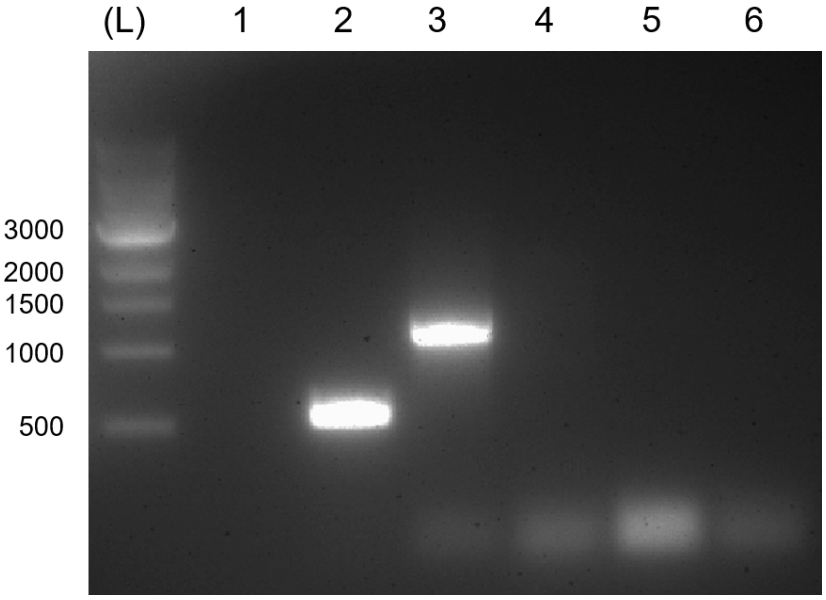
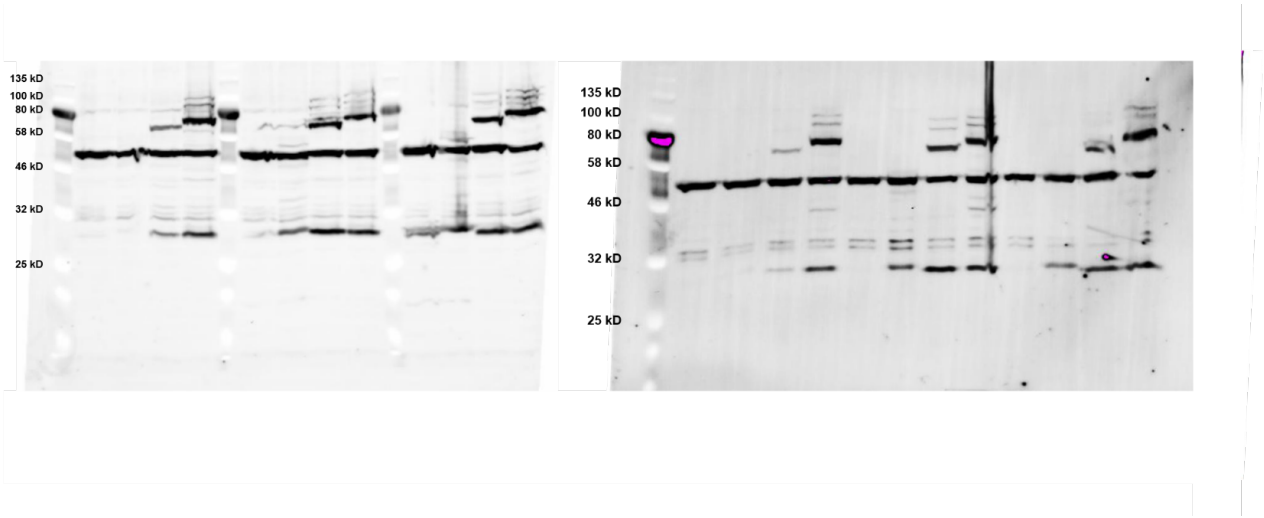
Supplementary Information

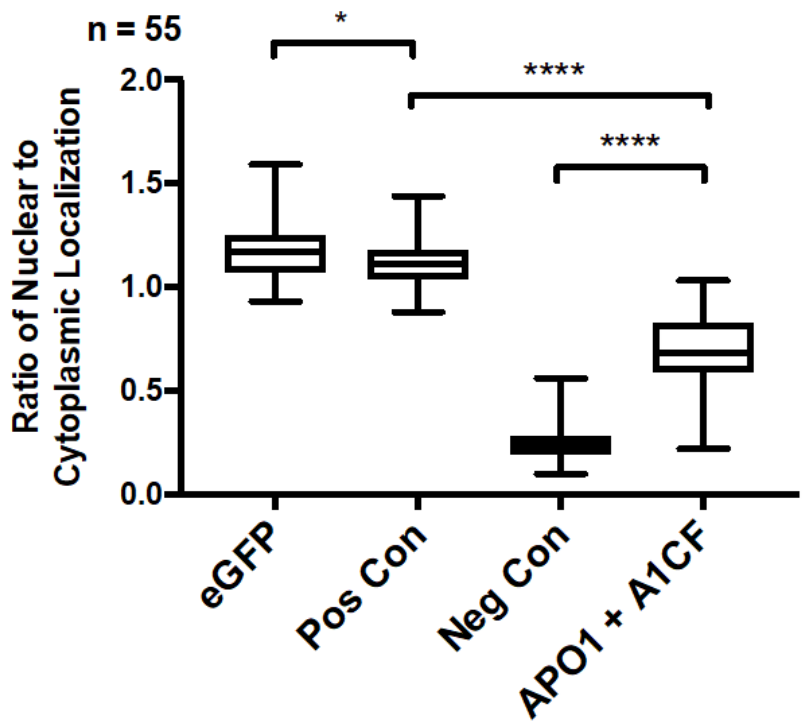
Comparison of RNA editing activity of APOBEC1-A1CF and APOBEC1-RBM47 complexes reconstituted in HEK293T cells



