Title

Library Screening Reveals Sequence Motifs That Enable ADAR2 Editing at Recalcitrant Sites

Journal

ACS Chemical Biology, 18(10)

ISSN

1554-8929

Authors

Jacobsen, Casey S

Salvador, Prince

Yung, John F

et al.

Publication Date

2023-10-20

DOI

10.1021/acschembio.3c00107

Peer reviewed

HHS Public Access

Author manuscript

ACS Chem Biol. Author manuscript; available in PMC 2023 October 23.

Published in final edited form as:

ACS Chem Biol. 2023 October 20; 18(10): 2188–2199. doi:10.1021/acschembio.3c00107.

Library Screening Reveals Sequence Motifs that Enable ADAR2 Editing at Recalcitrant Sites

Casey S. Jacobsen[1], **Prince Salvador**[1], **John F. Yung**[2], **Sabrina Kragness**[2], **Herra G. Mendoza**[1], **Gail Mandel**[2], **Peter A. Beal**[1,*]

[1]Department of Chemistry, University of California, Davis, Davis, CA, USA, 95616.

[2]Vollum Institute, Oregon Health and Science University, Portland, OR, USA, 97239.

Abstract

Adenosine Deaminases acting on RNA (ADARs) catalyze the hydrolytic deamination of adenosine to inosine in duplex RNA. The inosine product preferentially base pairs with cytidine resulting in an effective A-to-G edit in RNA. ADAR editing can result in a recoding event alongside other alterations to RNA function. A consequence of ADARs' selective activity on duplex RNA is that guide RNAs (gRNAs) can be designed to target an adenosine of interest and promote a desired recoding event. One of ADAR's main limitations is its preference to edit adenosines with specific 5' and 3' nearest neighbor nucleotides (e.g. 5' U, 3' G). Current rational design approaches are well-suited for this ideal sequence context, but limited when applied to difficult-to-edit sites. Here we describe a strategy for the *in vitro* evaluation of very large libraries of ADAR substrates (En Masse Evaluation of RNA Guides, EMERGe). EMERGe allows for a comprehensive screening of ADAR substrate RNAs that complements current design approaches. We used this approach to identify sequence motifs for gRNAs that enable editing in otherwise difficult-to-edit target sites. A guide RNA bearing one of these sequence motifs enabled the cellular repair of a premature termination codon arising from mutation of the *MECP2* gene associated with Rett Syndrome. EMERGe provides an advancement in screening that not only allows for novel gRNA design, but also furthers our understanding of ADARs' specific RNA-protein interactions.

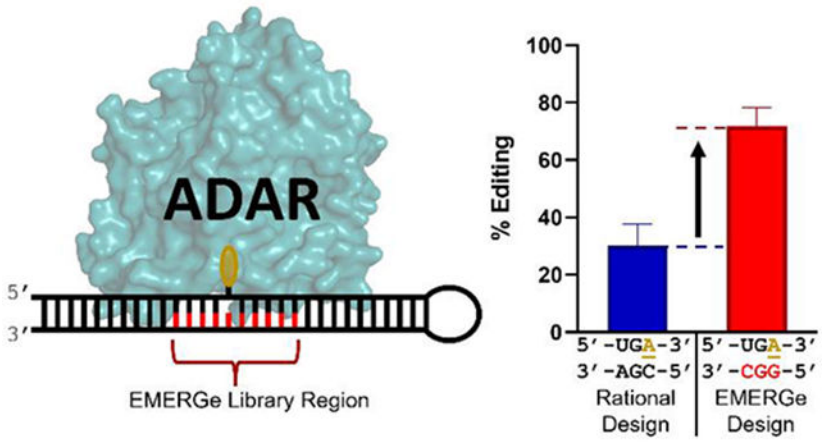
Graphical Abstract

*Corresponding author pabeal@ucdavis.edu.

Author Contributions

P.J.S. conducted all deaminations for the MECP2 R255X target and purified hADAR2. J.F.Y and S.M. completed cellular directed editing experiments. H.G.M. purified hADAR2-RD WT and hADAR2-RD E488Q. P.A.B. and C.S.J. designed the EMERGe methodology and C.S.J conducted all other experiments. C.S.J., G.M. and P.A.B. wrote and edited the manuscript.

P.A.B. is a consultant and holds stocks and stock options in ProQR Therapeutics and Beam Therapeutics, companies developing therapeutic editing technologies. G.M. is cofounder and stockholder of Vico Therapeutics, a company studying potential therapeutic use of editing approaches for neurological disease.



Keywords

RNA editing; ADAR2; adenosine deaminase acting on RNA; therapeutic RNA

Introduction

Adenosine Deaminases acting on RNA (ADARs) carryout the catalytic deamination of adenosine (A) in duplex RNA generating inosine (I) at the site of reaction[1,2,3,4]. Because the A-to-I conversion changes the Watson-Crick hydrogen bonding specificity of the base, the consequences on RNA function are wide-ranging and include changing the meaning of specific codons (recoding). This form of RNA editing is widespread in higher organisms with hundreds of thousands of A to I sites identified in the human transcriptome[5-9]. The efficiency of editing varies dramatically across different ADAR substrate RNAs from nearly quantitative to barely detectable above assay noise[6,10]. Defining the features of a target RNA that distinguish it as a good substrate has been a long-standing challenge in the ADAR field[11-18]. We know that ADARs edit efficiently within base paired, A-form duplexes and have 5’ and 3’ nearest neighbor nucleotide preferences, yet natural editing sites exist within RNA structures that contain helix defects (e.g. base mismatches, bulges, internal loops, *etc.*) and do not have the optimal nearest neighbor nucleotides. Clearly other features of a substrate RNA beyond base pairing and 5’ and 3’ nearest neighbor nucleotides allow for efficient interaction with the enzyme such that ADAR can overcome the effect of a sub-optimal sequence context or an imperfect duplex. Furthermore, evidence exists in the literature that suggests that base mismatches distal to the editing site can be beneficial for editing[17]. This has been rationalized by suggesting the duplex imperfections limit the number of double stranded RNA binding domain (dsRBD) binding sites on the substrate RNAs to only those that promote editing at the target adenosine[19]. Defining how different RNA structures can facilitate ADAR editing advances our basic understanding of the ADAR reaction and can inform the design of ADAR guide strands for maximum efficiency of directed RNA editing for therapeutic applications. We previously described a phenotypic screen carried out in *S. cerevisiae* of libraries of hairpin substrates for ADARs where the editing site was positioned within a 5’-UAG stop codon upstream of coding sequence for α-galactosidase[20]. Sequences that supported efficient ADAR editing led to reporter enzyme

expression and colored yeast colonies. These early efforts illustrated the importance of the editing site complementary sequence (ECS) and an adjacent hairpin loop sequence[21]. However, the need to transform yeast cells and visualize individual colonies limited the size of libraries that could be practically screened. Here, we describe a variation on this screening approach that does not suffer from these limitations and is amenable to the screening of very large libraries (*e.g.* 10^5-10^6 sequences in a single experiment). Using this approach with human ADAR2, we have identified unique sequence motifs that enable editing within challenging sequence contexts including for the 5'-G<u>A</u>A triplet that bears suboptimal 5' and 3' nearest neighbor nucleotides[22].

