitation (RIP-seq) and crosslinking based methods such as CLIP[4], eCLIP[5], and PAR-CLIP[6] identify RBP binding sites transcriptome-wide. These methods have been developed for short-read RNA sequencing and require RNase digestion of RNA molecules. However, RNase digestion limits resulting data to short fragments of RNA. As human RNA molecules are heterogeneous and vary widely in translational potential[7,8], RBP binding information in short reads averaged across RNA molecules can be challenging to interpret. Moreover, it is only possible to measure binding of one RBP at a time with these approaches.

To distinguish RNA isoforms and measure RBP–RNA interactions at the single molecule level, recent RNA molecular recording methods including TRIBE[9] and DART-seq[10] or STAMP-seq[11] have been developed. These methods use RNA deaminase enzymes as RNA molecular recorders to mark RBP binding sites. TRIBE relies on the adenosine deaminase *Drosophila* ADAR[9], while DART/STAMP-seq use the C-to-U deaminase APOBEC1[10,11]. However, both ADAR and APOBEC1 have limitations as RNA molecular recorders. dADAR is a double-stranded RNA deaminase that has reduced efficiency on single-stranded RNAs[9,12]. APOBEC1 has endogenous targets[13], and thus may introduce background editing. These limitations motivate development of a synthetic RNA molecular recorder.

Pumilio family RBPs are important regulators of post-transcriptional gene expression. Humans express Pumilio paralogs, encoded by *PUM1* and *PUM2*. PUM1 and PUM2 proteins primarily interact with mRNA 3′ untranslated regions (UTRs) to regulate translation, RNA stability, and RNA localization[14]. PUM1 and PUM2 proteins are 90% identical in the Pumilio domain and have nearly identical binding affinity landscapes *in vitro*[15,16], but bind overlapping and distinct sites in cells[17]. Many RNA-binding proteins interact with a subset of transcript isoforms expressed from a gene, but it has been difficult to measure which transcript isoform Pumilio proteins bind to using short-read sequencing approaches such as CLIP. Moreover, it is unclear whether the overlapping and distinct binding selectivity of PUM1 and PUM2 proteins manifests at the single molecule level.

Here, we use RNA-based directed evolution to develop an RNA adenosine deaminase with high activity, low background, and low sequence or RNA structure preference –

which we term the RNA Adenosine Base Editor (rABE; Figure 1A). Using a stringent computational pipeline to identify editing sites in short- and long-read sequencing data, we validate that rABE recapitulates known binding of multiple RNA-binding proteins. To facilitate description of the approach, we name this strategy REMORA: RNA Editing Molecular Recording in Adenosines. REMORA is a new approach within the family of RNA molecular recorders, such as TRIBE[9], DART[10], or STAMP[11]. When combined with long-read sequencing rABE identified isoform specific binding events. We then use rABE and APOBEC1 to measure co-occupancy at the single-molecule level between Pumilio family RNA-binding proteins that recognize the same RNA sequence. We find that the Pumilio protein with higher expression dominates binding at the population level, but single-molecule data reveals a positive correlation between PUM1 binding and PUM2 binding. REMORA thus enables single-molecule detection of RNA-binding proteins in cells through genetically encodable RNA adenosine deamination. Our evolved rABE outperforms other RNA molecular recorders due to low bias and background editing. Moreover, our directed protein evolution platform can be broadly adapted to engineer for other RNA editing enzymes, accelerating the development of RNA editing toolkits for both research and therapeutic purposes.

Results

Identifying a single-stranded RNA adenosine deaminase

We surveyed RNA adenosine deaminase families for enzymes that might function on single-stranded RNA (ssRNA). We selected variants the human tRNA deaminase *ADAT2–ADAT3*, the *S. cerevisiae* tRNA deaminase TAD2–TAD3, the prokaryotic tRNA deaminase *S. aureus* tadA, the ADAR2-based engineered RNA deaminase hyperTRIBE[18], and an engineered single-stranded DNA A-to-I deaminase ABE7.10[19], which was developed from *E. coli tadA* for base editing as a dCas9 fusion. The selected tRNA deaminases target a loop region of tRNA in a single stranded conformation, while ADAR2 targets double-stranded RNA (dsRNA)[20].

To evaluate the selected RNA deaminases, we designed an A-to-I editing reporter as shown in Figure 1B, inspired by prior work[21]. Deaminases were fused to a lambdaN peptide that interacts with the BoxB RNA element to recruit the deaminase to an unstructured UAG stop codon between eGFP and mCherry (Extended Figure 1A). When deaminated, UAG codon is modified to UIG, which is decoded as UGG (tryptophan), allowing mCherry expression. We transfected six deaminases into HEK 293T cells with the eGFP-UAG-BoxB-mCherry reporter. A-to-I editing was measured using flow cytometry by the percentage of mCherry+ in the GFP+ population. As shown in Figure 1C,1D and 1E, all six deaminases tested showed more mCherry-positive cells compared to the empty vector control (25.1% ± 1.1%, 95% confidence interval of mean). ABE7.10 showed the highest mCherry-positive percentage (71.6% ± 3.4%), significantly higher than hADAR2 (45.1% ± 1.7%), hyperTRIBE (42.2% ± 7.3 %) or *S. cerevisiae* TAD2–TAD3 (46.8% ± 4.8%). While the percentage of mCherry cells in GFP+ population is a sensitive measurement to identify deaminases with minimal A-to-I editing activity, mCherry/GFP ratio in GFP + population is a more direct measurement of the A-to-I editing efficiency. Similarly, ABE7.10 showed the

highest mCherry/GFP ratio (1.54 ± 0.14; Figure 1E). The very low activity of hyperTRIBE for the ssRNA site in our reporter is consistent with bias against single-stranded RNA[12]. The RNA-targeting activity of ABE7.10 is consistent with reported data ABE7.10[22–24], and indicates it can be directed to RNA sites. We then measured editing by ABE7.10 using Sanger sequencing (Figure 1F) and RNA-seq (Figure 1G). ABE7.10 had a maximum efficiency by RNA-seq of 25.8% ± 7.5% at 11 nucleotides upstream of the BoxB structural element. We concluded ABE7.10 is a good candidate ssRNA A-to-I editor.

Directed evolution of an improved RNA molecular recorder

Since ABE7.10 was selected for ssDNA activity but started as a tRNA deaminase, we reasoned that mutations might increase its efficiency on ssRNA. We designed a directed protein evolution system to select for improved ssRNA variants of ABE7.10. A UAG-BoxB cassette was inserted at the 5′ end of the kanamycin resistance gene *kanR* (Figure 2A). Similar to the eGFP-UAG-BoxB-mCherry reporter, deamination at the UAG stop codon will express *kanR*. Since RNA editing is transient, this selection proved sensitive to the initial starting optical density and kanamycin concentration. To determine the optimal kanamycin concentration, we incubated *E.coli* expressing ABE7.10 and the inactive mutant ABE7.10-E59A with different concentrations of kanamycin in deep 96-well plates for 48 hours. As shown in Extended Figure 2A, 18.75 μg/ml kanamycin (37.5% of the normal working concentration) allowed for growth of *E. coli* expressing ABE7.10 but not the inactive mutant ABE7.10-E59A.

To search for ABE7.10 variants that might improve deamination efficiency, we randomly mutated the ABE7.10 coding region with error prone PCR. We cloned this error-prone PCR library into pET15b with the lac promoter driving both the BoxB-UAG-*kanR* selection cassette and ABE7.10 variants, electroporated it into *E.coli*, and incubated transformants in 18.75 μg ml^{-1} kanamycin-containing LB medium for up to 48 hours (Figure 2B). After 1 cycle of selection, ABE7.10 was amplicon sequenced to identify enriched mutations (Supplementary Table 1). We cloned individual enriched mutations and tested their efficiency in human cells using the mGFP-UAG-BoxB-mCherry reporter. Variants with higher efficiency than ABE7.10 were pooled for a second round of error prone PCR and selection, which did not identify additional variants. The mean of ratios of mCherry to GFP fluorescence was calculated as before, which showed good sensitivity and reproducibility when measuring A-to-I editing efficiency (Figure 2C and Extended Figure 2B). We found that the C146S mutation increased the A-to-I editing efficiency of ABE7.10 on ssRNA by 1.64 (±0.07) fold. Combining multiple mutations that increased editing efficiency (rABE1.1 and rABE1.2) did not outperform ABE7.10^{C146S}.

Interestingly, C146S reverts a cysteine back to serine as in *E. coli* tadA. This mutation modestly increased expression (Extended Figure 2C). However, a codon optimized version of ABE7.10 (ABEmax[25]) had higher expression level than ABE7.10^{C146S}, but lower editing efficiency (Figure 2D), indicating the effect of C146S is not strictly due to expression. Structural modeling using *S. aureus* tadA (PDB: 2B3J)[26] shows that Thr145 (corresponding to *E. coli* Ser146) is about 4.5 angstroms from the phosphate backbone of RNA, suggesting that substitutions at this position may influence RNA–protein interactions (Extended Figure

2D). We also observed that *S. aureus* Thr145 is in the vicinity of U33, the −1 nucleotide prior to the adenosine to be deaminated. To explore the possibility that C146S influences sequence preferences around deamination sites, we analyzed Sanger sequencing traces on a version of the BoxB reporter with adenosines introduced adjacent to the stop codon. C146S increased deamination at adenosines with a −1 uracil and in other sequence contexts (such as CAG or GAA), suggesting that C146S does not induce an absolute dependence for a −1 uracil (Extended Figure 2E).

The deaminase tadA naturally functions as a homodimer[26]. To mimic this, ABE7.10 was designed as a constitutive dimer[19]. However, recent work found that a single tadA domain is sufficient[27]. We tested whether mono-ABE7.10^{C146S}, with one tadA monomer, can achieve similar efficiency to dimeric tadA-ABE7.10^{C146S}. We found that mono-ABE7.10^{C146S} has 1.72 (±0.09, CI 95%) fold higher A-to-I editing efficiency on ssRNA than the tadA-ABE7.10^{C146S} dimer and 2.47 (±0.13) fold of the original ABE7.10 (Figure 2D). We named mono-ABE7.10^{C146S} the RNA Adenosine Base Editor (rABE).

rABE has low background and minimal sequence dependence

To assess background and the impact of rABE on gene expression, we generated HEK 293T cell lines expressing rABE under a doxycycline inducible promoter. Cells were treated with 0, 50, or 1000 ng ml^{-1} doxycycline for 48 hours before performing RNA-seq (Figure 3A). We selected 48 hours as it yielded the highest signal in our colorimetric reporter (Figure 1). Expression of rABE alone had no effect on gene expression (Extended Figure 3I; odds ratio threshold = 2, $p < 0.05$, BH adjusted) and we observed no cell growth difference between rABE expressing cells and wild type HEK 293T cells under doxycycline treatment (Extended Figure 3A, 3B).

We developed a stringent computational pipeline to call rABE editing sites in RNA-seq data based on significant editing over background in all three replicates. Briefly, we filtered for coverage (average count per sample ≥ 20) and removed SNPs or naturally-occurring A-to-I sites (allele count ≥ 2 in controls). We then performed a Cochran–Mantel–Haenszel (CMH) test on the total read count and allele count at each nucleotide position. For every significant editing event ($p < 0.05$, BH adjusted), the editing strength was calculated as the average of the difference in editing rate between treated (dox+) and control (dox-) samples.

To benchmark rABE, we fused rABE to the C-terminus of Rbfox2, a well-studied RNA–binding protein involved in splicing regulation[28–30]. We generated stable HEK 293T cell lines with tetracycline-inducible expression of Rbfox2-rABE. Significant A-to-I editing events were identified as described above, by comparing Rbfox2-rABE and rABE-Flag expressing cells under the same doxycycline concentration and incubation time. The number of significant A-to-I editing events increased with doxycycline concentration and duration (Figure 3B and Extended Figure 3C–E), supporting the selection of 48 hours as a timepoint to achieve maximum coverage. Following high doxycycline induction, RNA extraction and RNA-seq, we identified 4662 A-to-I editing events over background introduced by Rbfox2-rABE across 487 genes (Figure 3C–E).

Previous work used hyperTRIBE[31], a double-stranded A-to-I deaminase and APOBEC1, a C-to-U RNA deaminase, for RNA molecular recording[10,11,32]. To compare the performance of rABE, hyperTRIBE and APOBEC1, we also expressed Rbfox2-hyperTRIBE and Rbfox2-APOBEC1 fusions in the same experimental format and performed RNA-seq. We used codon-optimized full-length APOBEC1 as in BE4max[25]. Rbfox2-hyperTRIBE identified only 338 significant A-to-I editing events, and Rbfox2-APOBEC1 identified 1587 significant C-to-U editing events under high doxycycline condition, about one-third of Rbfox2-rABE (Figure 3D, 3E and Supplementary Table 2), despite similar expression levels (Figure 3F–H and Extended Figure 3F–H). In addition, hyperTRIBE alone or Rbfox2-hyperTRIBE induced modestly more gene expression changes than rABE or APOBEC1 (Extended Figure 3I and J). Motif analysis within 100 nt regions centered around Rbfox2-rABE A-to-I editing sites or Rbfox2-APOBEC1 C-to-U editing sites recovered the known GCAUG motif of Rbfox2, while Rbfox2-hyperTRIBE did not (Figure 3K). We performed CentriMo[33] analysis on Rbfox2-rABE and Rbfox2-APOBEC1. For Rbfox2-rABE, the GCAUG motif occurs at the highest frequency at 10 nucleotides downstream of A-to-I editing sites, but also has high frequency about 20 nt upstream of A-to-I editing sites (Figure 3L), possibly suggesting Rbfox2-rABE tends to introduce A-to-I editing on both sides of Rbfox2 binding sites. Similar patterns were observed for Rbfox2-APOBEC1 (Extended Figure 3K).

Consistent with previous studies[12], hyperTRIBE biased towards double-stranded RNA regions, resulting in fewer editing events and failure to identify the Rbfox2 motif. As shown in Figure 3M, for both hyperTRIBE and Rbfox2-hyperTRIBE, the predicted RNA minimum free energy (MFE) around editing sites are lower than that of rABE and APOBEC1. In contrast, MFE around Rbfox2-rABE is higher than rABE alone ($p < 9.21e\text{-}218$, t test), suggesting rABE editing has less dependence on RNA secondary structure. *In silico* RNA structure prediction is consistent with published human transcriptome-wide icSHAPE data[34] (Figure 3N and Extended Figure 3L), regions around Rbfox2-hyperTRIBE editing sites showed lower SHAPE reactivity scores than that of Rbfox2-rABE or Rbfox2-APOBEC1, indicating hyperTRIBE's preference for double-stranded RNA. All three enzymes had increased SHAPE reactivity score at their immediate editing sites, suggesting some degree of accessibility is required for efficient RNA editing.