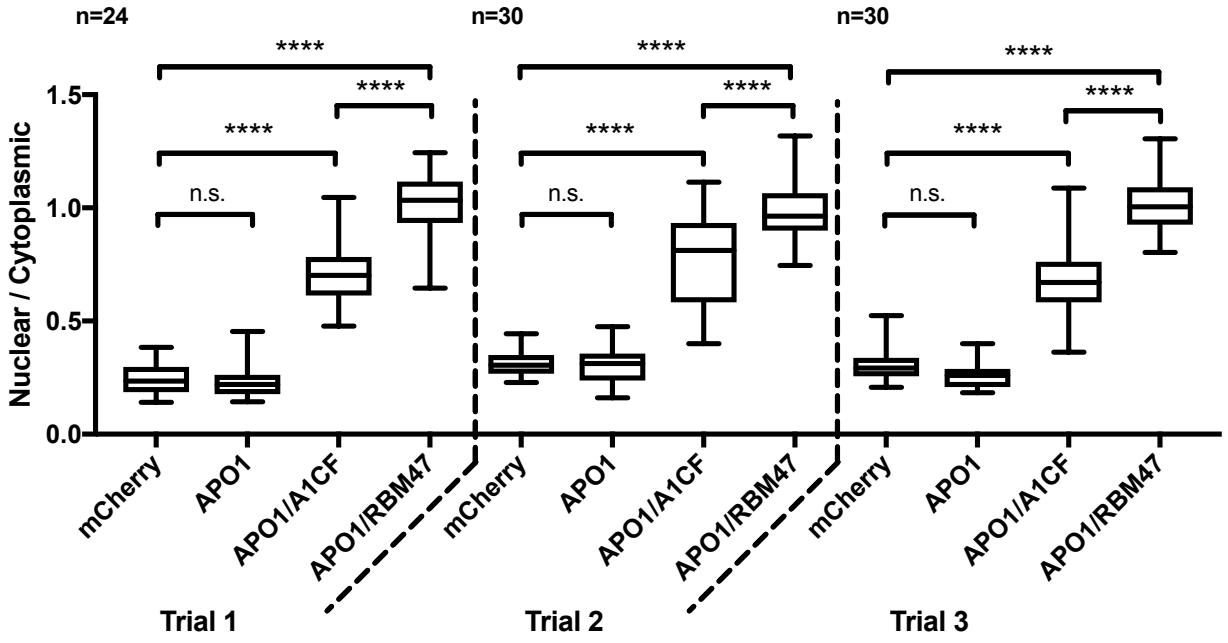
Supplemental Figure 1: RT-PCR analysis of un-transfected HEK293T cells. After RT of the total extracted mRNA (using an anchored poly-T primer), primers for detection of transcripts for *E1A* (lane 2: 550 nt predicted), *β-actin* (lane 3: 1120 nt predicted), *RBM47* (lane 4: 1785 nt predicted), *A1CF* (lane 5: 1785 nt predicted), and *APO1* (lane 6: 770 nt predicted) were used for amplification from the total cDNA. Lane 1 is a negative control of cDNA alone. Lanes 2 and 3 were considered positive controls for genomic proteins of HEK293T cells.