Results

Development and testing the EMERGe workflow

Our lab had previously described screening methods, such as an X-Gal screen[20] and Sat-FACS-Seq[23], to probe combinations of mutant ADARs and substrate RNA libraries in a high throughput manner. As a logical extrapolation from those approaches, we developed a new screening assay, En Masse Evaluation of RNA Guides (EMERGe), to identify guide RNA (gRNA) sequences that enable ADAR editing at difficult-to-edit sites. EMERGe uses a hairpin RNA substrate with the target editing site covalently linked to a sequence-randomized guide region (Figure 1A). The target RNA is designed to fold into a nearly complementary duplex structure that contains three distinct 10 nt regions: (1) a fully complementary 10 bp region 5' to the edit site; (2) a 10 nt region which contains the edit site and 10 variable nucleotides (N_{10}) across from the edit site; and (3) a fully complementary 10 bp region 3' to the edit site (Figure 1A). Additionally, a loop linker 3' to the edit site allows for hairpin formation. The hairpin sequence can be varied to allow for the discovery of enabling guide strands for different therapeutically-relevant targets, such as the MECP2 R168X mutation that causes the neurodevelopmental disorder Rett Syndrome[24] (Figure 1B). Importantly, the R168X mutation creates a UGA premature termination codon in the *MECP2* transcript that could be converted to a codon for tryptophan (UGI) by deamination of the adenosine[24,25]. However, the adjacent 5' G makes this a difficult-to-edit site for ADARs[22]. An N_{10} sequence-randomized region in the RNA library allows for ~10^6 possible guide sequences to be queried through the EMERGe screen simultaneously. The RNA hairpin library is subjected to an ADAR-catalyzed deamination reaction. The resulting products are converted to DNA through reverse transcription PCR (RT-PCR) and sequenced by Next Generation Sequencing (NGS). The sequencing data sets are processed using HTStream and subsequent NGS aligning to identify two key regions: (1) the editing site (A or G) and (2) the N_{10} region. These two variable regions are then used to calculate the % editing for each identified N_{10} region that enabled editing within the dataset. This provides a sortable dataset that is used to identify winning sequences that support efficient editing at the target site (Figure 1c). In order to minimize false positives, we applied a minimum read number cut-off, and then ordered the sequences based on the calculated % editing. For the R168X target substrate shown in Figure 1b, only sequences with 10 or more combined A and G reads in the dataset were considered further. This resulted in the selection of 30 top sequences (Table S1) that varied in % editing at the target site from 78% to 40%.

Importantly, when a no-ADAR control screen was carried out, none of the top 30 sequences showed editing above 0.1% (Table S2).

Given the large number of sequences tested at once in an EMERGe screen and with sufficient NGS read depth, the effect of single nucleotide variants of a top performing sequence can be readily ascertained without additional experiments. This is done by simply querying the data set for the single nucleotide variants of the sequence of interest and calculating the corresponding % editing for those variants. For instance, by completing a dataset look-up of single nucleotide variants of one of the 30 top sequences described above (R168X-5), we generated an activity heat map that shows how a single nucleotide change in the N_{10} region of R168X-5 affects editing efficiency (Figure 2). Through this, we found that specific nucleotides at certain positions are required for efficient editing. For instance, varying the nucleotides at the −2, −1, and the orphan positions opposite the target A from those found in R168X-5 is highly detrimental to editing (Figure 2). However, perturbations in the terminal regions of the selected sequence, particularly at the 5’ end, were not as detrimental.

We were particularly interested to know if any of the top 30 sequences from the *MECP2* R168X screen would support an *in vitro* ADAR2 reaction in a duplex formed *in trans* with an antisense oligonucleotide guide and longer target RNA as is typical in site directed RNA editing applications (Figure 1C and Table S1). Therefore, selected sequences were inserted into 30 nt antisense gRNA oligonucleotides to represent one strand of the hairpin from the EMERGe screen. The gRNA was then hybridized to a 300 nt target RNA bearing the *MECP2* R168X target site to form an approximately 30 bp dsRNA substrate mimic of the hairpin tested in the screen (Figure 3A). For comparison, a control guide (R168X-AC) was also generated based on known design principles for ADAR gRNAs. This control gRNA is fully complementary to the target except for an A:C pair at the edit site (Figure 3B)[22]. Six representative sequences (Figure 3C-H) were tested in an *in vitro* editing assay with ADAR2 and compared to the control guide. The sequence that showed the most effective editing (70% ±1%) in these studies was R168X-5 (Figure 3F). Importantly, this guide contains two features known to enable ADAR editing at 5’-G<u>A</u> sites: an A:C pair at the target adenosine and a G:G pair involving the 5’ G[26,27] (Figure 4A). In addition, the R168X-5 guide sequence contained two other notable features. First, it is predicted to form several G:U wobble pairs; and second, it bears an A:A mismatch at the 5’ end of the N_{10} region. To evaluate which features are necessary for editing, structure activity relationship (SAR) studies were completed by designing five new antisense guide strands with key differences from R168X-5 (Figure 4B-F). First, the G:G pair at the −1 position was replaced by a G:C pair. This guide, R168X-5a (Figure 4B) showed a reduction in editing in comparison to R168X-5, but editing was still greater than the control guide, highlighting the importance of the other selected features (*i.e.* the G:U wobbles and A:A mismatch). It is important to note here that the suffix “a” has been added to the name of the guide to denote a change from the EMERGe-derived sequence. Next, a guide was designed to evaluate the impact of converting the G:U wobbles to Watson-Crick pairs and the A:A mismatch to an A:U pair leaving only the G:G pair and A:C mismatch (R168X-5b, Figure 4C). This change did not reduce on-target editing, but it did increase bystander adenosine deamination within the guide-target RNA duplex. Finally, to evaluate the effect of removing the G:U wobble pairs and A:A

mismatch in different combinations, guides R168X-5c-e were designed (Figures 4D-F). Through deamination experiments with these three guides, we found that the A:A mismatch plays a role in reduction of off-target editing and that divergence from full complementarity within the 5' side of the N_{10} can be beneficial.

Enabling Editing in a 5'-GAA Context

One of the most challenging sites for ADAR editing is when the target A is within a 5'-GAA context[22]. This sequence has both suboptimal 5' and 3' nearest neighbor nucleotides. Therefore, applying an EMERGe screen to the 5'-GAA target sequence may provide insight into improving editing in this poorly edited sequence. Importantly, two common nonsense mutations in Rett Syndrome (R255X and R270X) generate premature termination codons where editing within a 5'-UGAA could restore a form of the full-length protein by converting the termination codons to codons for tryptophan. Therefore, we generated a new hairpin bearing the *MECP2* R255X sequence as the target. For this hairpin, we also replaced the GUAA loop present in the R168X hairpin with a longer and potentially more flexible loop sequence (Figure 5A). This new hairpin was subjected to an EMERGe screen and editing of the top 30 sequences was identified (Figure 5B). The dataset from this screen had a lower read depth per N_{10}, so the cut-off was lowered to three reads per sequence for consideration (Table S3). Among the top 30 sequences, a subset was tested independently for on-target and bystander editing *in trans* (Figure S1). In this subset, for example, R255X-11 led to the most specific and efficient on-target editing (Figure S1F). When R255X- 11 (Figure 5C) was compared directly to a guide containing the 5'-GA enabling G:G pair and A:C pair (Figure 5D), R255X-11 supported target site editing to a similar level (86 ± 12% vs 86 ± 6%) and reduced editing at two bystander sites from 20 ± 4% and 40 ± 9% to undetectable under the conditions of the assay. Understanding the exact mechanisms by which the bystander editing is reduced will require additional studies. However, the R255X-11 guide is predicted to pair with the target in a manner that causes the −4 bystander A to bulge out of the RNA duplex. Reducing bystander editing in this manner has recently been described by others.[13] Overall, these results again verify EMERGe's ability to identify N_{10} sequences that enable specific editing. Notable features of the predicted secondary structure for the R255X-11 guide-target RNA duplex include the single bulged A present in the target strand four nucleotides 5' of the target site (−4 bystander A) and a three nucleotide GUG loop opposite the target A and the G on its 5' side.