To explore the difference in the number of identified Rbfox2 sites between rABE and APOBEC1 fusions, we measured background editing by expressing rABE or APOBEC1 alone. We identified 3085 significant background C-to-U editing events introduced by APOBEC1 under high doxycycline conditions, 17 fold higher than rABE (182 A-to-I editing events; Figure 3C and 3E), despite comparable expression levels (Figure 3H and Extended Figure 3H). APOBEC1 induced background editing were enriched in 3′ UTR (Extended Figure 3M), consistent with its endogenous targets[13]. We compared the local sequence preference around editing sites for Rbfox2-rABE and Rbox2-APOBEC1 and find that rABE showed lower sequence bias around editing sites than APOBEC1 (Figure 3I, 3J, Extended Figure 3N and Supplementary Table 3), which prefers −1 A (69.6% frequency) and +1 U (54.3%). The sequence preference for APOBEC1 is consistent with endogenous targets[13]. Notably, G and C are almost absent from the −1 position of APOBEC1 C-to-U editing sites, with frequency of 1.1% and 0.8% respectively, suggesting APOBEC1 has stronger

local sequence dependence than rABE (13% A, 9.7% C, 13.7% G and 63.3% U at −1 position). RNAfold[35]-predicted structure around APOBEC1 editing sites also had higher MFE (less structured), which is consistent with icSHAPE scores (Figure 3N), suggesting some preference of APOBEC1 for unstructured RNA regions.

We identified known Rbfox2 targets in RNA editing events introduced by Rbfox2-rABE or Rbfox2-APOBEC1, as exemplified by the 3′ UTR of *APP*[11,36] (Figure 3O). Editing signals from Rbfox2-rABE and Rbfox2-APOBEC1 were proximal. Transcriptome-wide, we found that 83.6% of Rbfox2-rABE and a similar percentage of Rbfox2-APOBEC1 editing sites were in 3′ UTRs (Extended Figure 3O). This differs from prior work that used CLIP to identify Rbfox2 binding sites and found the majority of Rbfox2 sites are intronic[30,37]. One possibility for this difference is that REMORA and related approaches measure binding in the whole transcriptome, where intron-containing messages are comparatively rare, versus pull-down enrichment for RBP-associated RNAs in CLIP.

rABE recovers known binding patterns of Pumilio proteins

To test the generality of rABE, we fused rABE to the C-terminus of PUM1 and PUM2, which bind to Pumilio binding elements (PBEs; UGUANA)[38,39]. We generated stable HEK 293T cell lines expressing tetracycline inducible PUM1-rABE and PUM2-rABE, and treated them with 0, 50, or 1000 ng ml^{-1} doxycycline for 48 hours before RNA-seq (Figure 4A).

We found that PUM1-rABE introduced A-to-I edit sites above background at PUM1 targets in cells, as exemplified by the 3′ UTR of *PCNA*[40] (Figure 4B). PUM1-rABE A-to-I editing events were adjacent to PBE motifs and previously identified CLIP-seq peaks[17]. Transcriptome-wide, we identified 10793 significant A-to-I editing events in 1420 genes under high doxycycline condition in PUM1-rABE expressing cells (Figure 4C). 879 out of 1420 (61.9%) of PUM1 targeted genes were also identified in previous work using CLIP-seq[17] (Extended Figure 4A and Supplementary Table 4). Consistent with previous PUM1 footprinting with CLIP-seq, the majority of PUM1 binding events were in 3′ UTRs (9445/10793, 87.5%) (Figure 4D). We performed de novo motif identification within ±50 nt around A-to-I editing sites and found that the consensus PBE motif UGUANA was enriched (Figure 4E). For the 541 sites identified with PUM1-rABE but not in CLIP-seq, de novo motif discovery also recovered the UGUANA PBE motif, as well as a UGUAUUA near-cognate motif[15] (Extended Figure 4D), suggesting these sites may be false negatives in CLIP-seq. PUM1 targeted genes identified by CLIP-seq but not PUM1-rABE had lower expression level (p-value < 2.2 e-16, t-test, Extended Figure 4E). CentriMo[33] analysis showed that UGUANA motifs were 10–15 nt downstream of A-to-I sites (Figure 4F), suggesting PUM1-rABE introduces A-to-I edits 10–15 nt upstream of a binding site, consistent with rABE-lambdaN (Figure 1).

The number of PUM1-rABE A-to-I editing events identified over background increased with PUM1-rABE expression level. Doxycycline treatment (1000 ng ml^{-1}) increased expression of PUM1-rABE at both the RNA (Figure 4G) and protein level (Figure 4H and Extended Figure 4F). Under high doxycycline condition, PUM1-rABE protein level was increased by about 2.5 (Figure 4H and Extended Figure 4F), increasing signal but not to a degree likely to induce artifactual binding[15]. We identified 4362 significant A-to-I editing events in 799

genes under low doxycycline treatment, 40.4% of the high doxycycline condition (Figure 4I and Extended Figure 4H), and the average editing rate at each site was slightly lower (Figure 4J). This suggests that rABE deamination frequency is a measure of concentration dependent RBP binding in cells.

We fused rABE to the C-terminus of PUM2, another Pumilio RBP. We recovered the known PBE (Figure 4E), and observed that PUM2 sites have a moderate preference for cytidine at the 5th position (Figure 4E). The PUM2-rABE fusion identified only 419 significant A-to-I editing events even under high dox condition (Figure 4C). Consistent with previous work[15,17], the large majority of PUM2 targets were also PUM1 targets (134 out of 138 genes) (Extended Figure 4B). In the example of the *PCNA* 3′ UTR (Figure 4B), we also observed PUM2-rABE editing events near PBEs. PUM2-rABE identified fewer A-to-I events than PUM1-rABE, and fewer PUM2-rABE targeted genes than identified previously with CLIP-seq[17] (Extended Figure 4C). The lower amount of PUM2 binding events may be due to lower PUM2-rABE expression level than PUM1-rABE (Figure 4G), even under the same doxycycline condition. Notably, endogenous PUM2 also has lower expression level than PUM1 (Figure 4G). Together, we conclude that the REMORA approach is a faithful RNA molecular recorder that can be used to identify RNA-binding protein sites through introduced adenosine deamination.

To understand how protein binding affinity and RNA sequence features contribute to RBP-rABE induced editing rate, we built a model with xgboost[41], a machine learning algorithm based on gradient boosted trees, to predict PUM1-rABE editing rate. Previous work found that the ΔΔG of motifs from the optimal motif binding free energy ΔG was the determining factor of *in vitro* Pumillio protein binding affinity[15]. Other features included in our xgboost model, were sequencing read depth, −1 nt nucleotide from the editing site, nucleotide composition around PBE motif and editing site, the distance between editing site and the nearest PBE motif, RNA secondary structure around PBE (ss_ΔΔG)[15], and number of PBE motifs +-50 nt around editing site. Our xgboost model explained a moderate amount of variability in PUM1-rABE A-to-I editing rate (Figure 4K and L, R2=0.55 in training set and 0.31 in test set), and the most important predictor is sequencing read depth (Figure 4M), likely due to higher statistical power for significant edit calling in regions with higher read depth. The second and third predictors are ΔΔG (*in vitro* binding affinity) and −1 nucleotide, consistent with previous study[15] and our observation in Figure 3K.

Long read sequencing identifies isoform specific PUM1 binding

One of the advantages of REMORA is that it does not require RNA fragmentation. Thus, REMORA is compatible with long-read sequencing to identify RBP binding events in full-length RNA at the single molecule level, making it possible to measure heterogeneity in RBP-RNA interactions and isoform specific RBP binding events. To identify isoform specific PUM1 binding events, we performed PacBio long-read sequencing (Figure 5A) on the same RNA samples from doxycycline inducible PUM1-rABE HEK293T cells (Figure 4A). PUM1-rABE events were identified by filtering background editing in rABE-Flag controls. Isoform specific binding events were identified as follows: for two non-overlapping exons A and B in a gene, we counted the number of A-to-I mutations in exon A in all

reads containing exonA, and group reads by whether they contain exon B. Enrichment of PUM1 binding events on exon A was calculated as odds-ratio between groups, significance was calculated with Mann–Whitney U test, and p-values were adjusted with Bonferroni method. Transcriptome-wide, we identified 85 isoform specific PUM1 binding events on 16 genes (Figure 5B, Supplementary Table 5). This observation is a lower bound on the possible isoform-specific PUM1 binding events, since our long-read sequencing is lower depth compared to our short-read sequencing.

Figure 5C and 5D shows an example of isoform specific PUM1 binding on *DDX5*. The canonical isoform of DDX5 (DDX5-ca) has 13 exons, and the majority of PUM1-rABE signal was located in its 3′ UTR in exon 13. Of 509 Pacbio reads that aligned to *DDX5*, 61 exhibit intron retention between exon 11 and 12 (DDX5-IR). We observed 7-fold more A-to-I editing on exon 13 in DDX5-ca than DDX5-IR (Figure 5C, p-adj = 7.38e-14). Figure 5E shows another example in *KHDC4*, a pre-mRNA splicing factor. The canonical isoform of KHDC4 has 14 exons, and the majority of PUM1 signal was in its 3′ UTR in exon 14. We observed alternative splicing events on exon 10 (101 nt difference) and exon 12 (28 nt difference), shorter versions of either or both exons led to a frameshift and premature stop. The canonical isoform of KHDC4, which had longer exon 12 and longer exon 10, was associated with significantly higher PUM1 signal (Figure 5F). These examples suggests that proper processing of some mRNA may be required for their interaction with PUM1, and demonstrate the ability of REMORA to identify RNA–protein interactions unique to transcript isoforms due to distal RNA processing events even when the binding site is shared.

Dual labeling for simultaneous RBP molecular recording

Since rABE and APOBEC1 can be used for RBP molecular recording and introduce distinguishable mutations, we tested whether rABE and APOBEC1 can achieve simultaneous RBP molecular recording. We generated cell lines expressing dox inducible PUM1-rABE, PUM1-APOBEC1, PUM2-rABE, or PUM2-APOBEC1, and a cell line that expresses PUM1-rABE and PUM2-APOBEC1. Consistent with Rbfox2-APOBEC1 results above, we identified more significant editing events with PUM1-rABE (10793 sites) and PUM2-rABE (419 sites) (Figure 4C) compared to PUM1-APOBEC1 (6555 sites) and PUM2-APOBEC1 (84 sites) (Figure 6A) due to lower background editing from rABE. The choice of deaminase for RNA molecular recording had a larger impact on PUM2 than on PUM1 (60.1% as many sites identified in PUM1 versus 20.0% in PUM2). Thus, we used PUM1-APOBEC1 and PUM2-rABE for the following experiments.

At least three possibilities could occur between PUM1 and PUM2: independent, cooperative, or competitive binding. We first checked whether dual expression of both PUM1-APOBEC1 and PUM2-rABE changed binding patterns compared to individual expression. Figure 6B shows an example of PUM1-APOBEC1 and PUM2-rABE editing patterns on *RPL7A*, where we observed similar editing patterns in dual and single labeling. In another example, *PCNA*, we observed decreased PUM2-rABE editing signal when both PUM1 and PUM2 were overexpressed (Extended Data Figure 5). In contrast, we observed an increase of both PUM1 and PUM2 binding signal in the dual labeling sample

for the gene *UBB* (Extended Data Figure 5). Transcriptome-wide, as shown in Figure 6C, editing rates from dual and single labeling samples were highly correlated for PUM1-APOBEC1 (Pearson's r = 0.82) and moderately for PUM2-rABE (Pearson's r = 0.67); only a few genes showed differential signal. In cells expressing PUM1-APOBEC1 and PUM2-rABE (dual labeling cells), we observed slightly higher PUM1-APOBEC1 signal than PUM1-APOBEC1 alone, which may be due to slightly higher expression level of PUM1-APOBEC1. Intriguingly, in dual labeling cells, although the overall signal for PUM2-rABE was lower than PUM2-rABE only cells (Figure 6C), we found more genes increased rather than decreased in PUM2 binding (Figure 6D). These results suggesting when both PUM1 and PUM2 were expressed, PUM1 out-competed PUM2 due to its higher expression, however a subset of genes had increased PUM2 binding. Further analysis showed that genes with increased number of PUM2-rABE induced editing events had zero or low number of the canonical PBE binding motif (Figure 6E), suggesting PUM2 binding increase may be caused by a noncanonical mechanism – such as deamination of RNA residues that are nearby in three-dimensional space.

To evaluate whether a single mRNA can be bound by PUM1 and PUM2, we also performed PacBio long-read sequencing. Indeed, we observed that a single RNA molecule can have both A-to-I editing events and C-to-U events, suggesting binding by both PUM1 and PUM2 (Figure 6F). To evaluate PUM1 and PUM2 binding at the single molecule level, we correlated the A-to-I and C-to-U editing on each read of a single gene. Many genes had positive correlations (Figure 6G), indicating a subset of transcripts were bound by both PUM1 and PUM2, while others were bound by neither. For comparison, we generated a background distribution of correlation scores by shuffling of PUM1-APOBEC1 and PUM2-rABE single labelled cells, which were centered around zero. One explanation for this positive correlation between PUM1 and PUM2 binding is that certain conditions, such as mRNA processing, mRNA localization, or interactions with cofactors, are prerequisites for transcripts to be bound by Pumilio proteins. We performed GSEA analysis with gene list ranked by their PUM1-PUM2 correlation score, and found that the positively correlated genes are enriched in certain miRNA targets (Figure 6H). This resonates with previous findings that Pumilio proteins are involved in microRNA regulation[17,42]. Taken together, rABE and APOBEC1 can be used for robust orthogonal footprinting of two RNA binding proteins at the single molecule level.

Discussion

The systematic understanding of gene regulation has been advanced by high-throughput sequencing approaches. For RNA-binding proteins, diverse approaches for capturing RNA-binding sites have been developed for second-generation, short-read sequencing[2–6]. In recent years, new methods are being developed to enable measurement of RNA-binding proteins using third-generation, long-read sequencing, which preserves RNA transcript isoform information and can measure binding at the single-molecule level[43]. In this work, we developed a new adenosine deaminase for RNA molecular recording, an approach we term REMORA (Figure 6I).

We find that the highly similar Pumilio proteins PUM1 and PUM2 only sometimes compete for the same binding sites in cells. Jarmoskaite et al.[15] found that PUM1 and PUM2 have indistinguishable sequence preferences *in vitro* and found excellent agreement between their experimentally-derived thermodynamic model and eCLIP PUM1/PUM2 binding. Our finding that PUM1 and PUM2 compete is consistent with the conclusion that human PUM1 and PUM2 levels are subsaturating compared to binding sites, and therefore occupancy is highly sensitive to expression changes[15]. Conversely, our data showed a subset of genes with few PBE motif had increased PUM2 signal, hinting at complexity in cellular binding. Understanding factors that drive competition (or cooperation) between RBPs in cells will be the basis for exciting future work.