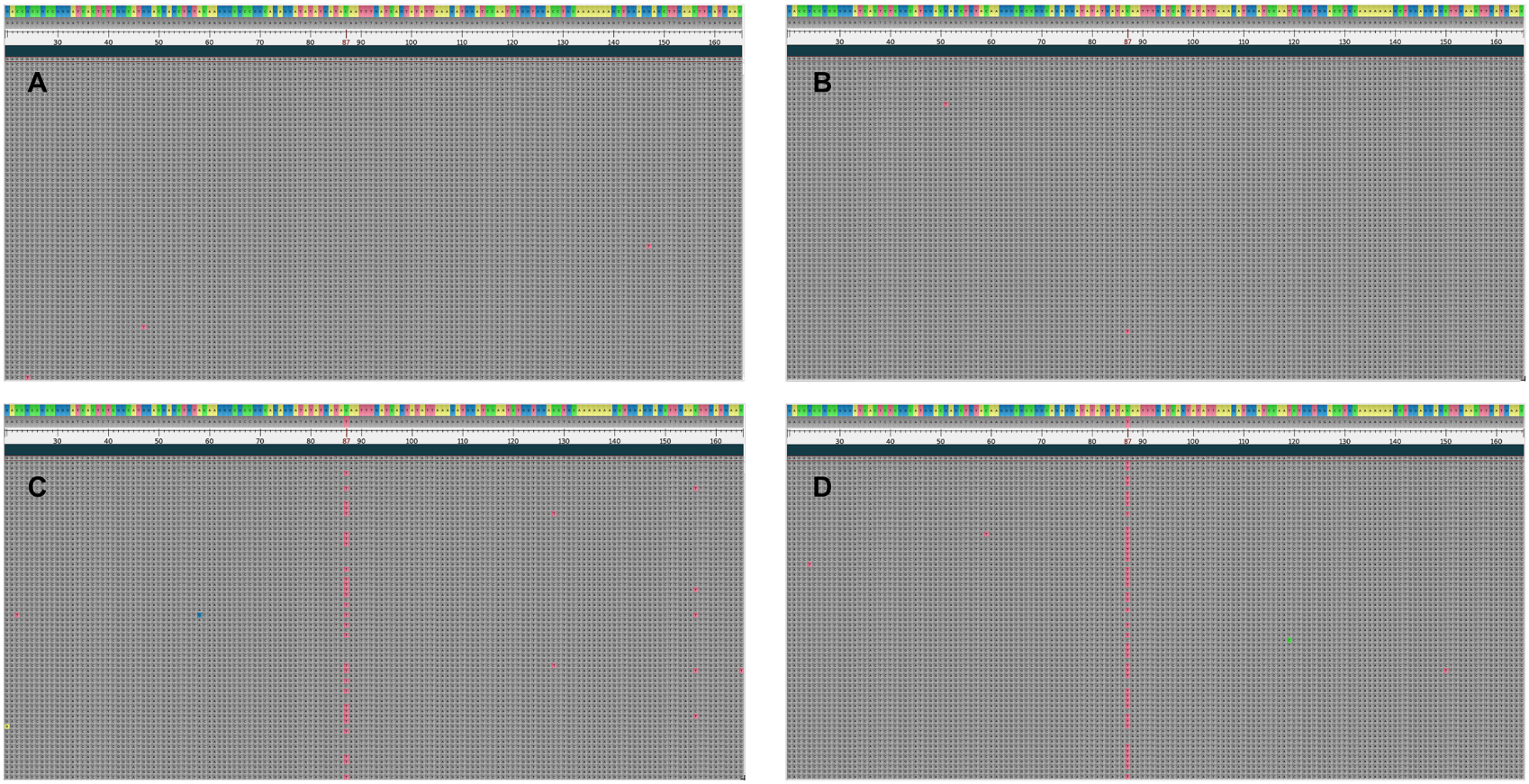
Supplemental Figure 2: Uncropped image of the western shown in Figure 1C. Two gels showing three repeats of the western shown in Figure 1C**.** Each repeat contains four samples: the eGFP reporter was co-expressed in the presence of either mCherry alone (not stained on blot), FLAG-APO1 (predicted 31 kDa) + mCherry, FLAG-APO1 + HA-A1CF (predicted 69.2 kDA) + mCherry, or FLAG-APO1 + HA-RBM47 (predicted 67.9 kDA) + mCherry. Despite being the smaller predicted molecular weight, RBM47 appears to consistently run a bit higher on SDS-page. The same set of 12 samples was loaded on both gels, and the second gel was an attempt to get a cleaner image. Figure 1C was cropped from the center set of the rightmost gel. α-tubulin is included as a loading control and runs at approximately 50 kDa.