We also carried out an EMERGe screen with a hairpin bearing the *MECP2* R270X site (Figure 6a). To avoid the low read depth for the R255X dataset described above, the R270X library was subjected to NGS with an estimated output of 350M paired end reads. This deeper read dataset allowed for a larger pool of edited sequences to be evaluated and the cut-off to be raised to 20 reads per sequence (Figure S2). Again, the top 30 sequences from the screen were selected (Table S4) and sorted according to editing efficiencies (Figure 6B). Representative sequences from this group were used for follow up verification studies as described above (Figure S3). The selected sequence leading to the highest editing efficiency (64% ± 2%) in an antisense guide strand, R270X-24, is predicted to form Watson-Crick or G:U wobble pairs at six of the ten variable nucleotide positions. The remaining four nucleotides are part of a 4 nt x 4 nt internal loop including the editing site adenosine

(Figure 6C). Interestingly, analyzing the editing efficiency of single nucleotide variants of the R270X-24 hairpin along with two closely related hit sequences R270X-6 and R270X-7 indicated that a key element enabling editing is the 3'-CGG across from the 5'-UGA editing site (Figure S4). Therefore, the antisense oligonucleotide guide R270X-24a was designed to contain only that 3 nt motif while keeping the remainder of the duplex Watson-Crick base paired (Figure 6D). We saw a significant increase in on-target editing (from 30 ± 7% to 71 ± 7%) and a reduction in bystander editing at one site (from 46 ± 5% to 25 ± 8%) for the R270X-24a guide in comparison to the R270X-GG;AC control (Figure 6D-F and Figure S5).

To better understand how the 3'-CGG motif enables editing (Figure 7a), we analyzed the effect of single nucleotide mutations at each position in two different ways. First, we generated an activity heat map for variants of the R270X-24 hairpin present in the original screening data set (Figure 7B). Second, we synthesized 30 nt antisense oligonucleotide guides bearing single nucleotide changes at each position in the 3'-CGG motif found in R270X-24a, measured ADAR2 editing efficiencies on a target RNA bearing the R270X sequence and generated the corresponding activity heat map (Figure 7B). Remarkably, on one hand, both data sets show the 3' C of the 3'-CGG motif is essential for enabling editing at the target site. On the other hand, there is a tolerance for U at the center position and A at the 5' position of the 3'-CGG motif. Again, both data sets show the same results indicating that the effect of mutations within this motif is not dependent on the presence of the hairpin structure.

High resolution structures are available for ADAR2-RNA complexes that reveal details of the interactions between the protein and nucleotides around the editing site[29,30]. Indeed, our lab recently published a structure of ADAR2 bound to substrate RNA where the editing site is adjacent to a G:G pair[25]. From these structures, we can define the "orphan" base found in the guide strand opposite the editing site. This nucleotide makes multiple contacts to the protein including direct hydrogen bonding to the side chain of the amino acid at position 488 of hADAR2 (Figure 7C). At this point, it is not clear which nucleotide of the 3'-CGG motif functions as the orphan base. In guide strands where cytidine is the orphan base, as is the case for the control guide R270X-GG;AC, the ADAR2 mutant E488Q is hyperactive (Figure 7D; 3'-AGC). However, with R270X-24a guide, this mutant is actually less active than the wild type protein (E488) (Figure 7D, 3'-CGG). Orphan base recognition is clearly different with this motif and will require additional studies to understand fully. This is important since our earlier studies have shown how a detailed understanding of the orphan base/ADAR interaction can suggest nucleoside analogs for this position that can further increase editing efficiency[31,32]. It is also possible that mutations other than E488Q may be uniquely beneficial with the 3'-CGG motif targeting the 5'-UGA site[33,34].

The top-performing sequence from the R270X screen and *in vitro* validation experiments described above (R270X-24a) was tested for its ability to recruit an exogenous ADAR2 deaminase domain for directed editing in a cultured human cell line. For this purpose, we turned to an engineered Editase[28], a system we used previously for editing *MECP2* mutations[25]. An *MECP2* guide RNA recruits the hADAR2 catalytic domain to the region in *MECP2* containing the target adenosine. The ADAR2 catalytic domain has been fused

to the λ N peptide from bacteriophage lambda (Editase) that recognizes two BoxB hairpins flanking the guide region (Figure 8A). The Editase is directed to the BoxB hairpins by λN delivering the hADAR2 deaminase domain to the gRNA-target RNA duplex for editing. Guides were designed to enable recoding at the MECP2 R270X site with the Editase system. We compared results from transfecting all three components into HEK cells expressing either a guide bearing the R270X-24a sequence or the guide bearing R270X-GG;AC (Figure 8C-D). The control transfections lacked a guide sequence (Figure 8B). The R270X-GG;AC guide bears two previously known enabling features for editing a 5'-GA site; an A:C pair at the target site and a G:G pair involving the 5' G[26,27], but is otherwise Watson-Crick complementary to the target site. After transfection of plasmids and allowing 72 hours for expression, total RNA was isolated, reverse transcribed, PCR amplified, and subjected to Sanger sequencing. Importantly, we found the guide bearing the 3'-CGG motif from R270X-24a supported significantly more efficient editing (21.4% ± 0.6% SD) than the designed guide R270X-GG,AC (9.0% ± 0.5% SD) (Figure 8C-D). Thus, enabling sequences discovered in an EMERGe screen can direct ADAR editing in human cells with improved performance.

It is clear that ADARs have 5', 3', and opposite base preferences within base paired, A-form duplexes[1,2]. Indeed, these preferences are well understood, and they can be used to inform the design of ADAR guide strands for site-directed RNA editing applications[13,15,27,31,32,34]. However, these preferences are not absolute as ADARs can deaminate adenosines in natural substrate RNAs that contain suboptimal nearest neighbors and/or are adjacent to helix defects, *etc*[19,21,22]. Furthermore, many therapeutically relevant target adenosines for directed RNA editing applications do not conform to ADAR's known preferred nearest neighbors[24,25]. In such cases where rational design based on current knowledge of ADAR-RNA recognition is insufficient, it is sensible to screen for sequences that can enable editing by forming beneficial, non-Watson-Crick structural features in the RNA. Different screening strategies for ADAR-RNA combinations have been published, but each require transfection of plasmid libraries, limiting the size of libraries that can be practically screened[14,20,23]. Here we describe a type of screen that does not require plasmid transfections, and that allows one to query very large libraries. By linking an editing site covalently through a hairpin loop to a site where the sequence is randomized, we could use NGS to quantify the number of reads of G or A associated with each specific sequence within the randomized region. One can imagine many different iterations of the strategy described here with different ADARs, different lengths of the randomized region, different lengths of duplex, and different loop sequences and/or lengths, *etc*. We have demonstrated here that EMERGe screens using human ADAR2 with hairpins containing a 30 bp duplex region, a 10 nt variable region and loops of four or twelve nucleotides can identify editing-enabling motifs. However, given the length and stability of the 30 bp duplex region, it is possible that features found in our winning sequences arise because they provide a benefit in the reverse transcription step of the EMERGe workflow. For this reason, we chose to test hit sequences in the context of duplexes formed *in trans* with 30 nt antisense oligonucleotides bearing the selected sequences in the center. The importance of this validation step is underscored by the observation that the winning sequences in the EMERGe hairpins (R168X-1, R255X-1, and R270X-1) did not produce the best antisense guide strands.