We demonstrate that combining REMORA with long-read sequencing identifies RBP-RNA interactions at the single molecule level. With PUM1-rABE and long-read sequencing, we identified isoform specific PUM1 binding patterns. Distal mRNA processing events, such as alternative splicing and intron retention, contribute to PUM1 binding regulation without changing the local context around PBE motifs. Using both rABE and APOBEC1, we labelled PUM1 and PUM2 simultaneously, allowing investigation of the binding patterns of two RBPs on a single RNA molecule. REMORA with long-read sequencing opens exciting new opportunities for understanding the heterogeneity in RBP: RNA interactions.

ABE7.10 was evolved from *E. coli* tadA[19]. Prokaryotic tadA introduces an inosine into the wobble base of $tRNA^{Arg\text{-}ACG,44}$. We explored related tRNA deaminases, including human ADAT2 and ADAT3 that install wobble-base inosines in eight tRNAs[45]. However, we only detected mild activity with human ADAT2 and ADAT3 (Figure 1). It will be an interesting direction for the future to explore development of human ADAT2-ADAT3 for RNA molecular recording. Moreover, therapeutic introduction of RNA deaminases for programmable deamination might be facilitated using human ADAT2-ADAT3.

Our UAG-BoxB based mRNA reporter can be broadly adapted to other RNA editing enzymes, including A-to-I editor ADAT2-ADAT3 or ADAR and U-to-C editor PPR-DYW proteins[46]. By replacing antibiotic resistant selection with toxic protein selection such as CcdB[47], our system could also be used for directed evolution of C-to-U RNA editors. In future work, exploring other mutagenesis methods such as sequence shuffling between closely related RNA editors, or shuffling between different RNA modification enzymes will further expand our RNA editing toolkits for both research and therapeutic purposes.

Interestingly, the offset in linear sequence space between rABE-induced adenosine deamination and RBP motifs differs in different constructs. In the lambda-BoxB system, dimeric ABE7.10 deaminated adenosines roughly 11 nt upstream from the BoxB hairpin. When fused to PUM1, rABE also introduced edits about 10–15 nt upstream from PUM1 motifs. In contrast, Rbfox2-rABE deaminated adenosines upstream and downstream of Rbfox2 motifs. An intriguing possibility to investigate in the future is that the structure or dynamics of the RBP that rABE is fused to imparts different distance dependences on deamination, or that RNAs or RBPs might generate conformation-dependent distance dependencies.

Expression of rABE for extended periods of time is a limitation with the present implementation of REMORA. Long duration expression of active deaminases integrates binding events across the experimental timecourse. As such, deamination events in this work should be interpreted as binding that was recorded at some point during the experiment (typically 48 hours post dox induction). Design of acutely inducible deaminase systems could enable RNA molecular recording with finer time resolution[48]. Alternatively, studying biological processes that can be frozen in time through chemical inhibition or other means could enable higher molecular resolution.

REMORA and related RNA molecular recording strategies do not enrich for bound sites as CLIP-seq or RIP-seq do, which is both a strength and a limitation. Enrichment of bound targets via purification of a protein of interest generates a more comprehensive binding profile. However, this enrichment process also has the potential to lose occupancy information, since purification will enrich for all targets including low abundance sites. Therefore, CLIP and related approaches may be useful to generate comprehensive binding maps, while RNA molecular recording approaches like REMORA may be useful for measuring near-native occupancy without enrichment. With REMORA and related approaches, it is important to express the RBP-rABE fusion as close to endogenous levels as possible. Furthermore, although we show that rABE has a low background editing rate, it is important to include rABE-only controls expressed at a similar level to eliminate possible spurious editing sites.

We anticipate that REMORA is compatible with a variety of other information-encoding strategies in RNA. Since rABE is an adenosine deaminase, REMORA is orthogonal to 4-thiouridine (4sU) or 5-ethynyl uridine (5-EU) incorporation and thus can be combined with diverse metabolic labeling approaches[49,50]. Moreover, subcellular localization can be encoded through biotinylation of RNA via APEX-seq[51], which if combined with REMORA may enable subcellular decoding of RNA binding sites. REMORA combined with fractionated polysome profiling would enable measuring RNA-binding protein association as a function of translation level[7,52,53]. Lastly, as rABE is genetically encodable, cell-type specific RNA-binding protein information can be measured through lineage-specific promoters or other strategies to regulate rABE expression. Long-read compatible, genetically encodable information encoding strategies such as REMORA set the stage for an expanded understanding of single-molecule RNA biology with cell type resolution as well as manipulation of gene expression through targeted RNA deamination.

Methods

General molecular cloning

Gene fragments coding for tested adenosine deaminases (human *ADAR2*, human *ADAT2–ADAT3*, *S. cerevisiae* TAD2–TAD3, *S. aureus* tadA) were synthesized by Twist Biosciences. DNA amplification was performed with Phusion High-Fidelity PCR Master Mix (Thermo Fisher #F630) or KAPA HiFi HotStart ReadyMix (Roche) as described in manufacturer's protocol, unless noted otherwise. PCR products were digested with 1 μl DpnI (NEB # R0176L) for 1 hour and purified by agarose gel extraction (MinElute Gel Extraction Kit, Qiagen #28606). Gibson cloning was done with NEBuilder HiFi DNA Assembly Master

Mix (NEB #E2621L) or Gibson Assembly Master Mix (NEB E2611L) to insert deaminase regions into pICE backbone (Addgene #46960). Assembled plasmids were transformed into Mach1 competent cells (Invitrogen #C862003) for amplification and purified with miniprep (Qiagen) or midiprep (ZymoPURE II Plasmid Midiprep Kit). Plasmids encoding ABE7.10 (Addgene #102919) and hyperTRIBE (Addgene # 154786) were obtained from Addgene. The GFP-UAG-BoxB-mCherry was generated by inserting the UAG-BoxB cassette into pcDNA5-EGFP-NLS-P2AT2A-mCherry-PTS1 (Addgene #87828) with around-the-horn PCR cloning.

General mammalian cell culture

HEK 293T cell line was a gift from Andrei Goga lab. The cells were authenticated by RNA-seq profiling. Cells were cultured in DMEM with 4.5 g / L glucose, L-glutamine & sodium pyruvate (Corning #10–0130CV) and 10% FBS (Avantor #97068–085) at 37 °C, 5% CO_2. Cell passage was done with trypsin (Corning #25–053-CI) with 2–5 min digestion at room temperature.

Flow cytometry measurement of editing efficiency

Plasmids containing the GFP-UAG-BoxB-mCherry reporter and the putative deaminase were co-transfected into HEK 293T cells. Cells were seeded in 12 or 24 well plates at day 0 and reached 70–90% confluency at day 1 for transfection. Transfection was done with lipofectamine (Invitrogen) following manufacturer's protocol. Cells were collected after 48 hours, washed with 1xPBS, and resuspended in 1xPBS containing 0.2 μg mL^{-1} DAPI to mark dead cells. Flow cytometry was performed using Fortessa SORP to measure the florescent signal of GFP (488nm 50mW laser, Blue B detector), mCherry (561nm. 50mW laser, YG D detector) and DAPI (405nm 100mW laser, Violet F detector). Positive cell populations were gated with wild type HEK 293T cells as negative controls and mono-color HEK 293T cells expressing only GFP or mCherry, after filtering for single cells. A-to-I editing efficiency was measured as the percentage of GFP+mCherry+ population in GFP+ population, and as the ratio of the mCherry to GFP signal in the GFP+ population. 3 replicates of measurements were taken from distinct samples. Data analysis was done with Flowjo v10.8.1.

Sanger sequencing verification of editing rates

Cellular RNA extraction was performed with Direct-zol RNA miniprep Kit (Zymo Research #R2053) following manufacturer's protocol. Reverse transcription was done with Maxima First Strand cDNA Synthesis Kit for RT-qPCR (Thermo Scientific #K1641). cDNA regions around editing sites were amplified with primers 5′-GGGCATCGACTTCAAGGAGGACGG-3′ and 5′-CGCCGTCCTGCAGGGAGGAG-3′ and sent to Genewiz for sanger sequencing, with sequencing primer 5′ CATGGTCCTGCTGGAGTTCGTG-3′.

Directed protein evolution

To determine the optimal kanamycin concentration for selection, we generated pET15b-UAG-BoxB-KanR-ABE7.10 plasmid and pET15b-UAG-BoxB-KanR-ABE7.10-

E59A plasmid, which encodes an inactive mutant of ABE7.10 as a negative control. We also generated a pET15b-UGG-BoxB-KanR-ABE7.10 plasmid that allows constant expression of KanR as positive control. Each of these plasmids were transformed into Mach1 competent cells, cultured in LB medium with 0, 0.3, 0.375, 0.5, and 0.75x of normal concentration of kanamycin (50 mg/L), 37 °C. Cell growth was measured as OD600 over 48 hours incubation time. 0.375x kanamycin concentration was the highest concentration that allowed for pET15b-UAG-BoxB-KanR-ABE7.10 cell growth and completely inhibited the growth of pET15b-UAG-BoxB-KanR-ABE7.10-E59A negative control cells.

The ABE7.10 mutant library was generated with error prone PCR using GeneMorph II (Agilent #200550) in the high mutation rate condition as described in manufacturer's protocol. For each round of mutagenesis PCR, eight 25 μl reactions were performed, purified by agarose gel extraction (MinElute Gel Extraction Kit, Qiagen #28606) and pooled together. The ABE 7.10 mutant library was inserted into pET15b-UAG-BoxB-KanR plasmid downstream of lac promoter by large scale Gibson reaction and transformation. Ninety-six 12 μl Gibson reactions were performed in 96-well PCR plates using Gibson Assembly Master Mix (NEB E2611L). The Gibson reaction mix was transformed into Mach1 competent cells (4 μl Gibson into 40 μl cells, 288 reactions in 3×96-well plates) following manufacture's protocol (Invitrogen C862003). After transformation, cells were incubated in 100 ml LB medium (US Biological, #L1505) with 100ug/ml ampicillin at 37 °C overnight, and plasmids were purified with Midiprep (ZymoPURE II Plasmid Midiprep Kit). A small volume of transformed cells (40 μl from 5 ml recover medium) was diluted 1:100, 1:1000 and 1:10000 and plated on Amp+ LB-agarose plates to evaluate transformation efficiency, and from this we estimated we had ~1.75×10^4 colonies in total.

For kanamycin selection experiments, plasmids containing the ABE 7.10 mutant library were transformed back into Mach1 cells in a similar strategy and we obtained ~1×10^7 colonies in total. Cells were cultured in 100 ml LB medium with 0.375x kanamycin (18.75 mg/L), 37 °C overnight. Enriched plasmids were purified with midiprep (ZymoPURE II Plasmid Midiprep Kit), and the ABE 7.10 mutant region was PCR amplified with Phusion High-Fidelity PCR Master Mix (Thermo Fisher) for 20 cycles and sequenced by Genewiz amplicon-EZ sequencing. Mutations enriched in first round selection were cloned into the selection plasmid individually and used as template for GeneMorph error-prone PCR for second round selection.

Stable cell line generation

Lentiviral plasmids expressing rABE or codon-optimized APOBEC1 (BE4max[25]) were generated by inserting the deaminase region into pSLIK-TT-RBFOX2 plasmid backbone (Addgene #59770) with Gibson cloning. Lentiviral plasmids expressing RBP-rABE or RBP-APOBEC1 were generated with similar strategy. The Rbfox2 coding region was from the pSLIK-TT-RBFOX2 plasmid (Addgene #59770), PUM1 was PCR amplified from pMAL-PUM1 (Addgene #120384) and PUM2 was PCR amplified from pMAL-PUM2 (Addgene #120385). For PUM1 plasmids, the hygromycin resistance gene in the plasmid was replaced with a puromycin resistance gene for dual selection with PUM2 plasmids. These plasmids were transfected together with lentiviral packaging plasmids (1.5 μg lentiviral plasmid, 0.1

μg pGag/Pol, 0.1 μg pREV, 0.1 μg pTAT, 0.2 μg pVSVG per 1 reaction in a 6-well plate well) into HEK 293T cells with TransIT-LT1 transfection reagent (Mirus #2304) following manufacturer's protocol. Cell culture supernatant containing lentivirus was collected and filtered after 2 days for spinfection to transduce HEK 293T cells. Briefly, 200–400 μl viral supernatant and 8 μg ml^{-1} polybrene was added in DMEM medium to total volume of 1 ml per well in 12-well plates, cells were centrifuged at 1000xg, 30 °C for 2 hours and returned to normal cell culture condition. 2 Days after spinfection, cells were treated with either 200 ug ml^{-1} hygromycin B (Gold Bio #H-270–1) or 2 μg ml^{-1} puromycin (Gold Bio #P-600–100) for at least 5 days. For dual labeling cells, we first generated HEK 293T-Tet-PUM2-rABE or HEK 293T-Tet-PUM1-APOBEC1 cells with hygromycin selection, PUM1-rABE or PUM1-APOBEC1 were then introduced by a second round of puromycin selection. Cells were treated with 0, 50, 100, 1000 ng / mL doxycycline (abcam # ab141091) for 2 days before western blot or RNA extraction.

Cell proliferation assays

Doxycycline inducible rABE expressing HEK293T cells and wt HEK293T cells were used for cell proliferation assays. Cells were seeded in 12-well plates with low density (1×10^3 cells in 1 ml medium per well). After 0, 24, 48, 72 hours of doxycycline treatment (0, 50, or 1000 ng/ ml), cells were counted with flow cytometry and CountBright™ Absolute Counting Beads (Invitrogen, C36950) following manufacturer's protocol. After 48 hours, cells were supplied with fresh medium with corresponding doxycycline concentration without passing. Cell proliferation was also measured with CFSE staining (CellTrace™ CFSE Cell Proliferation Kit, Invitrogen C34554). At day 0, cells were incubated with CFSE (1:500 dilution in 1x PBS) for 20 min at 37 °C. After CFSE labelling, cells were spin down and resuspended in DMEM medium, seeded in 12-well plates. After doxycycline treatment for 0, 24, 48, 72 hours, cells were collected for CFSE measurement with flow cytometry. Data analysis was done with Flowjo v10.8.1.