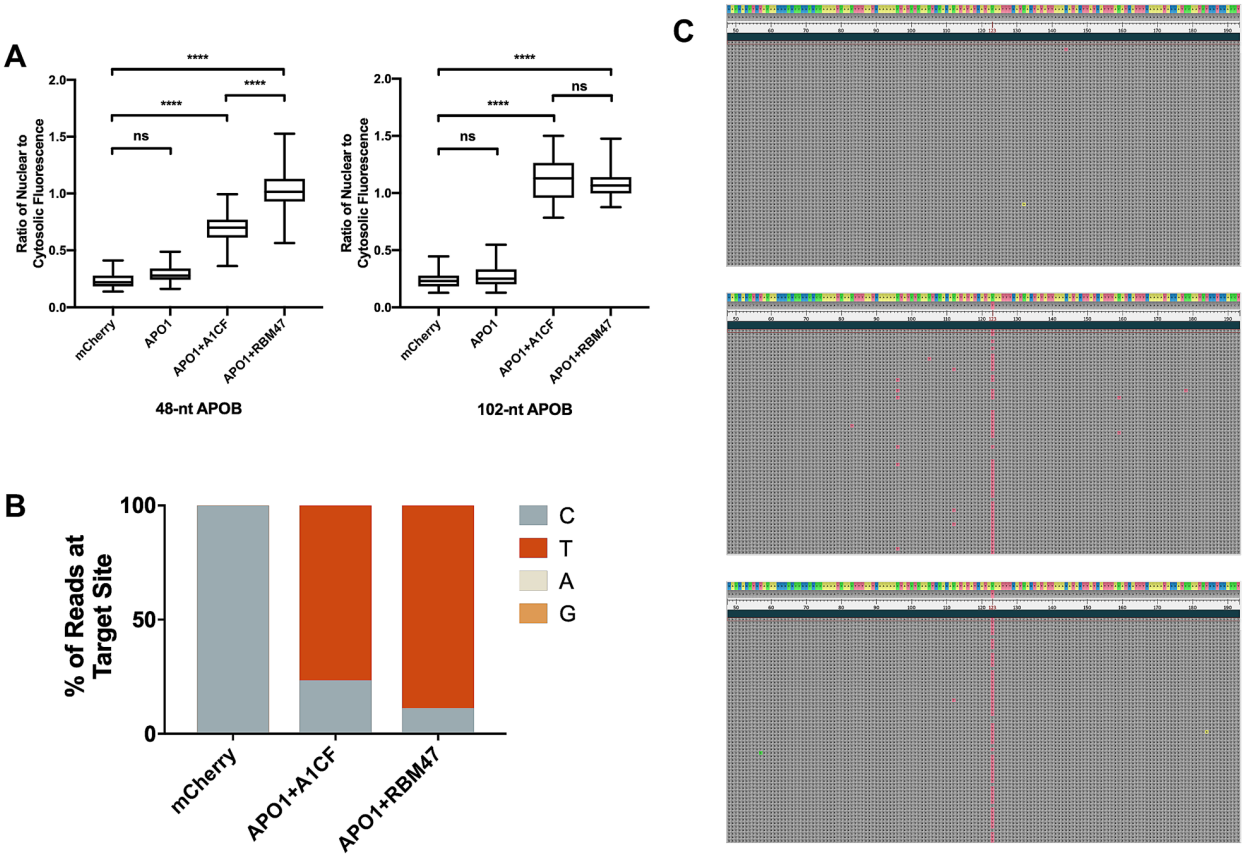
Supplemental Figure 3: An initial assessment of the reporter and editor system to confirm the relative sensitivity of the assay. The fluorescence localization ratios of wild-type eGFP was compared to the reporter coexpressed with mCherry alone but with the target cytosine mutated mutated to thymine, coexpressed with mCherry alone, and coexpressed with the editor construct. These respectively acted as a positive control (a maximum baseline for the observed localization ratio), a negative control (giving a minimum baseline for the localization ratio) and a test condition (showing the editor results in a localization change that is statistically different from both controls). The positive control was also found to be significantly different from eGFP alone (multiplicity-adjusted P = 0.0308) but this was not considered a detrimental characteristic of the tag.



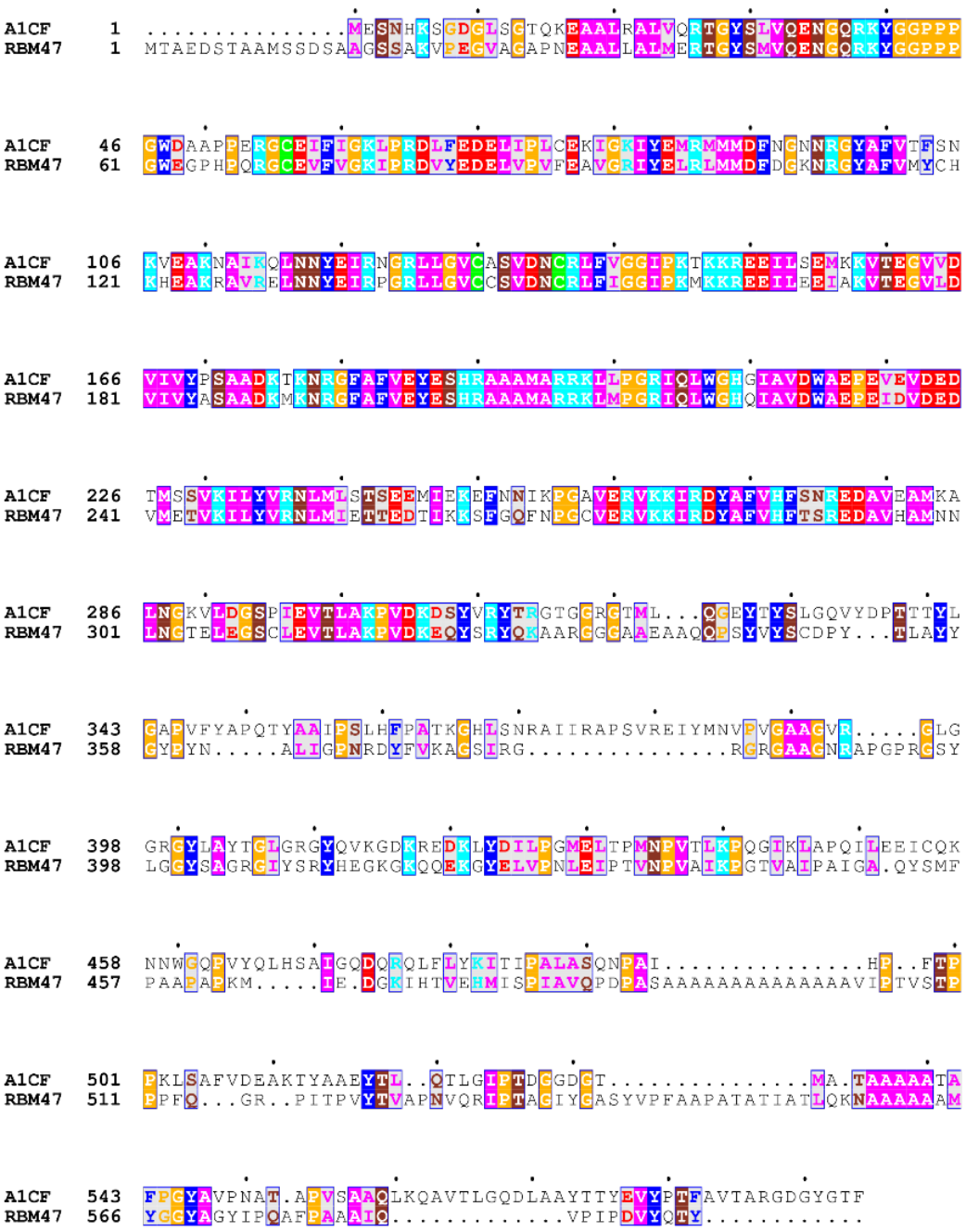
Supplemental Figure 4: Three independent trials comparing ApoB editing when cotransfected with mCherry alone, APO1 alone, APO1+A1CF, or APO1+RBM47 show the same consistent trend. These separate experiments were to confirm results are reproducible when a fresh batch of cells are transfected, ensuring results can be reliably compared across experiments if needed. RBM47 always shows increased fluorescence localization changes compared to A1CF ($P < 0.0001$). Significance was calculated as described elsewhere in this report. Some variation in the standard deviation across trials was observed, but this variance had no effect on the statistical analyses and could likely be mediated by analyzing greater numbers of cells.