CONCLUSION

We applied the EMERGe screening strategy to three sites in the *MECP2* transcript where mutations are known to cause Rett Syndrome. Each of these sites has a guanosine on the 5' side of the target adenosine: ADARs' least preferred nearest neighbor nucleotide. In all three screens, sequences that supported efficient editing at the target site were predicted to form non-Watson-Crick features around the editing site. Indeed, the most effective sequence identified in the R168X screen (R168X-5) includes an A:C pair at the editing site and a G:G pair involving the 5' G. Both of these features are known to facilitate editing at 5'-GA sites by known mechanisms[25]. This result validated the EMERGe approach for identifying editing-enabling features. In addition, antisense guide strands generated with sequences from EMERGe screens typically showed reduced bystander editing compared to designed guides (see Figures 3, 5 and 6). We believe this is at least in part due to the manner in which the NGS data was analyzed. Sequences leading to efficient editing at nearby adenosines on the substrate RNA would have been excluded in the alignment steps in our data processing workflow. The structural basis for the effects on ADAR selectivity are currently unknown. However, certain helix defects have been shown to increase ADAR specificity.[35]

Interestingly, the R255X and R270X screens each identified G-rich loops opposite the target site that enabled editing. Understanding the exact mechanism by which the G-rich loops enable editing at the R255X and R270X sites will require additional studies. Nevertheless, given the sequence similarities between the two targets, it is likely that features identified in the screens that are proximal to each editing site could be used for similar target sequences. In the case of the R270X target, the 3'-CGG loop was selected opposite the 5'-UGA target sequence and the 3' C was found to be essential for activity. It is possible that this C forms a Watson-Crick pair with the 5' G while also creating a single U bulge on the edited strand, but structural studies will be necessary to test this hypothesis and to define the nature of the ADAR-orphan base interaction. Such follow up analysis will aid further optimization of these editing-enabling features. Nevertheless, we were able to show here that an editing-enabling sequence motif discovered by the EMERGe strategy can be introduced into an ADAR guide strand for directed editing in a cultured human cell line with the Editase system (Figure 8). This guide strand directed more efficient on-target editing than did a guide designed based on our current understanding of ADAR-RNA recognition. It is important to note, however, that directed editing experiments in cultured cell lines provide limited information about efficacy in living organisms. Future work will be needed to establish the efficacy of EMERGe-derived features for directed editing *in vivo* by, for instance, AAV-mediated delivery of the Editase and guide as has been described before with a mouse model for Rett Syndrome.[36]

MATERIALS AND METHODS

Protein overexpression and purification

hADAR2, hADAR2-RD WT and hADAR2-RD E488Q with N-terminal His_{10} tag were overexpressed in *S. cerevisiae* and purified as follows. Cells were lysed using a microfluidizer in lysis buffer containing 20 mM Tris-HCl pH 8.0, 5% (v/v) glycerol, 750 mM NaCl, 35 mM imidazole, 0.01% (v/v) Triton X-100 and 1 mM β-mercaptoethanol

(BME). The clarified lysate was then passed over a 5 mL Ni-NTA column, washed once with 50 mL lysis buffer, then with 50 mL each of the following wash buffers: (1) 20 mM Tris-HCl pH 8.0, 5% (v/v) glycerol, 300 mM NaCl, 35 mM imidazole, 0.01% (v/v) Triton X-100 and 1 mM BME; then (2) 20 mM Tris-HCl pH 8.0, 5% (v/v) glycerol, 100 mM NaCl, 35 mM imidazole, 0.01% (v/v) Triton X-100 and 1 mM BME. Bound proteins were eluted by gradient elution with imidazole (30 to 400 mM). Protein fractions were pooled, concentrated, then dialyzed against a storage buffer containing 20 mM Tris-HCl pH 8.0, 20% (v/v) glycerol, 100 mM NaCl and 1 mM BME. Protein concentration was determined by running the sample alongside bovine serum albumin (BSA) standards in a sodium dodecyl sulfate-PAGE gel, followed by SYPRO Orange (Invitrogen) staining.

Generation and deamination of sequence-randomized hairpin substrates

MECP2 R168X, R255X, and R270X libraries, each containing a T7 promoter sequence upstream, were initially obtained as ssDNA synthesized by IDT. The ssDNAs were then converted into dsDNA by PCR (NEB M0530L). Each dsDNA library was used as a transcription template for T7 RNA polymerase transcription *in vitro* to generate ssRNA libraries (NEB E2040). The ssRNA libraries were purified using denaturing PAGE. Each RNA library (100 nM) was folded by heating to 95 °C for 5 min and subsequent cooling to room temperature over 2 h in 100 mM NaCl, 1 mM EDTA, and 10 mM Tris pH 7.4. The ssRNA library substrates were deaminated by the following reaction. hADAR2 (200 nM) was mixed with 10 nM RNA in a deamination buffer containing 15 mM Tris–HCl, pH 7.8, 60 mM KCl, 3 mM $MgCl_2$ 1.5 mM EDTA, 3% glycerol, 0.003% Nonidet P-40, 0.6 mM DTT, 160 U/mL RNase inhibitor, and 1.0 μg/mL yeast tRNA in a final volume of 100 μL. Reaction solutions were incubated at 30 °C prior to addition of hADAR2, then held at 30 °C for 30 min in the presence of hADAR2. The reaction was quenched by adding 20 μL aliquots to 180 μL of water preheated to 95 °C, held at this temperature for 5 min, then placed on ice for 3 min. The quenched reaction aliquots were then combined and concentrated *in vacuo* to a total volume of 350 μL. The concentrated RNA was used for RT-PCR (Promega A1280) to a total volume of 500 μL to generate cDNA. The cDNA was then purified by gel extraction using 2% agarose (Qiagen 28706) and dried *in vacuo*. The pellet was resuspended in nuclease free water and 2.5 μg of sample was submitted for library preparation and Illumina sequencing. The three samples (R168X, R255X, and R270X) were submitted for NGS under differing Illumina conditions. First, R168X was submitted to Genewiz for 10M (paired-end) PE reads. Second, R255X was submitted to the UC Davis DNA Technologies Core for 10M PE reads. Third, R270X was submitted to Genewiz for 350M PE reads.

Identification of hit sequences from NGS data

Pair-end reads from Miseq were first de-complexed according to the terminal 5’ and 3’ regions using HTStream (https://s4hts.github.io/HTStream/) to produce a preprocessed FASTQ output. This FASTQ output was trimmed to contain sequences that were the approximate known length of the libraries (3 nt shorter and 1 nt longer). Using an RStudio script (https://github.com/csjacobsen/EMERGe/tree/bioinformatics), two regions are selected as regions of interest, the N_{10} and the codon region. The N10 region is the variable guide RNA region. The codon region is 3 nt long, defined as the edited A and the

nucleotides directly 5’ and 3’. Once these two regions are selected, the script calculates the number of reads for each codon that is present for each unique N_{10} region. The comma separated variables file that is outputted by the script is first limited to two codon sequences, the edited and the unedited codon sequences. A table is generated containing each N_{10} region and its corresponding codon reads for both edited and unedited codons. Editing percentage and overall reads for the N_{10} sequences are calculated using the codon reads. The data were then processed to remove background noise. The background filter consisted of each N_{10} sequence requiring at least two edited codon reads and three reads overall. This then resulted in a sortable table of the winning N10 sequences. A subset of sequences, based on overall reads, editing %, and computationally predicted secondary structure, was taken forward to verification.