Western blot

Cells were lysed with RIPA buffer (Thermo Fisher # 89900) following manufacturer's protocol. Western blots were performed with 4–20% Mini-PROTEAN(R) TGX(tm) Precast Protein Gels (Bio-rad, #4561095 or #4561096) and Trans-Blot ® Turbo™ Mini PVDF Transfer Packs (Bio-rad #1704156). After membrane transfer, PVDF membranes were blocked with 5% milk in PBST buffer for 1 hour, incubated with primary antibody at 4 °C overnight, then incubated with secondary antibody at room temperature for 1 hour. For western blot quantification, 3 replicates of measurements were taken from distinct samples. Antibodies and concentration used in this study are: Monoclonal ANTI-FLAG® M2-Peroxidase (HRP) (Sigma-Aldrich #A8592) at 1:50k; anti-RBFox2 (abcam ab57154) at 1:500 – 1:1k; anti-PUM1 (abcam #ab3717) at 1:5k – 1:10k; anti-PUM2 (abcam #ab92390) at 1:1k; anti-β-Actin-HRP (Santa Cruz Biotech #sc-47778) at 1:5k; anti-β-Actin-680 (Santa Cruz Biotech #sc-47778-AF680) at 1:5k; IRDye Goat anti mouse 800CW (Licor #926–32210), IRDye Goat anti-Rabbit IgG 800CW (Licor #926–32211), IRDye Donkey anti goat 800CW (Licor #926–32214).

RNA extraction and high-throughput RNA sequencing

Cellular RNA extraction was performed with Direct-zol RNA miniprep Kit (Zymo Research #R2053) following manufacturer's protocol. RNA quality was verified with Bioanalyzer (Agilent RNA 6000 Pico) before library prep. For Illumina high-through sequencing library preparation, we used NEBNext® rRNA Depletion Kit (Human/Mouse/Rat) (NEB #E7405), NEBNext® Ultra™ II Directional RNA Library Prep Kit for Illumina® (NEB #E7765) and NEBNext Multiplex Oligos for Illumina (NEB #E6440S and NEB #E6442S) following manufacturer's protocol. 3 replicates of RNA-seq libraries were generated from distinct samples. Sequencing was performed with a NovaSeq (S4 300 flow cell, paired-end 2×150bp) at the Center For Advanced Technology at the University of California, San Francisco.

Long-read sequencing

Pac-bio long-read sequencing library was done with the SMRTbell Express Template Prep Kit 2.0 (Pacbio) with the Iso-seq protocol as described by manufacturer. RNA samples used in long-read sequencing were the same as used in Illumina sequencing, as described above. We used a customed TSO oligo containing 8 nt UMI barcodes for de-duplication: 5′ GCAATGAAGTCGCAGGGTTG [UMI N*8] HHHHHHHHrGrGrG 3′. Sequencing was performed with a shared PacBio Sequel II system on a SMRT Cell 8M.

High throughput sequencing data analysis

All scripts used in this study are available: https://github.com/linyz/remora. Sequencing reads were aligned with STAR (v2.7.10a)[54] to GRCh38 human genome reference and processed with samtools (ver 1.13)[55], picard (v2.27.1) MarkDuplicates, and GATK (v4.2.6.1) SplitNCigarReads[56]. Differential analysis was performed with Rsubreads (v2.8.2) and edgeR (v3.36)[57]. Variant calling was performed with bcftools (v1.15)[58]:

```
bcftools mpileup -f [input_reference.fa] -R [input_annotation.bed] -d
10000000 -I -a DP,AD,ADF,ADR,SP,INFO/AD,INFO/ADF,INFO/ADR [input.bam]|
bcftools filter -i 'INFO/AD[1-]>2 & MAX(FORMAT/DP)>50 -O v - > [output.vcf]
```

Variants were annotated with VariantAnnotation (v1.40)[59] (R Bioconductor).

Significant editing site calling was done on three replicates of treated cells and three replicates of control cells using the Cochran–Mantel–Haenszel test using allele read counts and total read counts at each nucleotide location similar to previously described[60,61]. P-values were adjusted with the Benjamini-Hochberg method.

Sequence motif analysis around editing sites was performed with seqLogo (v1.16, R, Bioconductor) in +/− 5nt region. De novo sequence motif discovery was done with Homer[62] in +/− 50 nt region around editing sites. CentriMo[33] analysis was performed with the MEME Suite (v5.4.1) (https://meme-suite.org/meme/tools/centrimo).

RNA secondary structure modeling was done with RNAfold v2.4.10[35] in +/− 30 nt region around editing sites with default parameters.

GSEA analysis was performed with GSEA v4.3.2[63] (GSEAPreranked).The input gene list was generated by ranking genes with correlation scores of numbers of PUM1 and PUM2 binding induced editing events.

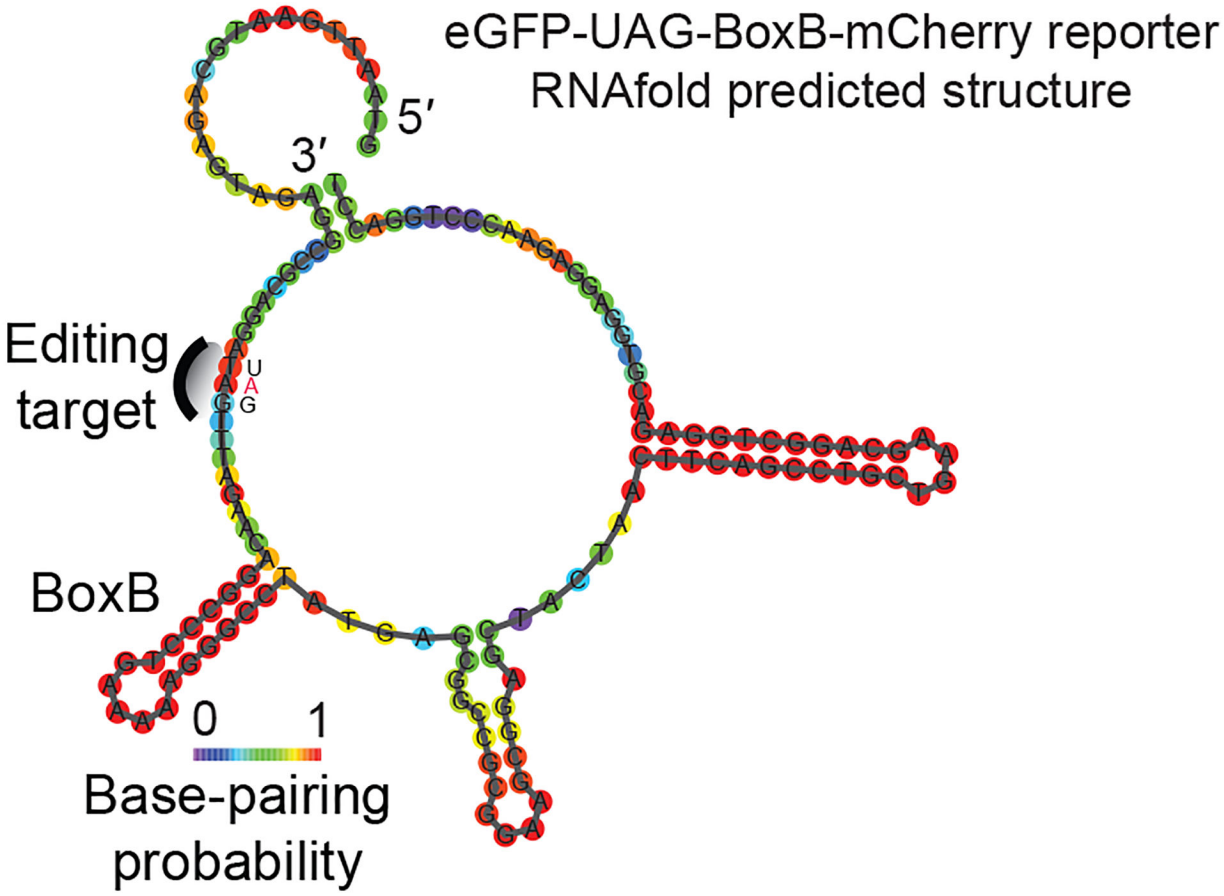
Long-read sequencing data analysis

Long-read sequencing alignment was performed with Pacbio's IsoSeq3 pipeline (https://github.com/ylipacbio/IsoSeq3). Single molecule level mutations were identified with custom python scripts, which are available at https://github.com/linyz/remora. RBP-rABE or RBP-APOBEC1 induced RNA editing events were identified by first filtering out mutations exists in rABE-Flag or APOBEC1-Flag control samples.

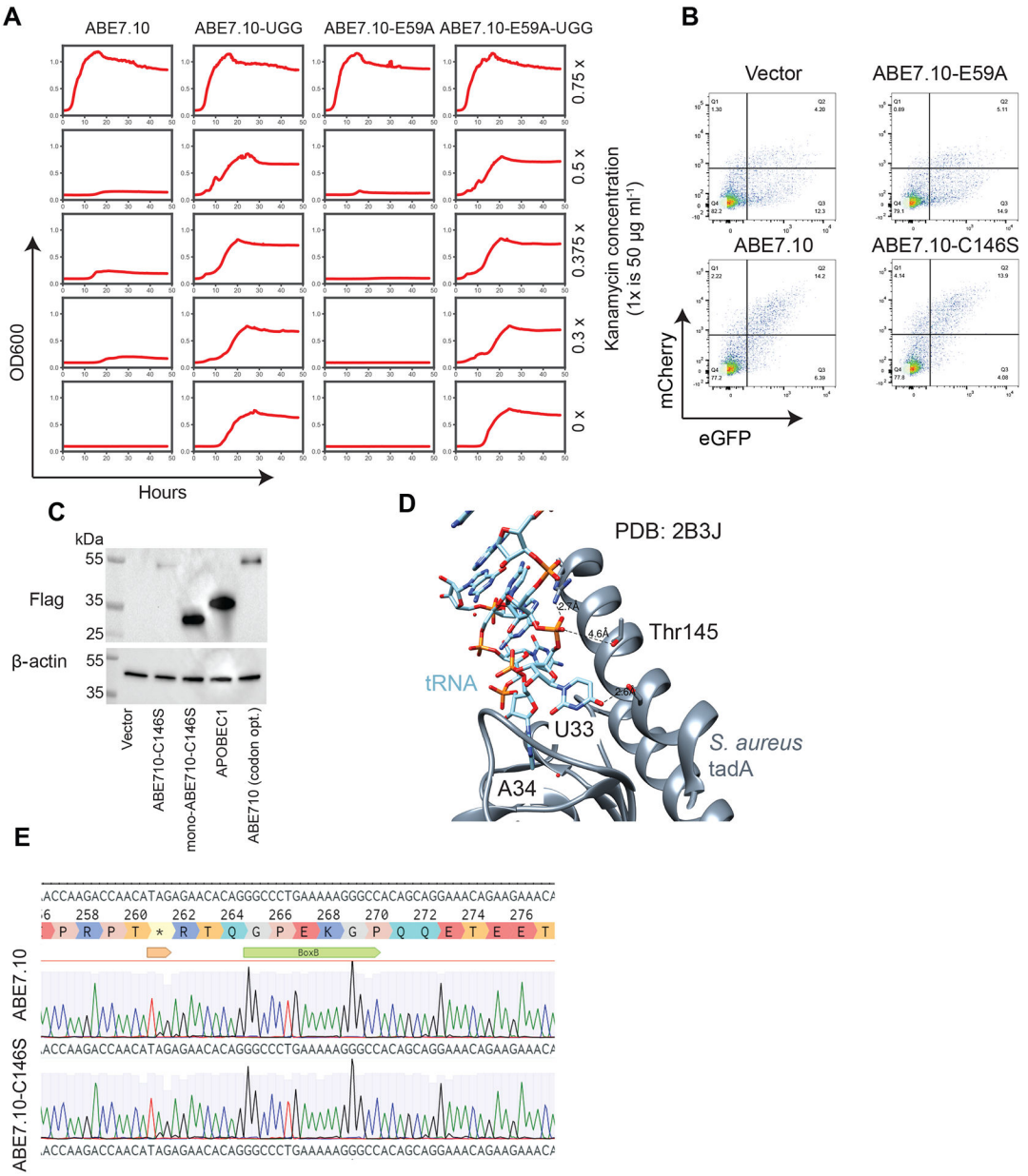
Isoform specific binding events were identified by a custom python script. Briefly, for each gene, we count number of RNA editing events and number of reads for each exon; we then calculate odds-ratio enrichment of RNA editing events in an exon A, associated with inclusion or exclusion of a different exon B, for all pairs of exons in the given gene. Significance and p-value of such enrichment was calculated with Mann Whiney U test, and corrected with BH adjustment.

Single-molecule level co-binding analysis was performed with a custom python script. Briefly, we count the number of A-to-I and C-to-U edits in each read and calculated the spearman correlation of A-to-I and C-to-U edits for a given gene. For the random shuffle control, we count the number of RNA editing events in PUM1-APOBEC1 or PUM2-rABE single labelled samples, and calculated correlation scores after random shuffling to create synthetic dual labeling data where PUM1 induced editing events were independent of PUM2.

Extended Data



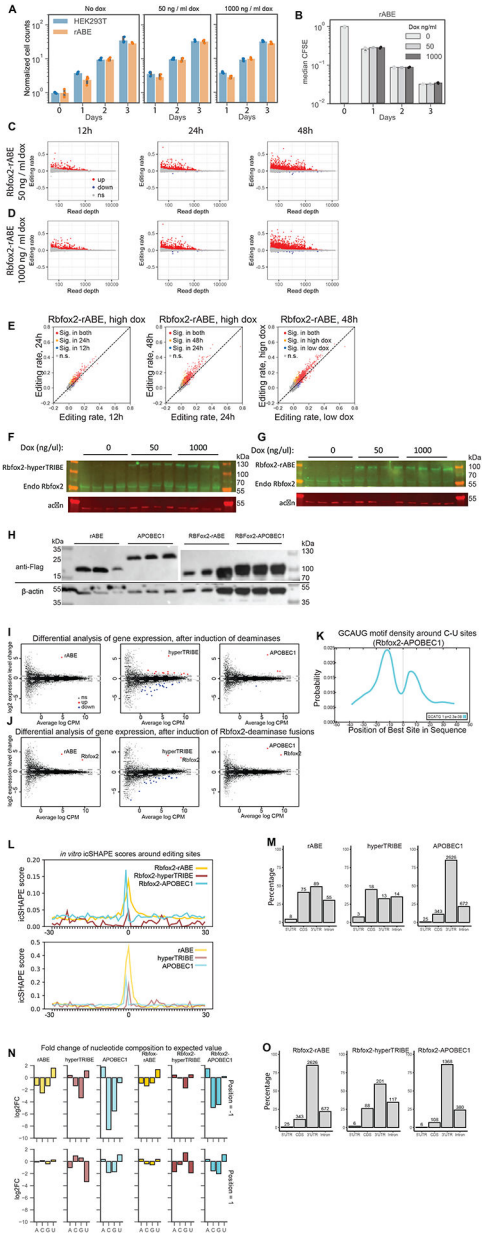
Extended Data Figure 1.
RNA secondary structure modeling around UAG-BoxB cassette with RNAfold. BoxB forms a stable stem-loop structure, while the UAG stop codon was predicted to be unstructured.