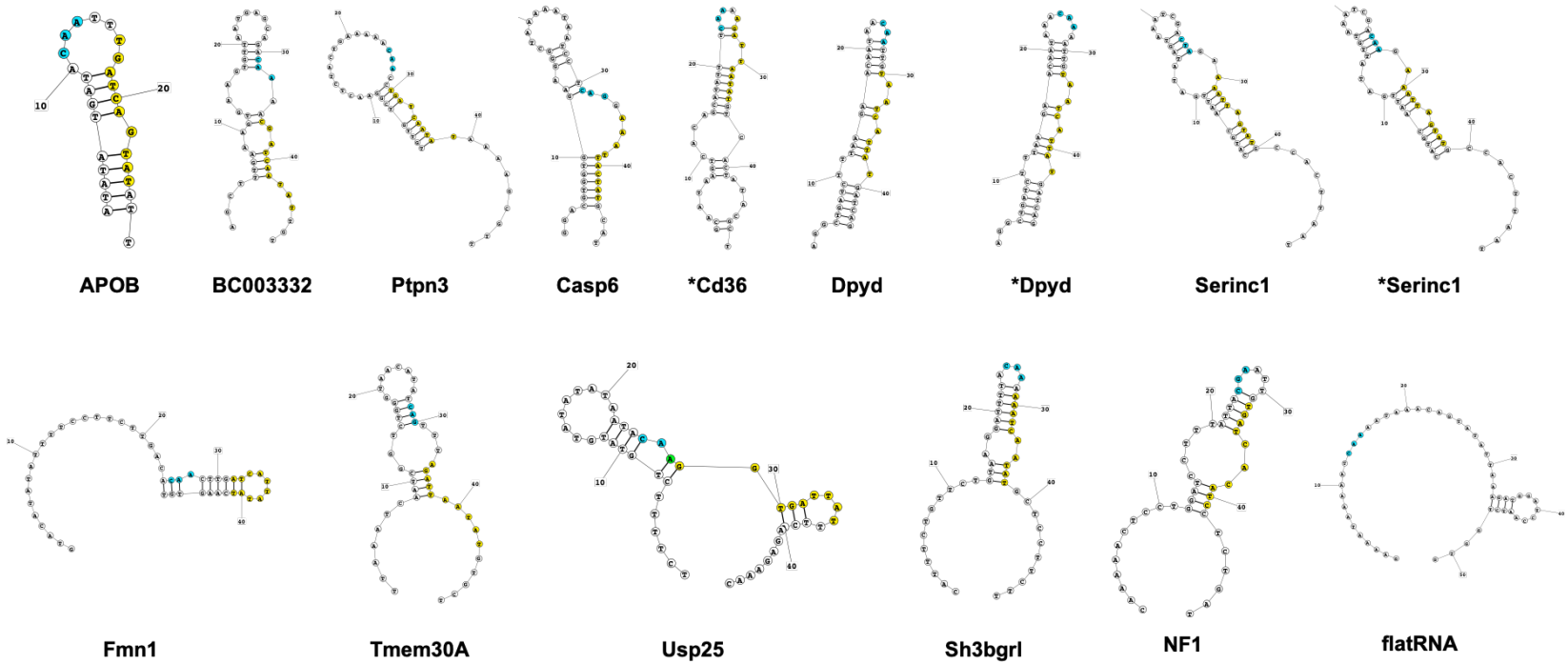
Supplemental Figure 5: Example of a next-gen sequencing alignments made in UGENE [42], showing a small fraction of the total reads sequenced. The four samples shown are the *APOB* reporter co-expressed with **(A)** mCherry alone, **(B)** APO1 alone, **(C)** APO1+A1CF, and **(D)** APO1+RBM47. In the amplified segment the edited residue is located at position 87. The number of A, T, C, and G residues at this position were calculated as a percentage of all reads and the results for each of the four co-expression experiments (mCherry, APO1 alone, APO1+A1CF, and APO1+RBM47) are plotted in Figure 2C. Importantly, none of the other positions within the sequenced region showed any significant degree of editing activity.