Testing hit sequences and mutants in antisense oligonucleotide guide strands

Human *MECP2* R168X, R255X and R270X targets were initially obtained as dsDNAs synthesized by IDT. These dsDNAs contain 300 nt excerpts of their respective sequences, centered on the mutation. They also contain a T7 promoter sequence upstream. The dsDNA fragment was PCR amplified (NEB M0530L) to provide 1 μg of dsDNA. Each dsDNA target was used as a transcription template for T7 RNA polymerase transcription *in vitro* to generate deamination-targets (NEB E2040). The ssRNA targets were purified using denaturing PAGE. To maximize yields during T7 RNA polymerase transcription, the sequence GGG was added to the 5’ end of each guide strand followed by 10 nt of target complementary sequence, a N_{10} sequence, and finally 10 nt of target complementary sequence for a total length of 33 nt. Winning guide and mutant guide sequences and their reverse complements were initially obtained as ssDNAs synthesized by IDT. These complementary ssDNAs were then hybridized to a final concentration of 5 μM dsDNA by heating at 95 °C for 5 min and subsequent cooling to room temperature over 2 h in 500 mM NaCl, 1 mM EDTA, and 10 mM Tris pH 7.4. Each winning guide dsDNA was used as a transcription template for T7 RNA polymerase transcription *in vitro* to generate gRNAs (NEB E2040), which were purified using denaturing PAGE. The gRNAs were then hybridized to the targets, with a final concentration of 1 μM gRNA and 100 nM target, by heating to 95 °C for 5 min and subsequent cooling to room temperature over 2 h in 500 mM NaCl, 1 mM EDTA, and 10 mM Tris pH 7.4. The hybridized dsRNA substrate was then deaminated using the following conditions. hADAR2 (200 nM) was mixed with 10 nM substrate dsRNA in a deamination buffer containing 15 mM Tris–HCl, pH 7.4, 60 mM KCl, 1.5 mM EDTA, 3% glycerol, 0.003% Nonidet P-40, 0.5 mM DTT, 160 U/mL RNase inhibitor, and 1.0 μg/mL yeast tRNA in a final volume of 20 μL. Reaction solutions were incubated at 30 °C prior to addition of hADAR2, then held at 30 °C for 30 min in the presence of hADAR2. The reaction was quenched by adding 180 μL of water preheated to 95 °C, held at this temperature for 5 min, then placed on ice for 3 min. The deaminated dsRNA was subjected to RT-PCR (Promega A1280) to generate cDNA. The cDNA was then purified by gel extraction using 2% agarose (Qiagen 28706) and submitted to Genewiz for Sanger sequencing. Percent editing was determined by recording peak values for A to G conversion from Sanger sequencing traces. For comparative analysis between WT and the E488Q mutant ADAR, the same procedure was used as described above with minor changes. First, 100 nM of hADAR2-RD WT or E488Q was used for deaminations. These

proteins bear deaminase domains and dsRBD2 and lack the N-terminal dsRBD of human ADAR2[30]. Second, the reaction was quenched at 15 min to better observe differences in editing.

Directed editing in HEK293T cells using λN-BoxB Editase

All cloning primers and guide sequences are included in Supplemental Tables S24-25. All plasmids were completely sequenced. Plasmid pGM1090 contains the wild-type hADAR2 deaminase domain under control of the CMV promoter (Editase)[37]. To generate an epitope tagged version of the *MECP2* R270X target plasmid (pGM1524), we first subcloned wild-type *MECP2* cDNA, from a plasmid kindly provided by Adrian Bird [38], into a 3xFlag-CMV-10 plasmid (Sigma; pGM1160). We then generated the R270X mutation in pGM1160 (pGM1524) using Q5 Site-Directed Mutagenesis (NEB E0554S) per manufacturer's instruction. As previously described[37], guide plasmids were generated by ligating annealed single-stranded synthetic oligonucleotides containing BsaI overhangs into the pENTR plasmid (Thermo Fisher Scientific; pGM1139) downstream from the *U6* promoter. pGM1139 without *MECP2* guides or BoxB sequences served as our control guide plasmid. The plasmids pGM1525 and pGM1526 contain the *MECP2* guide RNA sequences R270X-GG;AC and R270X-24a, respectively, as well as the λN-BoxB hairpins flanking the guide sequences as described[37]. The host HEK 293T cells (ATCC, CRL-1573) were maintained in DMEM (Thermo Fisher Scientific, 11965092), 10% FBS, and penicillin-streptomycin solution at 37°C in a 5% CO_2 humidified incubator. Transfection, processing, and quantification of RNA editing was performed as previously described[37]. Cells were seeded into a 12-well plate at 1.25 x 10^5 per well. After 24 h, cells were transfected with three plasmids: *MECP2* R270X target (pGM1524), the Editase (pGM1090) and either *MECP2* (pGM1525 or pGM1526) or control guides (pGM1139). We used a 2:1 ratio of Lipofectamine 2000 Reagent (Thermo Fisher Scientific, 11668019) and DNA in Opti-MEM I (Thermo Fisher Scientific, 31985070), then the following reagents were added per well: 125 ng target, 250 ng Editase, and 2.5 μg guide plasmid DNA. At 72 h post transfection, cells were harvested and total RNA was isolated using the Purelink RNA Mini Kit (Thermo Fisher Scientific, 12183025). Residual plasmid DNA was removed using the TURBO DNA-free Kit (Thermo Fisher Scientific, AM1907), and elimination was confirmed by PCR and agarose gel electrophoresis. Total RNA was reverse transcribed using the SuperScript III First-Strand Synthesis System (Thermo Fisher Scientific, 18080051) and primed using oligo dT. Exogenous *MECP2* cDNA was amplified by PCR using a forward primer in the 3xFLAG sequence and a reverse primer in the target sequence (Table S26). Target PCR product was isolated by agarose gel electrophoresis. The bands were purified using the QIAquick Gel Extraction Kit (Qiagen, 28706) and eluted in 1mM Tris pH 8.0 prior to Sanger sequencing. A to I editing efficiency was determined by Sanger sequence analysis of the purified target PCR product. Sequencing peak heights from the antisense strand were determined using the Bioedit Software package (www.mbio.ncsu.edu/BioEdit/bioedit.html; File > Batch Export of Raw Sequence Trace Data). The formula {[C/(C+T)] x 100%} was used to quantify editing percentages at any given cDNA site, where C and T are maximum heights of the edited and nonedited peaks, respectively. Quantification of C/T peak heights on the antisense strand are more accurate than A/G peak heights on the sense strand[16]. All transfection data is shown as reverse complement.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

ACKNOWLEDGMENT

P.A.B. acknowledges financial support from the Rett Syndrome Research Trust and National Institutes of Health in the form of grant R35GM141907. G.M. acknowledges financial support from the National Institute of Health in the form of grant R01NS110868. We would like to give special thanks to M. Settles and J. Li at the UC Davis bioinformatics core for technical help with initial NGS data processing.