Extended Data Figure 2.

Directed evolution for an improved ssRNA A-to-I editor. **(A)** OD600 measurement of E. coli growth curve after transformation of ABE7.10 under different kanamycin concentrations. 0.375 x concentration (18.75 mg/L) was the optimal concentration that allowed growth of ABE.10 expressing cells and inhibited growth of ABE7.10-E59A negative control cells. **(B)** Scatter plots of flow cytometry measurement of the editing efficiency of ABE7.10 and ABE7.10-C146S. **(C)**Western blot of ABE variants in Figure 2D, ABE710 had low expression and was not detected. Expression level and fluorescent reporter assay for rABE were independently tested at least three times, with consistent results. **(D)** Structural model of the S. aureus tadA protein near Thr145 (corresponding to E. coli Ser146). **(E)** Sanger sequencing measurement of A-to-I editing rates introduced by ABE7.10 and ABE7.10-

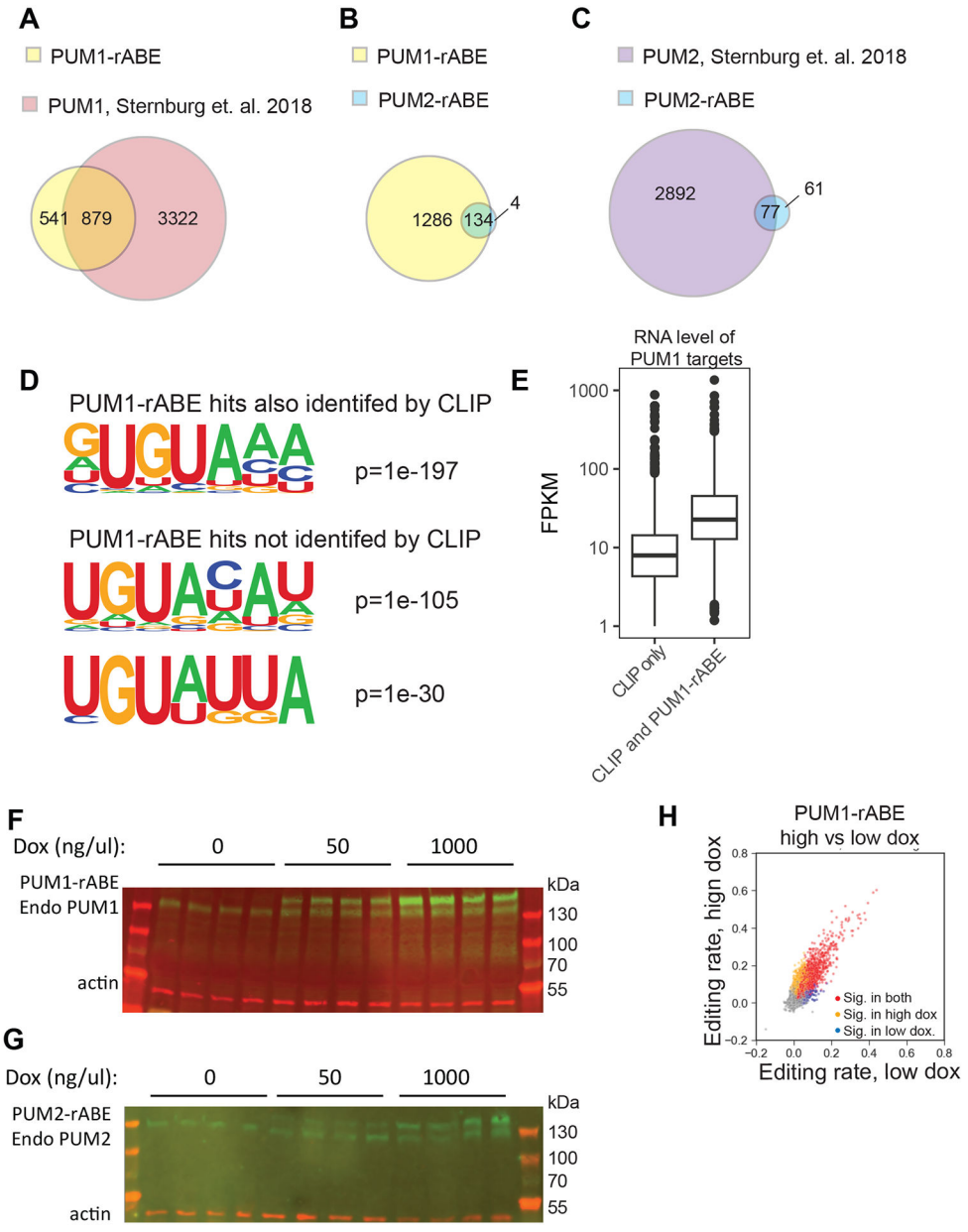
C146S on an adenosine-rich UAG-BoxBreporter. Several A-to-I editing were observed flanking both sides of BoxB in different sequence contexts.



Extended Data Figure 3.

Transcriptome-wide characterization of rABE. (A) Cell proliferation of wt HEK293T cells and doxycycline inducible rABE-Flag expressing HEK293T cells measured by cell counts. Cells were incubated with 0, 50, or 1000 ng/mL doxycycline for 3 days. Error bars are 95% CI around mean. n=3 biologically independent samples. (B) Cell proliferation measured by CFSE labeling of HEK293T stable cell lines expressing rABE under 0, 50, 1000 ng/mL doxycycline. Error bars are 95% CI around mean. n=3 biologically independent samples. (C, D) Scatter plots of significant A-to-I editing events induced by Rbfox2-rABE

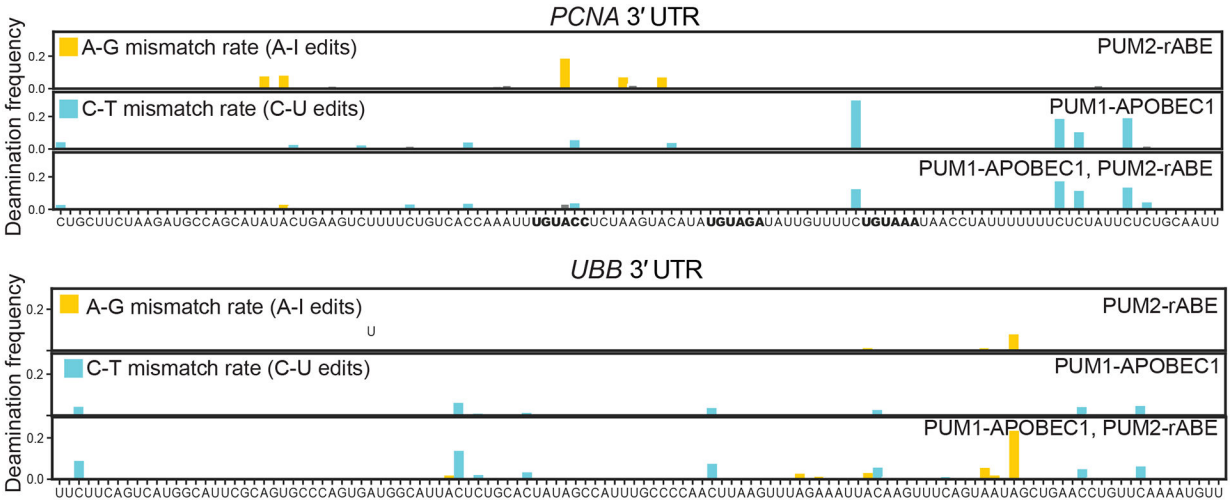
under 50 ng/mL (C) and 1000 ng/mL (D) after 12, 24 and 48 hours. (E) Scatterplot of Rbfox2-rABE induced A-to-I editing rate, comparing 24 hours incubation vs 12 hours incubation of 1000 ng/mL doxycycline; 48 hours incubation vs 24 incubation of 1000 ng/mL doxycycline; and 1000 ng/mL vs 50 ng/mL doxycycline after incubation for 48 hours. (F) Western blot of Rbfox2-hyperTRIBE and endogenous Rbfox2 in HEK293T dox inducible cell lines, 4 biologically independent replicates each. (G) Western blot of Rbfox2-rABE and endogenous Rbfox2 in HEK293T dox inducible cell lines, 4 biologically independent replicates each. (H) Western blot measurement of rABE, APOBEC1, Rbfox2-rABE, Rbfox2-APOBEC1 expression levels under 1000 ng/mL doxycycline condition, 3 replicates each. All proteins contain a Flag tag at their C-terminus and are blotted with anti-Flag-HRP. (I, J) Differential expression analysis comparing transcriptome-wide gene expression level after doxycycline induction (1000 ng/mL vs no doxycycline control) of of rABE, hyperTRIBE and APOBEC1 (I) and of Rbfox2-rABE, Rbfox2-hyperTRIBE and Rbfox2-APOBEC1 (J). Expression of rABE or APOBEC1 alone and their Rbfox2 fusions do not change gene expression levels, while hyperTRIBE induced more genes with expression level changes. (K) CentriMo analysis around Rbfox2-APOBEC1 induced C-to-U editing sites. (L) In vitro structure probing icSHAPE scores around editing sites. Rbfox2-hyperTRIBE and hyperTRIBE showed lower SHAPE reactivity around the editing sites, and lower SHAPE reactivity is a signal of increased structure. (M) Distribution of rABE, hyperTRIBE and APOBEC1 background editing sites across transcript regions. (N) Nucleotide composition at -1 or +1 position of editing sites. Y-axes are log2 fold change or expected value (0.25). (O) Distribution of Rbfox2-rABE, Rbfox2-hyperTRIBE and Rbfox2-APOBEC1 editing sites across transcript regions.



Extended Data Figure 4.

rABE recovers known binding patterns of pumillio proteins. (A) Venn diagram comparison of PUM1-rABE targets (this study) and PUM1 targets identified by CLIP-seq (Sternburg et al.37). (B) Venn diagram comparison of PUM1-rABE targets and PUM2-rABE targets. (C) Venn diagram comparison of PUM2-rABE targets (this study) and PUM2 targets identified by CLIP-seq (Sternburg et al.37). (D) De novo motif discovery in PUM1-rABE hits that are also idneitifed by CLIP-seq or not identified by CLIP-seq. Both populations successfully recovered the known PBE motif UGUANA. (E) PUM1 target genes identified by CLIP-seq but not PUM1-rABE showed significantly lower expression level ($p < 2.2e-16$, t test, two-sided, n=3322 for CLIP only group, n=879 for CLIP and PUM1-rABE group). Centre: median; box bounds: 25% and 75% percentile; whisker: 1.5x of interquartile range. All

outliers are marked as dots. (F) Western blot of PUM1-rABE and endogenous PUM1 in HEK293T dox inducible cell lines, 4 independent biological replicates each. (G) Western blot of PUM2-rABE and endogenous PUM2 in HEK293T dox inducible cell lines, 4 independent biological replicates each. (H) Scatter plot showing editing rate in PUM1-rABE cell lines, under 50 or 1000 ng/mL doxycycline conditions.



Extended Data Figure 5.
PUM1-APOBEC1 and PUM2-APOBEC1 induced editing patterns in the 3′ UTR region of UBB and PCNA 3′ UTR regions.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgements:

We thank members of the Floor lab for feedback on this work and their continued support. Computation was supported by the UCSF Wynton high-performance computing infrastructure, PacBio sequencing was supported through instrumentation shared with V. Ramani (Gladstone Institutes), high-throughput short-read sequencing was supported by the UCSF Center for Advanced Technology, and flow cytometry was supported by the Parnassus Flow Cytometry Core at UCSF. This work was supported by the UCSF Program for Breakthrough Biomedical Research, funded in part by the Sandler Foundation (to SNF) and the National Institutes of Health DP2GM132932 and R35GM149255 (to SNF). SNF is a Pew Scholar in the Biomedical Sciences, supported by The Pew Charitable Trusts.

Reagent and data availability

Plasmids encoding rABE are available through Addgene (#191383-#191386). Sequencing data is available at GEO accession number GSE216334.

References

1. Gerstberger S, Hafner M & Tuschl T A census of human RNA-binding proteins. Nat. Rev. Genet 15, 829–845 (2014). [PubMed: 25365966]
2. Dominguez D et al. Sequence, Structure, and Context Preferences of Human RNA Binding Proteins. Mol. Cell 70, 854–867.e9 (2018). [PubMed: 29883606]