Supplemental Figure 6: Effect of longer substrate insertion lengths of strength of editing activity. Using an *APOB* substrate of either 48- or 102-nucleotides long (inserted into the reporter construct) does not change the overall trend of RNA editing but can increase the efficiency. **(A)** The 48-nt substrate showed no apparent difference in editing efficiency when compared to the original 27-nt substrate (see Supplemental Figure 4). The 102-nt substrate showed much stronger editing for APO1+A1CF, with no statistical significance ($P > 0.05$) from what was observed for APO1+RBM47. However, NGS showed that RBM47 is still the preferred cofactor for this editing **(B,C)**, suggesting instead that the 102-nt editing for both samples has resulted in a fluorescence shift that exceeds the dynamic range of this assay. Notably, the off-target editing in the presence of A1CF has seemingly increased when the 102-nt substrate is used, with the second-most efficient editing occurring 27-nt upstream of the edited site (residue 96 in the sequence here or base 25 in the isolated segment as shown in Supplemental Figure 9).



Supplemental Figure 7: Pairwise sequence alignment of human A1CF vs RBM47 [45, 46], showing a high degree of identity between the two sequences up to around residue 318 on RBM47, which defines the end of the third RRM domain of both proteins. On A1CF, residue 391 was found to be the minimum C-terminal deletion that still had efficient RNA editing complementation. By comparison, Residue 406 of RBM47 was found to be roughly around the cutoff point for the minimum necessary regions needed for complementation activity.