ABBREVIATIONS

ADARs	Adenosine Deaminases acting on RNA
A	adenosine
I	inosine
U	uracil
G	guanosine
C	cytosine
dsRBD	double stranded RNA binding domain
ECS	editing site complementary sequence
EMERGe	En Masse Evaluation of RNA Guides
gRNA	guide RNA
N10	10nt region which contains the edit site and complementary 10 variable nucleotides
RT-PCR	reverse transcription PCR
NGS	Next Generation Sequencing
SAR	structure activity relationship

REFERENCES

(1). Bass BL RNA Editing by Adenosine Deaminases That Act on RNA. Annu Rev Biochem 2002, 71, 817–846. 10.1146/annurev.biochem.71.110601.135501. [PubMed: 12045112]

(2). Wang Y; Zheng Y; Beal PA Adenosine Deaminases That Act on RNA (ADARs). Enzymes 2017, 41, 215–268. 10.1016/bs.enz.2017.03.006. [PubMed: 28601223]

(3). Zinshteyn B; Nishikura K Adenosine-to-Inosine RNA Editing. Wiley Interdiscip Rev Syst Biol Med 2009, 1 (2), 202–209. 10.1002/wsbm.10. [PubMed: 20835992]

(4). Keegan LP; Leroy A; Sproul D; O'Connell MA Adenosine Deaminases Acting on RNA (ADARs): RNA-Editing Enzymes. Genome Biol 2004, 5 (2), 209. 10.1186/gb-2004-5-2-209. [PubMed: 14759252]

(5). Nishikura K. Functions and Regulation of RNA Editing by ADAR Deaminases. Annu Rev Biochem 2010, 79, 321–349. 10.1146/annurev-biochem-060208-105251. [PubMed: 20192758]

(6). Gabay O; Shoshan Y; Kopel E; Ben-Zvi U; Mann TD; Bressler N; Cohen-Fultheim R; Schaffer AA; Roth SH; Tzur Z; Levanon EY; Eisenberg E Landscape of Adenosine-to-Inosine RNA Recoding across Human Tissues. Nat Commun 2022, 13 (1), 1184. 10.1038/s41467-022-28841-4. [PubMed: 35246538]

(7). Eisenberg E. Proteome Diversification by RNA Editing. Methods Mol Biol 2021, 2181, 229–251. 10.1007/978-1-0716-0787-9_14. [PubMed: 32729084]

(8). Tang Q; Rigby RE; Young GR; Hvidt AK; Davis T; Tan TK; Bridgeman A; Townsend AR; Kassiotis G; Rehwinkel J Adenosine-to-Inosine Editing of Endogenous Z-Form RNA by the Deaminase ADAR1 Prevents Spontaneous MAVS-Dependent Type I Interferon Responses. Immunity 2021, 54 (9), 1961–1975.e5. 10.1016/j.immuni.2021.08.011. [PubMed: 34525337]

(9). Ramaswami G; Li JB RADAR: A Rigorously Annotated Database of A-to-I RNA Editing. Nucleic Acids Res 2014, 42 (Database issue), D109–13. 10.1093/nar/gkt996. [PubMed: 24163250]

(10). Buchumenski I; Roth SH; Kopel E; Katsman E; Feiglin A; Levanon EY; Eisenberg E Global Quantification Exposes Abundant Low-Level off-Target Activity by Base Editors. Genome Res 2021, 31 (12), 2354–2361. 10.1101/gr.275770.121. [PubMed: 34667118]

(11). Liu X; Sun T; Shcherbina A; Li Q; Jarmoskaite I; Kappel K; Ramaswami G; Das R; Kundaje A; Li JB Learning Cis-Regulatory Principles of ADAR-Based RNA Editing from CRISPR-Mediated Mutagenesis. Nat Commun 2021, 12 (1), 2165. 10.1038/s41467-021-22489-2. [PubMed: 33846332]

(12). Song Y; Yang W; Fu Q; Wu L; Zhao X; Zhang Y; Zhang R IrCLASH Reveals RNA Substrates Recognized by Human ADARs. Nat Struct Mol Biol 2020, 27 (4), 351–362. 10.1038/s41594-020-0398-4. [PubMed: 32203492]

(13). Yi Z; Qu L; Tang H; Liu Z; Liu Y; Tian F; Wang C; Zhang X; Feng Z; Yu Y; Yuan P; Zhao Y; Wei W Engineered Circular ADAR-Recruiting RNAs Increase the Efficiency and Fidelity of RNA Editing in Vitro and in Vivo. Nat Biotechnol 2022, 40 (6), 946–955. 10.1038/s41587-021-01180-3. [PubMed: 35145313]

(14). Katrekar D; Xiang Y; Palmer N; Saha A; Meluzzi D; Mali P Comprehensive Interrogation of the ADAR2 Deaminase Domain for Engineering Enhanced RNA Editing Activity and Specificity. Elife 2022, 11, e75555. 10.7554/eLife.75555. [PubMed: 35044296]

(15). Xiang Y; Katrekar D; Mali P Methods for Recruiting Endogenous and Exogenous ADAR Enzymes for Site-Specific RNA Editing. Methods 2022, 205, 158–166. 10.1016/j.ymeth.2022.06.011. [PubMed: 35779766]

(16). Eggington JM; Greene T; Bass BL Predicting Sites of ADAR Editing in Double-Stranded RNA. Nat Commun 2011, 2, 319. 10.1038/ncomms1324. [PubMed: 21587236]

(17). Riedmann EM; Schopoff S; Hartner JC; Jantsch MF Specificity of ADAR-Mediated RNA Editing in Newly Identified Targets. RNA 2008, 14 (6), 1110–1118. 10.1261/rna.923308. [PubMed: 18430892]

(18). Lehmann KA; Bass BL Double-Stranded RNA Adenosine Deaminases ADAR1 and ADAR2 Have Overlapping Specificities. Biochemistry 2000, 39 (42), 12875–12884. 10.1021/bi001383g. [PubMed: 11041852]

(19). Polson AG; Bass BL Preferential Selection of Adenosines for Modification by Double-Stranded RNA Adenosine Deaminase. EMBO J 1994, 13 (23), 5701–5711. 10.1002/j.1460-2075.1994.tb06908.x. [PubMed: 7527340]

(20). Eifler T; Chan D; Beal PA A Screening Protocol for Identification of Functional Mutants of RNA Editing Adenosine Deaminases. Curr Protoc Chem Biol 2012, 4 (4), 357–369. 10.1002/9780470559277.ch120139. [PubMed: 23788559]

(21). Phelps KJ; Tran K; Eifler T; Erickson AI; Fisher AJ; Beal PA Recognition of Duplex RNA by the Deaminase Domain of the RNA Editing Enzyme ADAR2. Nucleic Acids Res 2015, 43 (2), 1123–1132. 10.1093/nar/gku1345. [PubMed: 25564529]

(22). Kuttan A; Bass BL Mechanistic Insights into Editing-Site Specificity of ADARs. Proc. Natl. Acad. Sci. U. S. A 2012, 109 (48), E3295–E3304. 10.1073/pnas.1212548109. [PubMed: 23129636]

(23). Wang Y; Beal PA Probing RNA Recognition by Human ADAR2 Using a High-Throughput Mutagenesis Method. Nucleic Acids Res. 2016, 44 (20), 9872–9880. 10.1093/nar/gkw799. [PubMed: 27614075]

(24). Christodoulou J; Grimm A; Maher T; Bennetts B RettBASE: The IRSA MECP2 Variation Database-a New Mutation Database in Evolution. Hum. Mutat 2003, 21 (5), 466–472. 10.1002/humu.10194. [PubMed: 12673788]