3. Buenrostro JD et al. Quantitative analysis of RNA-protein interactions on a massively parallel array reveals biophysical and evolutionary landscapes. Nat. Biotechnol 32, 562–568 (2014). [PubMed: 24727714]
4. Ule J et al. CLIP Identifies Nova-Regulated RNA Networks in the Brain. Science (80-.) 302, 1212–1215 (2003).
5. Van Nostrand EL et al. Robust transcriptome-wide discovery of RNA-binding protein binding sites with enhanced CLIP (eCLIP). Nat. Methods 13, 508–514 (2016). [PubMed: 27018577]
6. Hafner M et al. Transcriptome-wide Identification of RNA-Binding Protein and MicroRNA Target Sites by PAR-CLIP. Cell 141, 129–141 (2010). [PubMed: 20371350]
7. Floor SN & Doudna JA Tunable protein synthesis by transcript isoforms in human cells. Elife 5, 1–25 (2016).
8. Blair JD, Hockemeyer D, Doudna JA, Bateup HS & Floor SN Widespread Translational Remodeling during Human Neuronal Differentiation. Cell Rep. 21, 2005–2016 (2017). [PubMed: 29141229]
9. McMahon AC et al. TRIBE: Hijacking an RNA-Editing Enzyme to Identify Cell-Specific Targets of RNA-Binding Proteins. Cell 165, 742–753 (2016). [PubMed: 27040499]
10. Meyer KD DART-seq: an antibody-free method for global m6A detection. Nat. Methods 16, 1275–1280 (2019). [PubMed: 31548708]
11. Brannan KW et al. Robust single-cell discovery of RNA targets of RNA-binding proteins and ribosomes. Nat. Methods 18, 507–519 (2021). [PubMed: 33963355]
12. Abruzzi K, Ratner C & Rosbash M Comparison of TRIBE and STAMP for identifying targets of RNA binding proteins in human and Drosophila cells. RNA rna.079608.123 (2023) doi:10.1261/RNA.079608.123.
13. Rosenberg BR, Hamilton CE, Mwangi MM, Dewell S & Papavasiliou FN Transcriptome-wide sequencing reveals numerous APOBEC1 mRNA-editing targets in transcript 3′ UTRs. Nat. Struct. Mol. Biol 18, 230–236 (2011). [PubMed: 21258325]
14. Goldstrohm AC, Hall TMT & McKenney KM Post-transcriptional Regulatory Functions of Mammalian Pumilio Proteins. Trends Genet. 34, 972–990 (2018). [PubMed: 30316580]
15. Jarmoskaite I et al. A Quantitative and Predictive Model for RNA Binding by Human Pumilio Proteins. Mol. Cell 74, 966–981.e18 (2019). [PubMed: 31078383]
16. Campbell ZT et al. Cooperativity in RNA-protein interactions: global analysis of RNA binding specificity. Cell Rep. 1, 570–581 (2012). [PubMed: 22708079]
17. Sternburg EL, Estep JA, Nguyen DK, Li Y & Karginov FV Antagonistic and cooperative AGO2-PUM interactions in regulating mRNAs. Sci. Rep 8, 1–13 (2018). [PubMed: 29311619]
18. Rahman R, Xu W, Jin H & Rosbash M Identification of RNA-binding protein targets with HyperTRIBE. Nat. Protoc 13, 1829–1849 (2018). [PubMed: 30013039]
19. Gaudelli NM et al. Programmable base editing of A-T to G-C in genomic DNA without DNA cleavage. Nature 551, 464–471 (2017). [PubMed: 29160308]
20. Thomas JM & Beal PA How do ADARs bind RNA? New protein-RNA structures illuminate substrate recognition by the RNA editing ADARs. Bioessays 39, (2017).
21. Montiel-Gonźalez MF, Vallecillo-Viejo IC & Rosenthal JJC An efficient system for selectively altering genetic information within mRNAs. Nucleic Acids Res. 44, e157–e157 (2016). [PubMed: 27557710]
22. Grünewald J et al. Transcriptome-wide off-target RNA editing induced by CRISPR-guided DNA base editors. Nature 569, 433–437 (2019). [PubMed: 30995674]
23. Zhou C et al. Off-target RNA mutation induced by DNA base editing and its elimination by mutagenesis. Nature 571, 275–278 (2019). [PubMed: 31181567]
24. Rees HA, Wilson C, Doman JL & Liu DR Analysis and minimization of cellular RNA editing by DNA adenine base editors. Sci. Adv 5, 1–11 (2019).
25. Koblan LW et al. Improving cytidine and adenine base editors by expression optimization and ancestral reconstruction. Nat. Biotechnol 36, 843–846 (2018). [PubMed: 29813047]
26. Losey HC, Ruthenburg AJ & Verdine GL Crystal structure of Staphylococcus aureus tRNA adenosine deaminase TadA in complex with RNA. Nat. Struct. Mol. Biol 13, 153–159 (2006). [PubMed: 16415880]

27. Richter MF et al. Phage-assisted evolution of an adenine base editor with improved Cas domain compatibility and activity. Nat. Biotechnol 38, 883–891 (2020). [PubMed: 32433547]
28. Conboy JG Developmental regulation of RNA processing by Rbfox proteins. Wiley Interdiscip. Rev. RNA 8, (2017).
29. Gehman LT et al. The splicing regulator Rbfox2 is required for both cerebellar development and mature motor function. Genes Dev. 26, 445–460 (2012). [PubMed: 22357600]
30. Lovci MT et al. Rbfox proteins regulate alternative mRNA splicing through evolutionarily conserved RNA bridges. Nat. Struct. Mol. Biol 20, 1434–1442 (2013). [PubMed: 24213538]
31. Xu W, Rahman R & Rosbash M Mechanistic implications of enhanced editing by a HyperTRIBE RNA-binding protein. RNA 24, 173–182 (2018). [PubMed: 29127211]
32. Flamand MN, Ke K, Tamming R & Meyer KD Single-molecule identification of the target RNAs of different RNA binding proteins simultaneously in cells. Genes Dev. 36, 1002–1015 (2022). [PubMed: 36302554]
33. Bailey TL & MacHanick P Inferring direct DNA binding from ChIP-seq. Nucleic Acids Res. 40, e128–e128 (2012). [PubMed: 22610855]
34. Chan D et al. Diverse functional elements in RNA predicted transcriptome-wide by orthogonal RNA structure probing. Nucleic Acids Res. 49, 11868–11882 (2021). [PubMed: 34634799]
35. Lorenz R et al. ViennaRNA Package 2.0. Algorithms Mol. Biol 6, 26 (2011). [PubMed: 22115189]
36. Alam S, Suzuki H & Tsukahara T Alternative splicing regulation of APP exon 7 by RBFox proteins. Neurochem. Int 78, 7–17 (2014). [PubMed: 25125370]
37. Wei C et al. RBFox2 Binds Nascent RNA to Globally Regulate Polycomb Complex 2 Targeting in Mammalian Genomes. Mol. Cell 62, 875–889 (2016). [PubMed: 27211866]
38. Lu G & Hall TMT Alternate modes of cognate RNA recognition by human PUMILIO proteins. Structure 19, 361–367 (2011). [PubMed: 21397187]
39. Bohn JA et al. Identification of diverse target RNAs that are functionally regulated by human pumilio proteins. Nucleic Acids Res. 46, 362–386 (2018). [PubMed: 29165587]
40. Yamada T et al. Systematic Analysis of Targets of Pumilio-Mediated mRNA Decay Reveals that PUM1 Repression by DNA Damage Activates Translesion Synthesis. Cell Rep. 31, 107542 (2020). [PubMed: 32375027]
41. Chen T & Guestrin C XGBoost: A Scalable Tree Boosting System. in Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining vol. 19 785–794 (ACM, 2016).
42. Kedde M et al. A Pumilio-induced RNA structure switch in p27–3' UTR controls miR-221 and miR-222 accessibility. Nat. Cell Biol 12, 1014–1020 (2010). [PubMed: 20818387]
43. Oikonomopoulos S et al. Methodologies for Transcript Profiling Using Long-Read Technologies. Front. Genet 11, 606 (2020). [PubMed: 32733532]
44. Wolf J, Gerber AP & Keller W tadA, an essential tRNA-specific adenosine deaminase from Escherichia coli. EMBO J. 21, 3841–51 (2002). [PubMed: 12110595]
45. Torres AG et al. Inosine modifications in human tRNAs are incorporated at the precursor tRNA level. Nucleic Acids Res. 43, 5145–5157 (2015). [PubMed: 25916855]
46. Ichinose M et al. U-to-C RNA editing by synthetic PPR-DYW proteins in bacteria and human culture cells. Commun. Biol. 2022 51 5, 1–11 (2022).
47. Xu W et al. One-Step, Highly Efficient Site-Directed Mutagenesis by Toxic Protein Selection. Biotechniques 32, 1266–1270 (2002). [PubMed: 12074156]
48. Seo KW & Kleiner RE Profiling dynamic RNA–protein interactions using small-molecule-induced RNA editing. Nat. Chem. Biol (2023) doi:10.1038/s41589-023-01372-9.
49. Herzog VA et al. Thiol-linked alkylation of RNA to assess expression dynamics. Nat. Methods 14, 1198–1204 (2017). [PubMed: 28945705]
50. Schofield JA, Duffy EE, Kiefer L, Sullivan MC & Simon MD TimeLapse-seq: adding a temporal dimension to RNA sequencing through nucleoside recoding. Nat. Methods 15, 221–225 (2018). [PubMed: 29355846]
51. Fazal FM et al. Atlas of Subcellular RNA Localization Revealed by APEX-Seq. Cell 178, 473–490.e26 (2019). [PubMed: 31230715]

52. Sterne-Weiler T et al. Frac-seq reveals isoform-specific recruitment to polyribosomes. Genome Res. 23, 1615–1623 (2013). [PubMed: 23783272]

53. Spies N, Burge CB & Bartel DP 3′ UTR-isoform choice has limited influence on the stability and translational efficiency of most mRNAs in mouse fibroblasts. Genome Res. 23, 2078–2090 (2013). [PubMed: 24072873]

54. Dobin A et al. STAR: Ultrafast universal RNA-seq aligner. Bioinformatics 29, 15–21 (2013). [PubMed: 23104886]

55. Li H et al. The Sequence Alignment/Map format and SAMtools. Bioinformatics 25, 2078–2079 (2009). [PubMed: 19505943]

56. McKenna A et al. The Genome Analysis Toolkit: A MapReduce framework for analyzing next-generation DNA sequencing data. Genome Res. 20, 1297–1303 (2010). [PubMed: 20644199]

57. Chen Y, Lun ATLL & Smyth GK From reads to genes to pathways: differential expression analysis of RNA-Seq experiments using Rsubread and the edgeR quasi-likelihood pipeline. F1000Research 5, 1438 (2016). [PubMed: 27508061]

58. Li H A statistical framework for SNP calling, mutation discovery, association mapping and population genetical parameter estimation from sequencing data. Bioinformatics 27, 2987 (2011). [PubMed: 21903627]

59. Obenchain V et al. VariantAnnotation : a Bioconductor package for exploration and annotation of genetic variants. Bioinformatics 30, 2076–2078 (2014). [PubMed: 24681907]

60. Lin Y, May GE & McManus CJ Mod-seq: A high-throughput method for probing RNA secondary structure. Methods in Enzymology vol. 558 (Elsevier Inc., 2015).

61. Talkish J, May G, Lin Y, Woolford JL & McManus CJ Mod-seq: high-throughput sequencing for chemical probing of RNA structure. RNA 20, 713–720 (2014). [PubMed: 24664469]

62. Heinz S et al. Simple Combinations of Lineage-Determining Transcription Factors Prime cis-Regulatory Elements Required for Macrophage and B Cell Identities. Mol. Cell 38, 576–589 (2010). [PubMed: 20513432]

63. Subramanian A et al. Gene set enrichment analysis: A knowledge-based approach for interpreting genome-wide expression profiles. Proc. Natl. Acad. Sci. U. S. A 102, 15545–15550 (2005). [PubMed: 16199517]

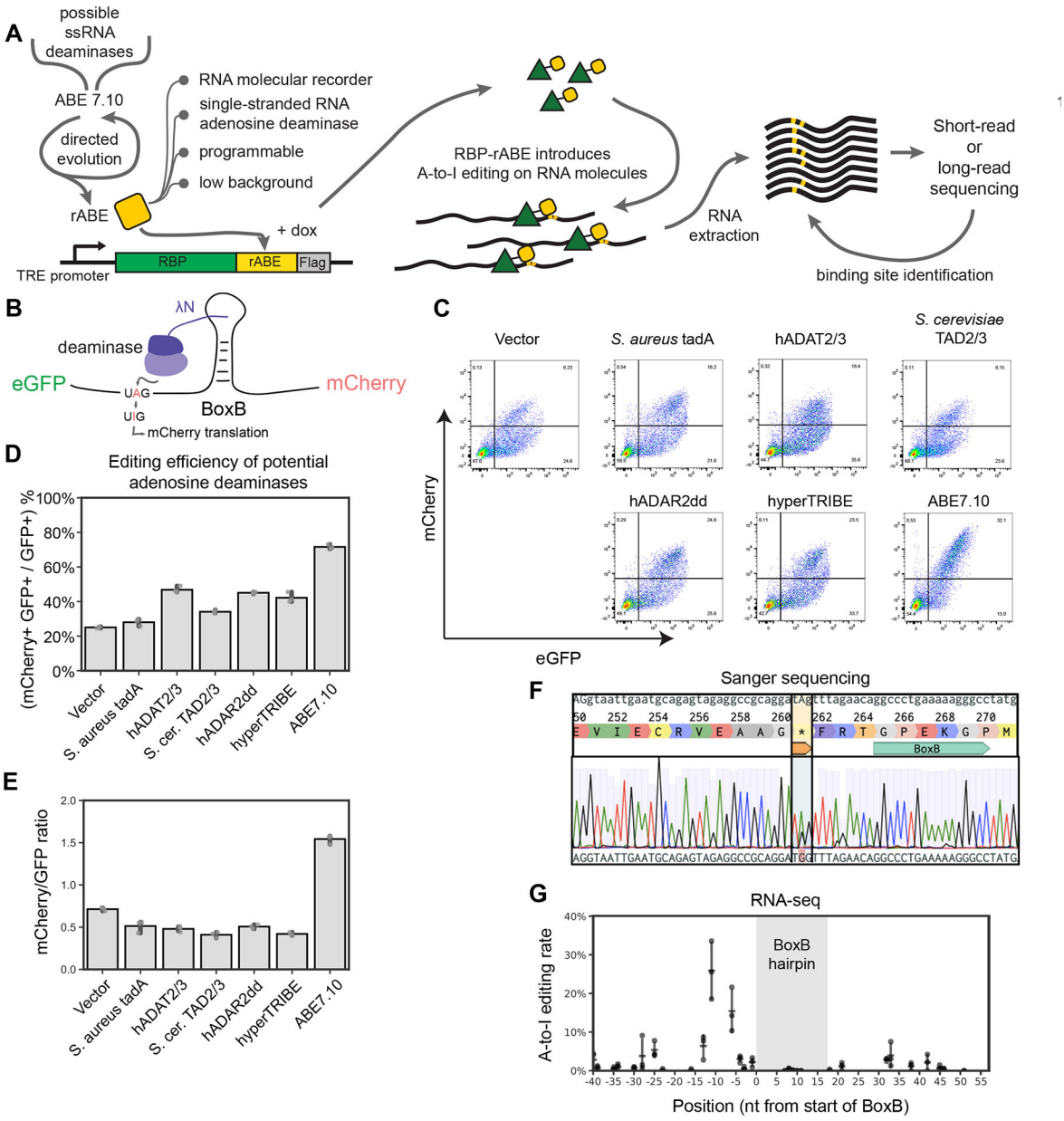


Figure 1:

Screening for a single-stranded RNA adenosine deaminase. **(A)** The design of rABE. Putative RNA adenosine deaminases were screened to identify ABE7.10 as a lead candidate. Directed evolution and construct design produced rABE. Fusion of rABE to RBPs enables detection of binding sites through high-throughput sequencing. **(B)** Design of the eGFP-UAG-BoxB-mCherry deamination reporter system. Putative ssRNA deaminases were fused to a lambdaN peptide to facilitate recruitment near a stop codon by binding to a BoxB stem-loop structure in the eGFP-UAG-BoxB-mCherry reporter mRNA. Adenosine deamination at the UAG stop-codon produces UIG (where I is Inosine), which is decoded as UGG (Trp) and increases expression of mCherry. **(C)** Flow cytometry measurements of HEK 293T cells co-transfected with the eGFP-UAG-BoxB-mCherry reporter and six putative

RNA deaminases fused to lambdaN. Percent of cells in each quadrant is indicated. **(D)** Quantification of flow cytometry measurements by dividing the number of double-positive mCherry+ GFP+ cells by GFP+ cells as indicated by gates shown in (C). Error bars are 95% CI around mean. n=3 biologically independent samples. S. cer: *Saccharomyces cerevisiae*. hADAR2dd: the deaminase domain of human ADAR2. **(E)** Quantification of flow cytometry measurements by GFP/mCherry ratio in GFP+ cells as indicated by gates shown in (C). Error bars are 95% CI around mean. n=3 biologically independent samples. **(F)** Sanger sequencing of a PCR amplicon around the targeted UAG stop-codon. Note that Inosine in RNA is converted to guanosine in DNA during reverse transcription (black trace). **(G)** RNA-seq based measurement of A-to-I editing rates of ABE7.10 on the eGFP-UAG-BoxB-mCherry reporter mRNA after co-transfection in HEK 293T cells. Error bars are SD; hash is at the mean. n=3 biologically independent samples.