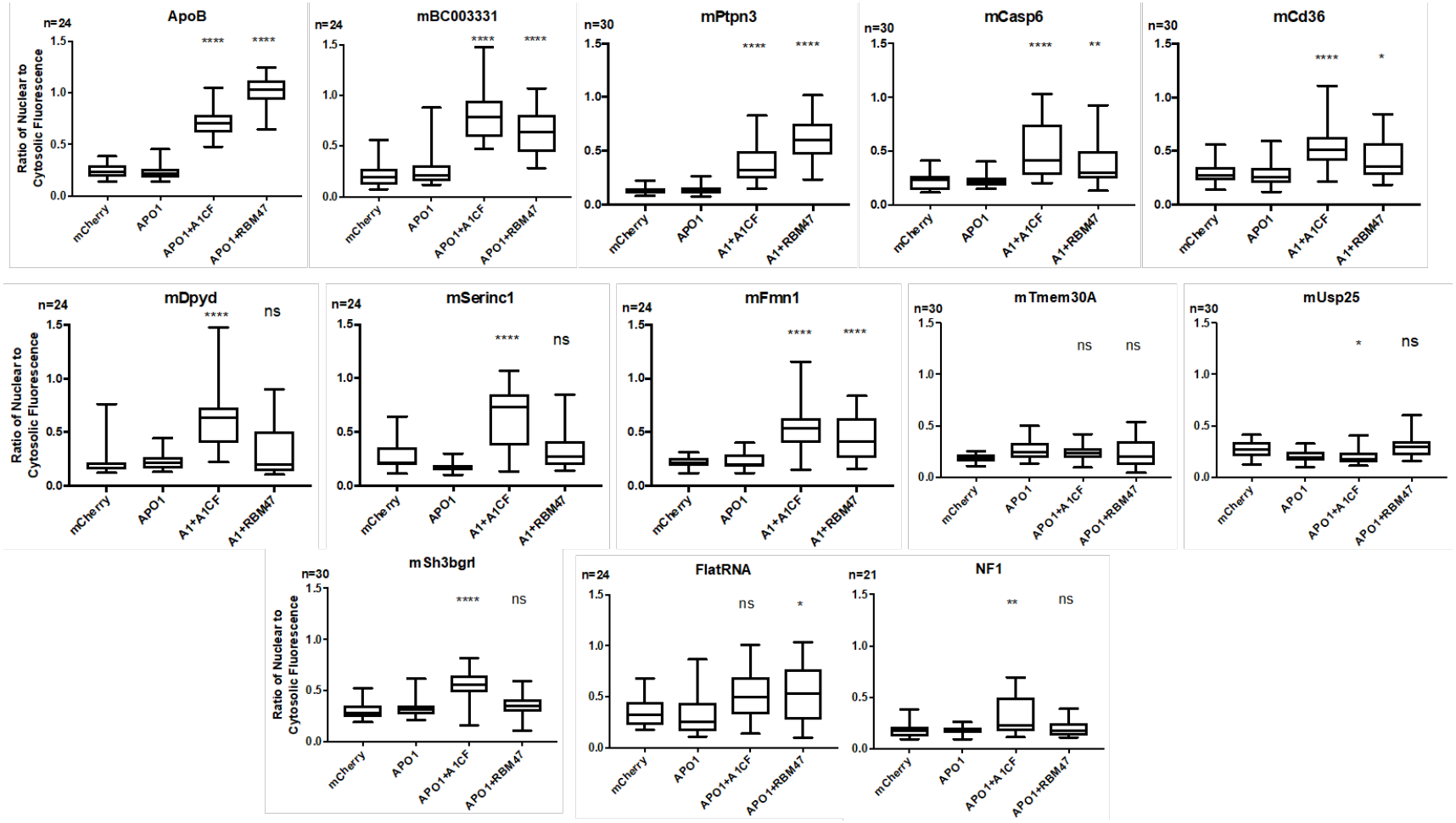


Supplemental Figure 8: Predicted RNA secondary structures as calculated by RNAstructure [35, 36] for the local RNA region as described elsewhere in this report. The edited cytosine and related codon are shown highlighted in blue; the consensus mooring sequence for each substrate is highlighted in yellow. *APOB* consists of a canonical stem loop as previously reported in literature. Interestingly, *Ptpn3* shows a comparable canonical stem loop and similarly favored editing by RBM47. Structures marked with an asterisk (*) are the modified sequences used in the actual experiment as displayed in Supplemental Figure 8, included to show intentional consistency in predicted structures when changes had to be made to the sequence.

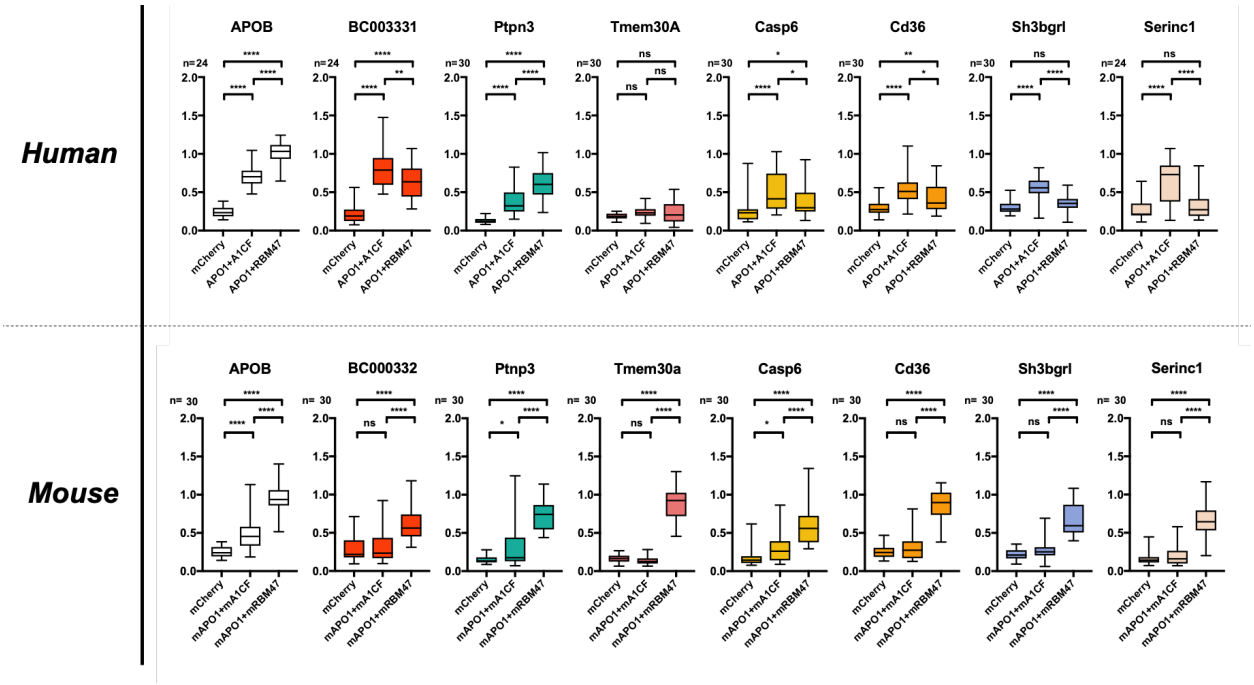
Target codon = blue | Mooring sequence = yellow | Altered from endogenous transcript = red