(25). Sinnamon JR; Jacobson ME; Yung JF; Fisk JR; Jeng S; McWeeney SK; Parmelee LK; Chan CN; Yee SP; Mandel G Targeted RNA Editing in Brainstem Alleviates Respiratory Dysfunction in a Mouse Model of Rett Syndrome. Proc Natl Acad Sci U S A 2022, 119 (33), e2206053119. 10.1073/pnas.2206053119. [PubMed: 35939700]

(26). Doherty EE; Karki A; Wilcox XE; Mendoza HG; Manjunath A; Matos VJ; Fisher AJ; Beal PA ADAR Activation by Inducing a Syn Conformation at Guanosine Adjacent to an Editing Site. Nucleic Acids Res 2022, 50 (19), 10857–10868. 10.1093/nar/gkac897. [PubMed: 36243986]

(27). Schneider MF; Wettengel J; Hoffmann PC; Stafforst T Optimal GuideRNAs for Re-Directing Deaminase Activity of HADAR1 and HADAR2 in Trans. Nucleic Acids Res 2014, 42 (10), e87. 10.1093/nar/gku272. [PubMed: 24744243]

(28). Ojha N; Diaz Quiroz JF; Rosenthal JJC In Vitro and in Cellula Site-Directed RNA Editing Using the ΛNDD-BoxB System. Methods Enzym. 2021, 658, 335–358. 10.1016/bs.mie.2021.06.009.

(29). Matthews MM; Thomas JM; Zheng Y; Tran K; Phelps KJ; Scott AI; Havel J; Fisher AJ; Beal PA; Biology C Structures of Human ADAR2 Bound to DsRNA Reveal Base- Flipping Mechanism and Basis for Site Selectivity. Nat Struct Mol Biol. 2016, 23 (5), 426–433. 10.1038/nsmb.3203.Structures. [PubMed: 27065196]

(30). Thuy-Boun AS; Thomas JM; Grajo HL; Palumbo CM; Park S; Nguyen LT; Fisher AJ; Beal PA Asymmetric Dimerization of Adenosine Deaminase Acting on RNA Facilitates Substrate Recognition. Nucleic Acids Res. 2020, 48 (14), 7958–7972. 10.1093/nar/gkaa532. [PubMed: 32597966]

(31). Doherty EE; Wilcox XE; van Sint Fiet L; Kemmel C; Turunen JJ; Klein B; Tantillo DJ; Fisher AJ; Beal PA Rational Design of RNA Editing Guide Strands: Cytidine Analogs at the Orphan Position. J Am Chem Soc 2021, 143 (18), 6865–6876. 10.1021/jacs.0c13319. [PubMed: 33939417]

(32). Monian P; Shivalila C; Lu G; Shimizu M; Boulay D; Bussow K; Byrne M; Bezigian A; Chatterjee A; Chew D; Desai J; Favaloro F; Godfrey J; Hoss A; Iwamoto N; Kawamoto T; Kumarasamy J; Lamattina A; Lindsey A; Liu F; Looby R; Marappan S; Metterville J; Murphy R; Rossi J; Pu T; Bhattarai B; Standley S; Tripathi S; Yang H; Yin Y; Yu H; Zhou C; Apponi LH; Kandasamy P; Vargeese C Endogenous ADAR-Mediated RNA Editing in Non-Human Primates Using Stereopure Chemically Modified Oligonucleotides. Nat Biotechnol 2022, 40 (7), 1093–1102. 10.1038/s41587-022-01225-1. [PubMed: 35256816]

(33). Wang Y; Havel J; Beal PA A Phenotypic Screen for Functional Mutants of Human Adenosine Deaminase Acting on RNA 1. ACS Chem Biol 2015, 10 (11), 2512–2519. 10.1021/acschembio.5b00711. [PubMed: 26372505]

(34). Monteleone LR; Matthews MM; Palumbo CM; Thomas JM; Zheng Y; Chiang Y; Fisher AJ; Beal PA A Bump-Hole Approach for Directed RNA Editing. Cell Chem. Biol 2019, 26 (2), 269–277.e5. 10.1016/j.chembiol.2018.10.025. [PubMed: 30581135]

(35). Lehmann KA; Bass BL The importance of internal loops within RNA substrates of ADAR1. J Mol Biol. 1999, 291(1), 1–13. 10.1006/imbi.1999.2914. [PubMed: 10438602]

(36). Sinnamon JR; Jacobson ME; Yung JF; Fisk JR; Jeng S; McWeeney SK; Parmelee LK; Chan CN; Yee SP; Mandel G Targeted RNA editing in brainstem alleviates respiratory dysfunction in a mouse model of Rett syndrome. Proc Natl Acad Sci U S A. 2022, 119 (33), e2206053119. 10.1073/pnas.2206053119. [PubMed: 35939700]

(37). Sinnamon JR; Kim SY; Corson GM; Song Z; Nakai H; Adelman JP; Mandel G Site-Directed RNA Repair of Endogenous Mecp2 RNA in Neurons. Proc. Natl. Acad. Sci. U. S. A 2017, 114 (44), E9395–E9402. 10.1073/pnas.1715320114. [PubMed: 29078406]

(38). Lyst MJ; Ekiert R; Ebert DH; Merusi C; Nowak J; Selfridge J; Guy J; Kastan NR; Robinson ND; De Lima Alves F; Rappsilber J; Greenberg ME; Bird A Rett Syndrome Mutations Abolish the

Interaction of MeCP2 with the NCoR/SMRT co-repressor. Nat. Neurosci 2013, 16 (7), 898–902. 10.1038/nn.3434 [PubMed: 23770565]

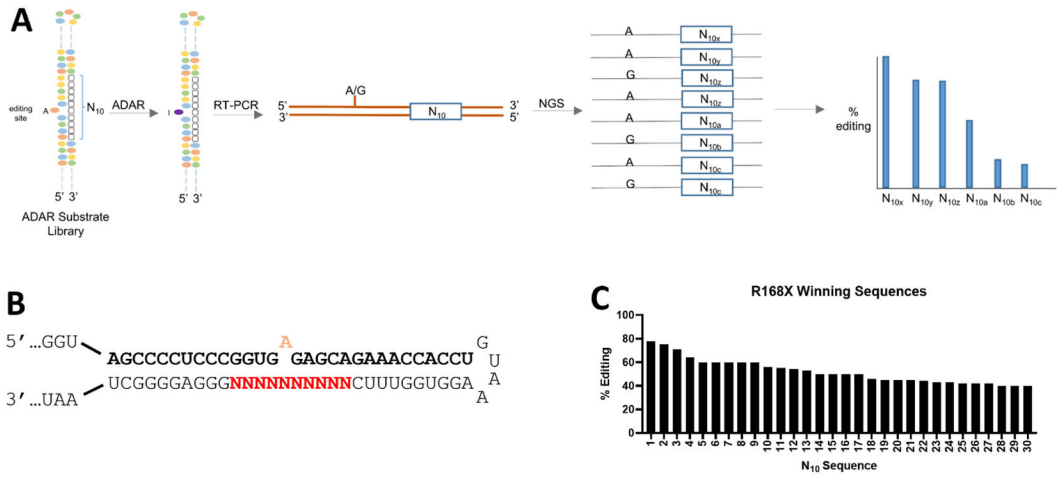


Figure 1:

En Masse Evaluation of RNA Guides (EMERGe). **A**: A library of RNA hairpin substrates is deaminated by ADAR and converted to DNA by RT-PCR. NGS is carried out to determine the number of reads associated with a given N_{10} sequence that has either A or G at the editing site. These values are then used to calculate % editing for each sequence present. **B**: An example of an EMERGe library corresponding to the sequence present in the mRNA for human *MECP2* bearing the R168X mutation (bold) showing the N_{10} sequence randomized region in red. The target A is shown in yellow. **C**: Calculated % editing by ADAR2 for the top 30 sequences having at least 10 reads from the R168X EMERGe dataset.