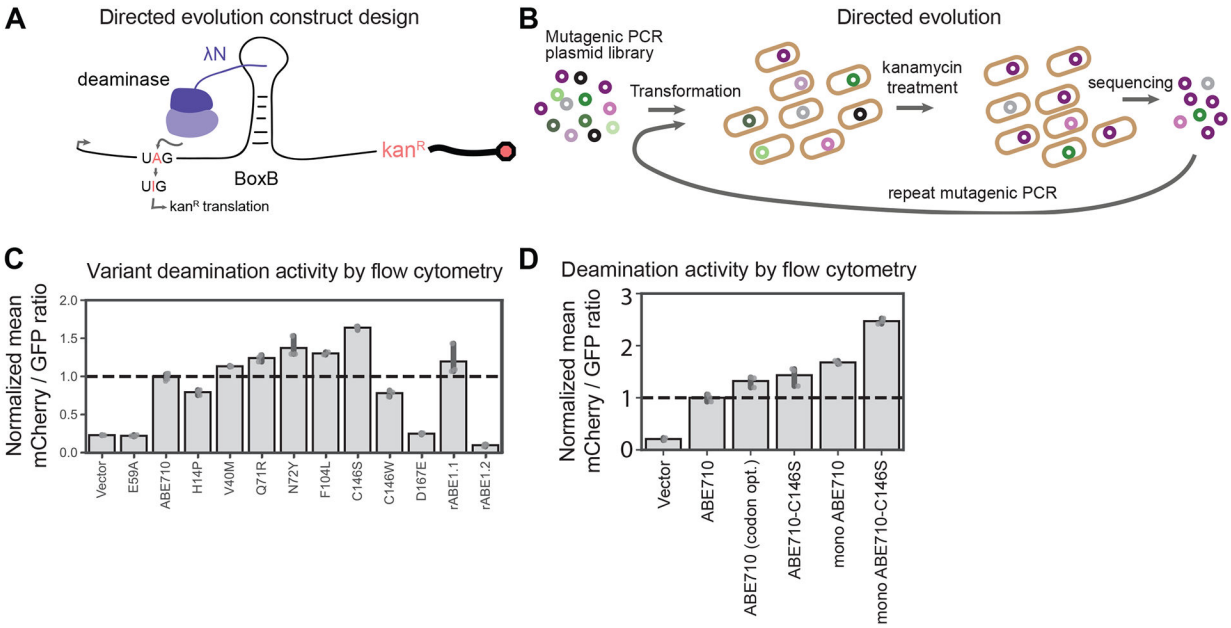


Figure 2:

Engineering an improved ssRNA adenosine deaminase by directed protein evolution. **(A)** Design of UAG-BoxB-kanR reporter for directed protein evolution in *E. coli*. **(B)** Directed protein evolution workflow. Random mutations were introduced into the ABE7.10 protein coding region by error prone PCR and were transformed into *E. coli* Mach1 cells together with the UAG-BoxB-KanR reporter. Cells were incubated with kanamycin overnight to select for ABE7.10 mutants that survived selection. Enriched mutants after selection were identified by amplicon sequencing, transformed back into E. coli cells for a second round of directed protein evolution. **(C)** Editing efficiency of enriched ABE7.10 mutants measured by flow cytometry with the GFP-UAG-BoxB-mCherry reporter. (rABE1.1: ABE7.10 + V40M, Q71R, N72Y, F104L, C146S; rABE1.2: ABE7.10 + V40M, R47S, G66V, Q71R, N72Y, F104L, R107L, C146S). Error bars are 95% CI around mean. n=3 biologically independent samples. **(D)** A comparison of deamination efficiency measured as the mCherry to GFP ratio of the monomeric and dimeric ABE7.10 and rABE variants using flow cytometry and the GFP-UAG-BoxB-mCherry reporter. Error bars are 95% CI around mean. n=3 biologically independent samples.

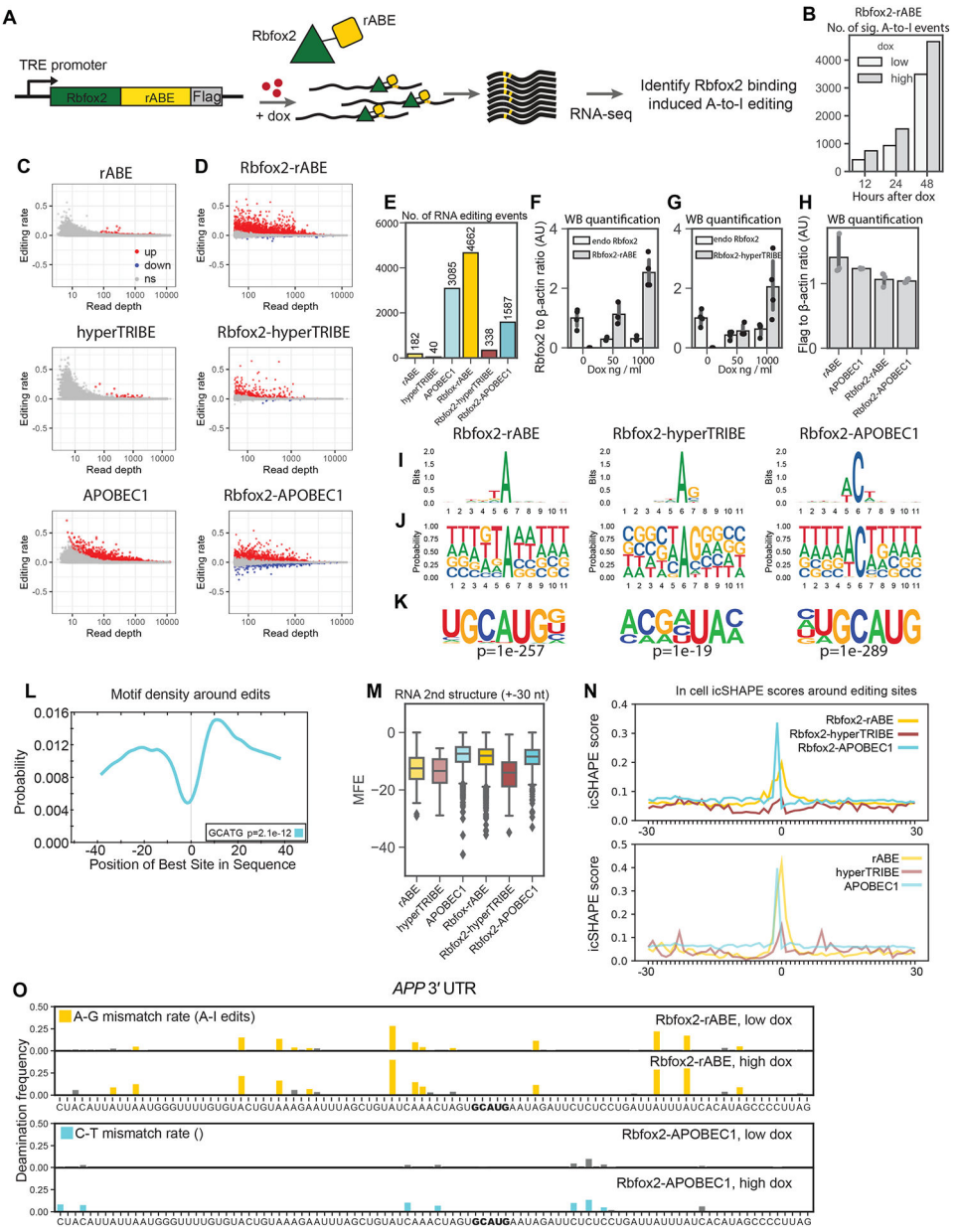


Figure 3:

Transcriptome-wide characterization of rABE. **(A)** Stable HEK293T cell line expressing rABE-Flag or Rbfox2-rABE-Flag was generated by lentiviral transduction. The expression of fusion protein was under regulation of Tet expression system and was induced with addition of doxycycline. Cells were incubated with 0, 50, or 1000 ng/mL doxycycline for 48 hours before RNA extraction and RNA-seq. **(B)** Number of significant A-to-I edits induced by Rbfox2 binding under 50 and 1000 ng/mL doxycycline after 12, 24 and 48 hours. Significant Rbfox2 binding induced A-to-I editing events were determined comparing Rbfox2-rABE-Flag samples and rABE-Flag samples after filtering out DNA mutations and endogenous A-to-I edits ($p_{adj} < 0.05$, Cochran-Mantel-Haenszel test, two-sided, Benjamini-Hochberg adjusted). **(C)** Scatter plots of rABE-Flag, hyperTRIBE-Flag and APOBEC1-Flag

induced background RNA editing under 1000 ng/mL doxycycline after 48 hours. **(D)** Scatter plots of Rbfox2-rABE-Flag, Rbfox2-hyperTRIBE-Flag and Rbfox2-APOBEC1-Flag induced RNA editing under 1000 ng/mL doxycycline after 48 hours. **(E)** Barplot showing number of significant editing sites in C and D. **(F,G)** Western blot protein expression level quantification of Rbfox2-rABE (F) and Rbfox2-hyperTRIBE (G). Error bars are 95% CI around mean. n=4 biologically independent samples. **(H)** Western blot protein expression level quantification with anti-Flag showing rABE, APOBEC1, Rbfox2-rABE, Rbfox2-APOBEC1 were expressed at similar level. Error bars are 95% CI around mean. n=3 biologically independent samples. **(I)** Sequence logo and **(J)** Sequence composition around significant editing sites of Rbfox2-rABE, Rbfox2-hyperTRIBE and Rbfox2-APOBEC1. **(K)** *De novo* motif discovery of Rbfox2-rABE, Rbfox2-hyperTRIBE and Rbfox2-APOBEC1 in +/− 50 nt around significant editing events. The Rbfox2 binding motif is GCAUG. **(L)** CentriMo analysis of Rbfox2-rABE around A-to-I editing sites. Rbfox2-rABE A-to-I editing events are between 15–20 nt both upstream and downstream of Rbfox2 binding motifs. **(M)** Minimum free energy (MFE) of predicted RNA structure in a window of +/− 30 nt around editing sites. Note a decrease in MFE indicates an increase in RNA structure propensity. Sample size: all significant editing sites are included in generating this boxplot, as stated in (E); centre: median; box bounds: 25% and 75% percentile; whisker: 1.5x of interquartile range. All outliers are marked as diamonds. **(N)** In-cell structure probing icSHAPE scores around editing sites. Rbfox2-hyperTRIBE and hyperTRIBE showed lower SHAPE reactivity around the editing sites, and lower SHAPE reactivity is a signal of increased structure. **(O)** Example of Rbfox2-rABE and Rbfox2-APOBEC1 introduced editing patterns around the GCAUG motif in the 3′ UTR region of APP.

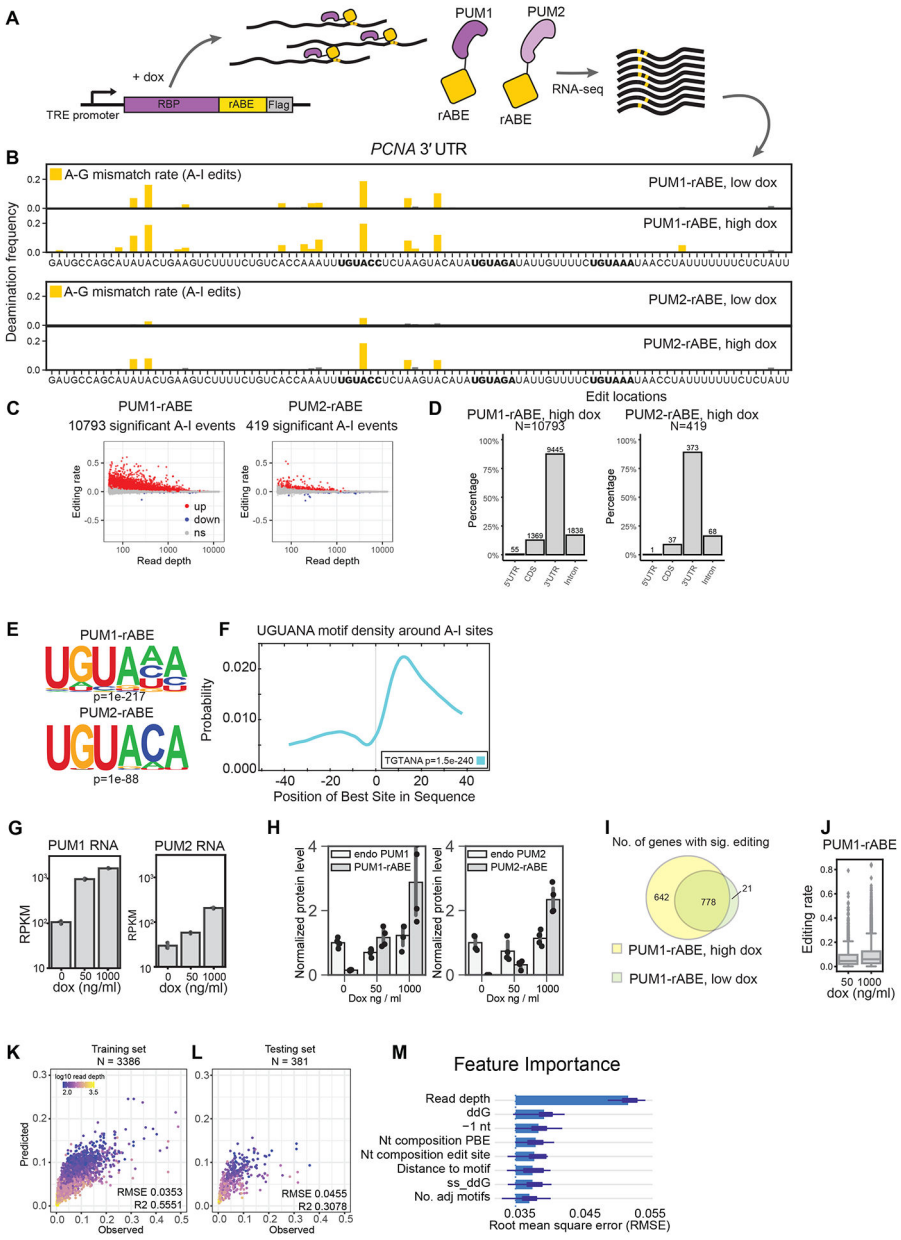


Figure 4:

rABE recovers known binding patterns of pumillio proteins. **(A)** Stable HEK293T cells lines expressing PUM1-rABE or PUM2-rABE under a doxycycline inducible promoter were generated by lentiviral transduction. Cells were incubated with 0 (control), 50 (low), or 1000 (high) ng/mL doxycycline before RNA extraction and RNA-seq. **(B)** PUM1-rABE or PUM2-rABE introduced A-to-I editing events adjacent to PUM1/2 binding sites that were identified as A-to-G mutations by high throughput sequencing, as shown in the 3′ UTR region of PCNA. The deamination frequency is the frequency of A-G transitions in PUM1/2-rABE at the indicated dox induction level over background A-G transitions in rABE alone at the corresponding dox level; A-G transitions result from A-to-I editing. Bar plots of deamination frequency are the average frequency across three replicates. Pum binding

elements (PBE) motifs are highlighted in bold. **(C)** Differential A-to-I editing rates from PUM1/2-rABE compared to rABE-Flag untagged control under 1000 ng/mL doxycycline induction. Significant A-to-I editing events ($p_{adj} < 0.05$, Cochran-Mantel-Haenszel test, two-sided, Benjamini-Hochberg adjusted) are marked in red. **(D)** Transcriptome-wide distribution of significant A-to-I editing events introduced by PUM1-rABE or PUM2-rABE. As some editing sites are annotated in multiple regions in different transcript isoforms, the sum of the percentage of four categories is greater than 100%. **(E)** Motif enrichment in the +/− 50 nt region around A-to-I editing sites identifies the canonical Pumilio binding element (UGUANA). **(F)** CentriMo analysis around PUM1-rABE A-to-I editing sites. The PRE motif was enriched at +15 nt downstream of A-to-I editing sites. **(G)** Expression level of PUM1 or PUM2 measured by RNA-seq, under 0, 50, 1000 ng/mL doxycycline induction. Both endogenous PUM2 and PUM2-rABE showed lower expression level than PUM1s. Error bars are 95% CI around mean. Error bars are 95% CI around mean. n=3 biologically independent samples. **(H)** Expression level of endogenous PUM1 and PUM1-rABE under 0, 50, 100, 1000 ng/ ul doxycycline induction, measured as a ratio of the western blot signal to actin. Error bars are 95% CI around mean. N=4 biologically independent samples. **(I)** A comparison of the editing sites identified at low and high dox for PUM1-rABE. **(J)** Higher doxycycline concentration led to higher editing rates as measured by RNA-seq. n=10793 for high dox and n=4362 for low dox; centre: median; box bounds: 25% and 75% percentile; whisker: 1.5x of interquartile range. All outliers are marked as diamonds. **(K, L)** Scatter plot in training set (K) and testing set (L) of predicted editing rate and measured editing rate of PUM1-rABE with a xgboost model. **(M)** Variables importance plot of PUM1-rABE editing rate xgboost model. Boxplot shows the distribution of variable importance in 10 permutations, bar length: mean; box bounds: 25% and 75% percentile; whisker: 1.5x of interquartile range.

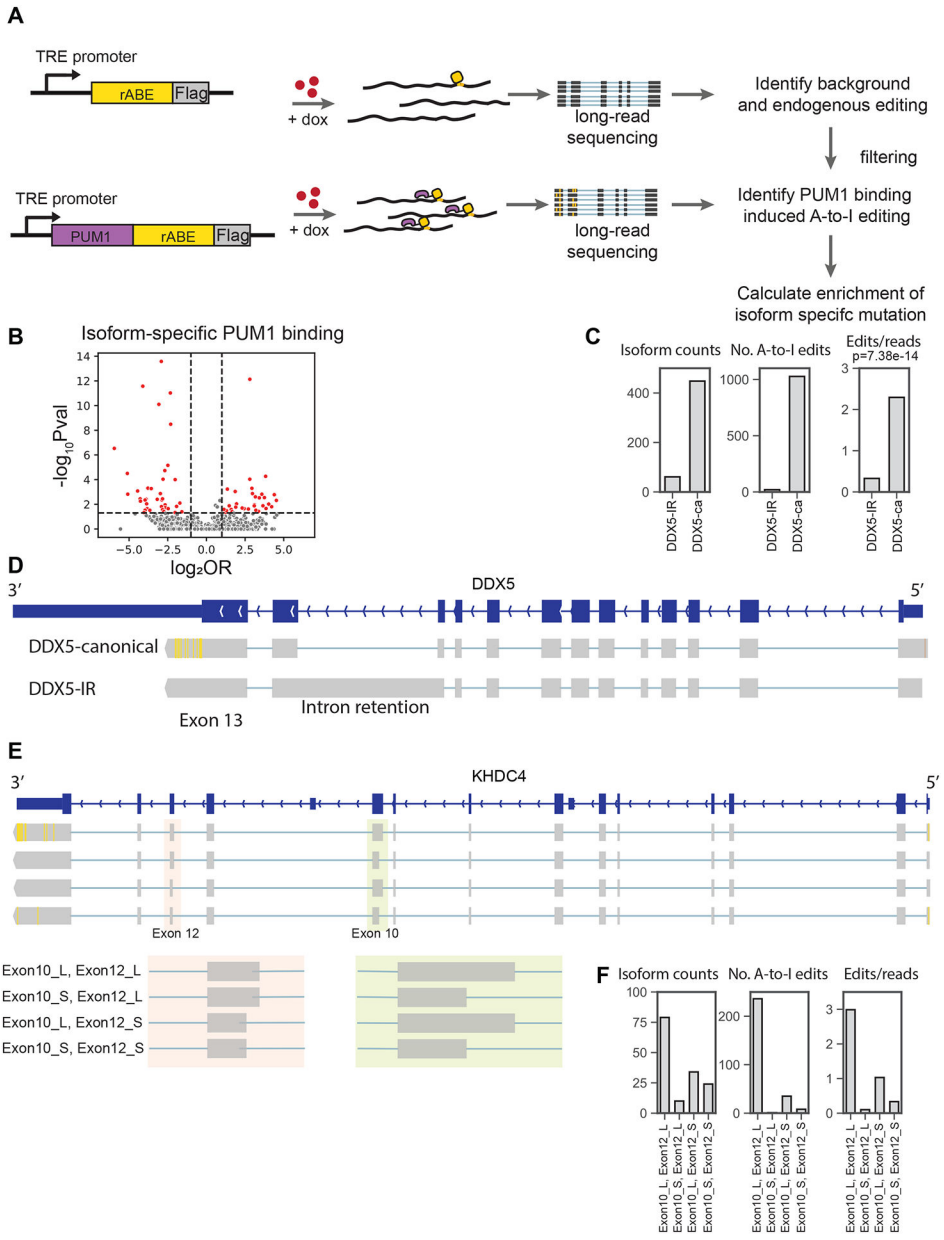


Figure 5:

Pacbio long-read sequencing identified isoform specific binding events of PUM1. **(A)** RNA samples from stable HEK293T cells lines expressing PUM1-rABE as described in Figure 4A were also used for Pacbio long-read sequencing. PUM1-rABE induced A-to-I editing were determined after filtering out locations with mutation in rABE-Flag samples. **(B)** Volcano plot of transcriptome-wide significant isoform specific PUM1 binding events. Isoform specific PUM1 binding events were determined by calculating fold enrichment of isoform specific mutations. ($p_{adj} < 0.05$, Mann-Whitney U test, two-sided, Benjamini-Hochberg adjusted). **(C)** Number of canonical DDX5 isoform and intron retention isoform (DDX5-IR) Pacbio reads, number of A-to-I editing events and number of edits per read in each isoform. P value calculated using Mann–Whitney U test, two-sided DDX5-IR: intron

retention isoform; DDX5-ca: canonical isoform **(D)** Transcript structure of the canonical DDX5 isoform (DDX5-ca) and intron retention isoform (DDX5-IR). **(E)** scheme of KHDC4 canonical isoform (Exon10_L, Exon12_L), and isoforms with alternative splicing in exon10, exon12, or both. **(F)** Number of KHDC4 isoforms Pacbio reads, number of A-to-I editing events and number of edits per read in each isoform.

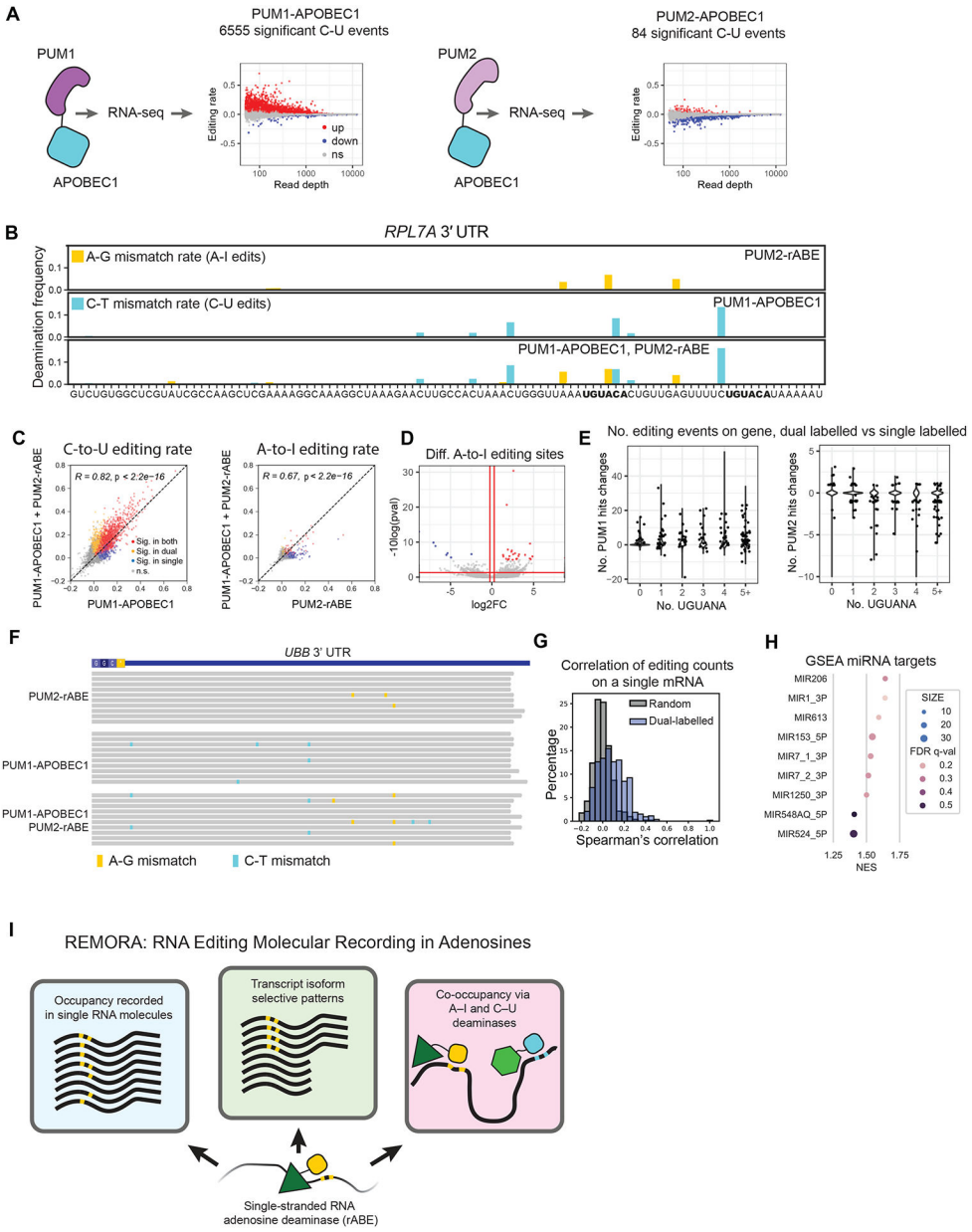


Figure 6.

Co-occupancy of PUM1 and PUM2 on RNA using rABE and APOBEC1 fusions. **(A)** Transcriptome-wide C-to-U editing events introduced by PUM1-APOBEC1 or PUM2-APOBEC1. Significant editing events ($p_{adj} < 0.05$, Cochran-Mantel-Haenszel test, two-sided, Benjamini-Hochberg adjusted)) are marked in red. **(B)** PUM1-APOBEC1 or PUM2-APOBEC1 induced editing patterns in the 3′ UTR region of RPL7A around Pumilio binding motifs. **(C)** Scatter plots showed correlation of editing rates of PUM1-APOBEC1 and PUM2-rABE between single labeling (individual expression) and dual labeling (co-expression) HEK 293T cells (t-test, two-sided.) **(D)** Volcano plot showing sites with differential editing rate in dual labelled cells and PUM2-rABE single labelled cells. We identified more sites with increased editing rate in dual labelled cells (red) than decreased

editing rates (blue). ($p_{adj} < 0.05$, Cochran-Mantel-Haenszel test, two-sided, Benjamini-Hochberg adjusted)**(E)** Number of significant A-to-I editing events increase in dual labeling cells compared to PUM1-APOBEC1 single labelled cells and PUM2-rABE single labelled cells across genes with different number of PRE motifs. Genes with higher number if increased PUM2 binding are enriched in no PRE motif group. All genes with changes of hits ≠ 0 are shown as dots. **(F)** Example of UBB 3′ UTR showing dual labeling of PUM1-APOBEC1 and PUM2-rABE on single RNA molecules using PacBio long-reads. **(G)** Histogram of correlation scores of numbers of PUM1-APOBEC1 edits and PUM2-rABE edits in PacBio long-reads. The majority of genes showed positive correlation, suggesting co-binding between PUM1 and PUM2, while random shuffle background from single labelled cells showed correlation scores around zero. **(H)** GSEA analysis of genes with positive correlation of PUM1 and PUM2 binding showed they are enriched in microRNA targeted genes. **(I)** A model of REMORA. Adenosine deamination with rABE enables high fidelity occupancy recording with transcript isoform and single-molecule resolution. When combined with a C-to-U deaminase such as APOBEC1, co-occupancy and competition between RNA-binding proteins in cells can be measured.