		EXPLANATION
APOB	ATATATGATACAATTTGATCAGTATATT	
NM_000384.2	ATATATGATACAATTTGATCAGTATATT	
APOB-48	TCTCAACTGCAGACATATATGATACAATTTGATCAGTATATTAAAGAT	
NM_000384.2	TCTCAACTGCAGACATATATGATACAATTTGATCAGTATATTAAAGAT	
APOB-102	GCCAAAATCAACTTTAATGAAAAACTATCTCAACTGCAGACATATATGATACAATTTGATCAGTATATTAAAGATAGTTATGATTTACATGATTTGAAAATA	
NM_000384.2	GCCAAAATCAACTTTAATGAAAAACTATCTCAACTGCAGACATATATGATACAATTTGATCAGTATATTAAAGATAGTTATGATTTACATGATTTGAAAATA	
BC003332	AGCTTTGAAAGTGAAGTGTTAATGAGCAGACAAAACGATCAATATTGT	
NM_145511.2	AGCTTTGAAAGTGAAGTGTTAATGAGCAGACAAAACGATCAATATTGT	
Ptpn3	TGTTGTCGGAACTCTACTGAAAAACAACCTGATCAATATAAAAGCGTT	
NM_011207.2	TGTTGTCGGAACTCTACTGAAAAACAACCTGATCAATATAAAAGCGTT	
Casp6	GGACGTGGTGGAAGGGCTAAAAATAATCCTCAGGAAATTACTATGCAT	
NM_009811.4	GGACGTGGTGGAAGGGCTAAAAATAATCCTCAGGAAATTACTATGCAT	
Cd36	GCAATAAGTCACAGCATATTTCAAAAGATTAATATGTCACTATACGCT	
NM_001159556.1	GCAATAAGTCACAGCATATTTCAAAAGATTAATATGTCACTATAGGCA	Avoids stop codon
Dpyd	AGGCTGATCTTTAAGAACAATAAACAAAATTGTAATCATTATGATCAG	
NM_170778.3	AGGCTGATCTTTAAGAACAATA-ACAA--TTGTAATCATTATGATCAG	Target codon reading frame had too many stop codons
Serinc1	CATGCAATTGATATTGTAAATCGACAAG-AAATTAGTATGCCACTTAAT	
NM_019760.3	CATGCAATTGATATAGTAAATCGACTAGAAAATTAGTATGCCACTTAAT	Avoid stop codon, L→Q to make stop with editing, and needed frame shift in spacer
Fmn1	GTACATATATTTTCCTTCTTGACACAACTTGATCATTATATCAAGTGT	
NM_001285458.1	GTACATATATTTTCCTTCTTGACACTACTTGATCATTATATCAAGTGT	Needed L→Q transition
Tmem30A	TTAAATCAATCGGTCTGGGTAACATATCAGTTTAGATTAATATGTGCT	
NM_133718.4	TTAAATCAATCGGTCTGGGTAACATATCAGTTTAGATTAATATGTGCT	
Usp25	TCTTTTCTGTATGTATATATAATACAAGGTGATTATTTCAAGAGAAAC	
NM_013918.2	TCTTTTCTGTATGTATAAATAATACAAGGTGATTATTTCAAGAGAAAC	Avoids stop codon
Sh3bgrl	CATTTCTGTTCTGTAAGGATTTTACAAAAAATCAATATGCTCCTTCTT	
NM_019989.2	CATTTCTGTTCTGTAAGGATTTTACAAAAAATCAATATGCTCCTTCTT	
NF1	CAAAAACTCCTGGATCCTTTATTACGAATTGTGATCACATCCTCTGAT	
NM_000267.3	CAAAAACTCCTGGATCCTTTATTACGAATTGTGATCACATCCTCTGAT	
flatRNA	AAATAAAAATCAAAATAAACAGTATATT	
ApoB	ATATATGATACAATTTCATCAGTATATT	

Supplemental Figure 9: Sequences of all the tested substrates, with changes from the wild-type sequences made to account for any protein incompatibilities highlighted in red. The edited cytosine codon and the canonical mooring sequence are highlighted in blue and yellow respectively. The flatRNA substrate was engineered based on mutations made to *APOB* and is shown accordingly.



Supplemental Figure 10: Complete box-and-whisker plots for all alternative substrates listed in Figure 3B. Significance levels are defined as when the multiplicity-adjusted P values from a one-way ANOVA with Tukey post-test fall within the range of n.s. if P > 0.05, * if P < 0.05, ** if P < 0.01, *** if P < 0.001, and **** if P < 0.0001. Shown here are the significance levels when comparing the means of APO1+A1CF or APO1+RBM47 to the mean of mCherry alone as a negative control.



Supplemental Figure 11. The side-by-side display of the editing efficiency of the tested RNA substrates by APO1/A1CF and APO1/RBM47 from human and mAPO1/mA1CF and mAPO1/mRBM47. Results are adapted from Figures 3,4 and Supplemental Figure 10. Significance are displayed as n.s. if P > 0.05, * if P < 0.05, ** if P < 0.01, *** if P < 0.001, and **** if P < 0.0001.