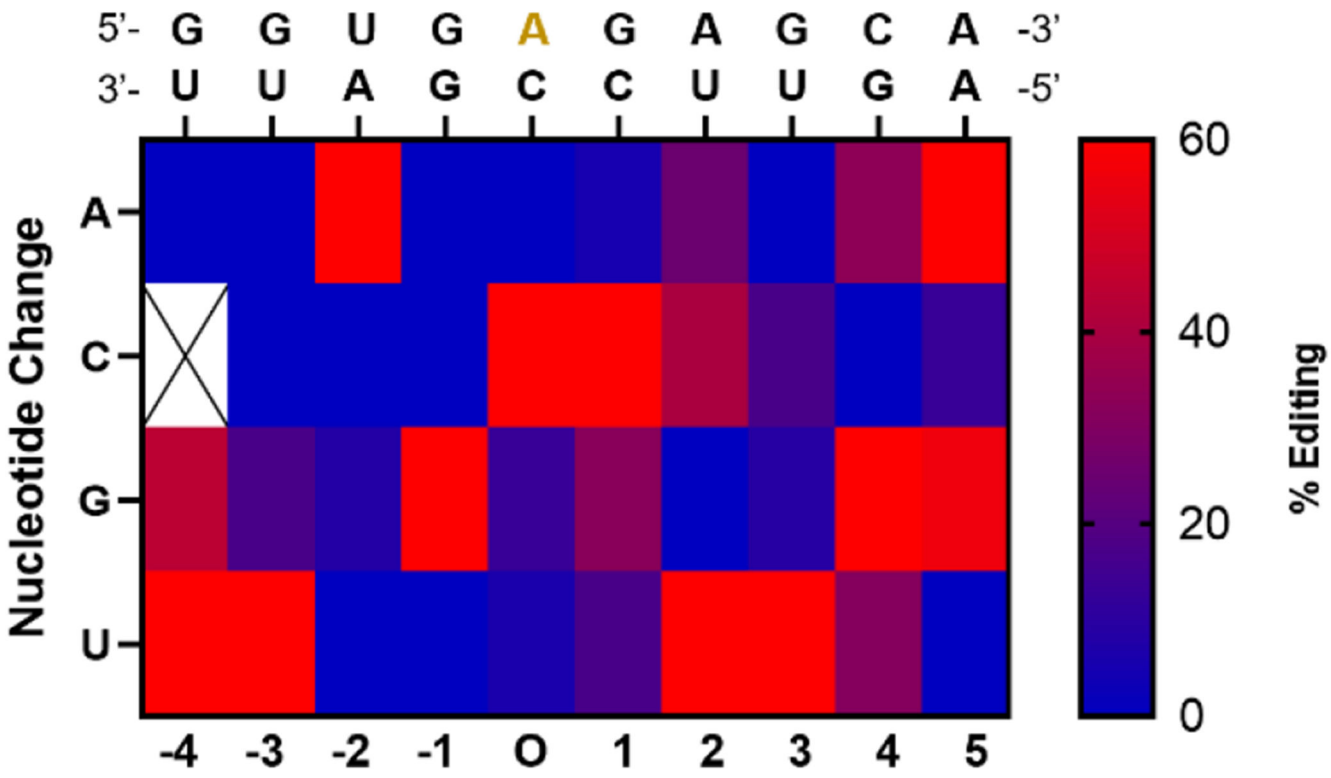


Figure 2:
Percent editing calculated for all the single nucleotide variants of the R168X-5 hairpin present in the EMERGe NGS dataset. Numbers below the heat map correspond to the nucleotide position in the guide strand relative to the orphan (O) position opposite the target A.

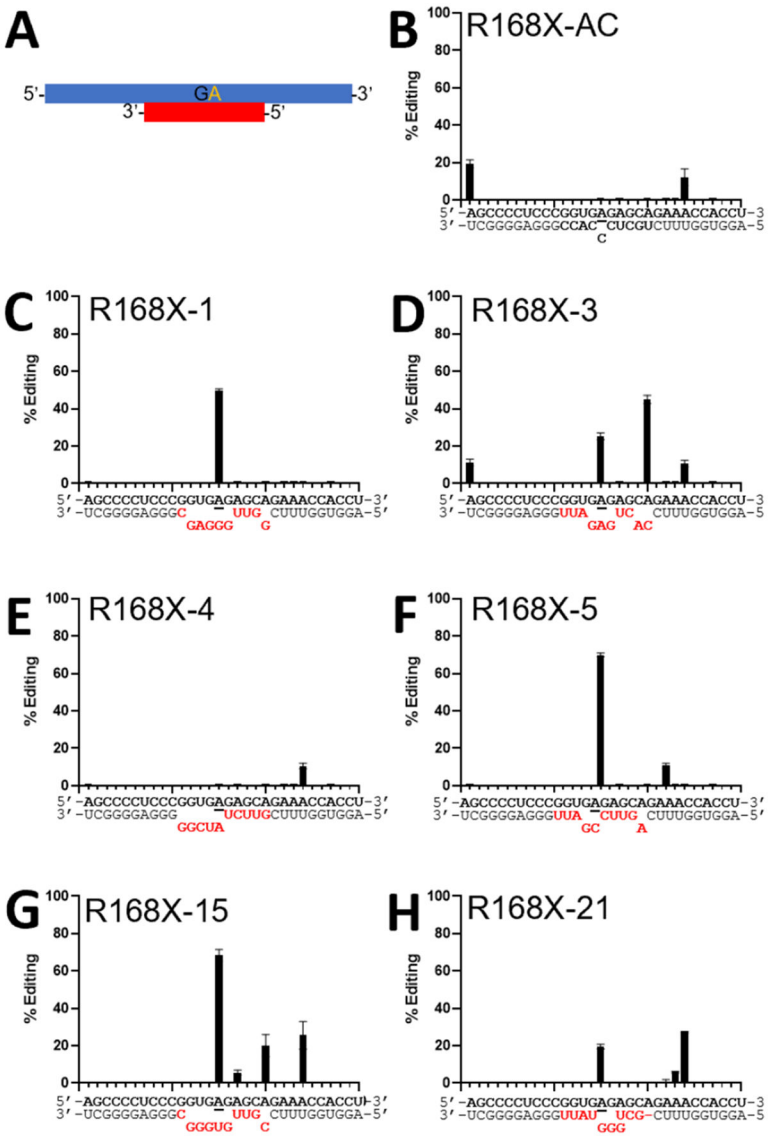


Figure 3:

Directed editing assay results with R168X hit sequences present in 30 nt antisense oligonucleotide guide strands **A**: The experimental design with 30 nt guide strands and a 300 nt target strand. **B-H**: Percent editing at target site and bystander sites observed with 200 nM hADAR2 in a 30 min reaction. The target adenosine is underlined in black. **B**: Complementary guide with A:C mismatch at target site. **C-H**: Guides R168X-1, 3, 4, 5, 15, and 21. Data are plotted as the mean ± s.d. from three independent experiments.

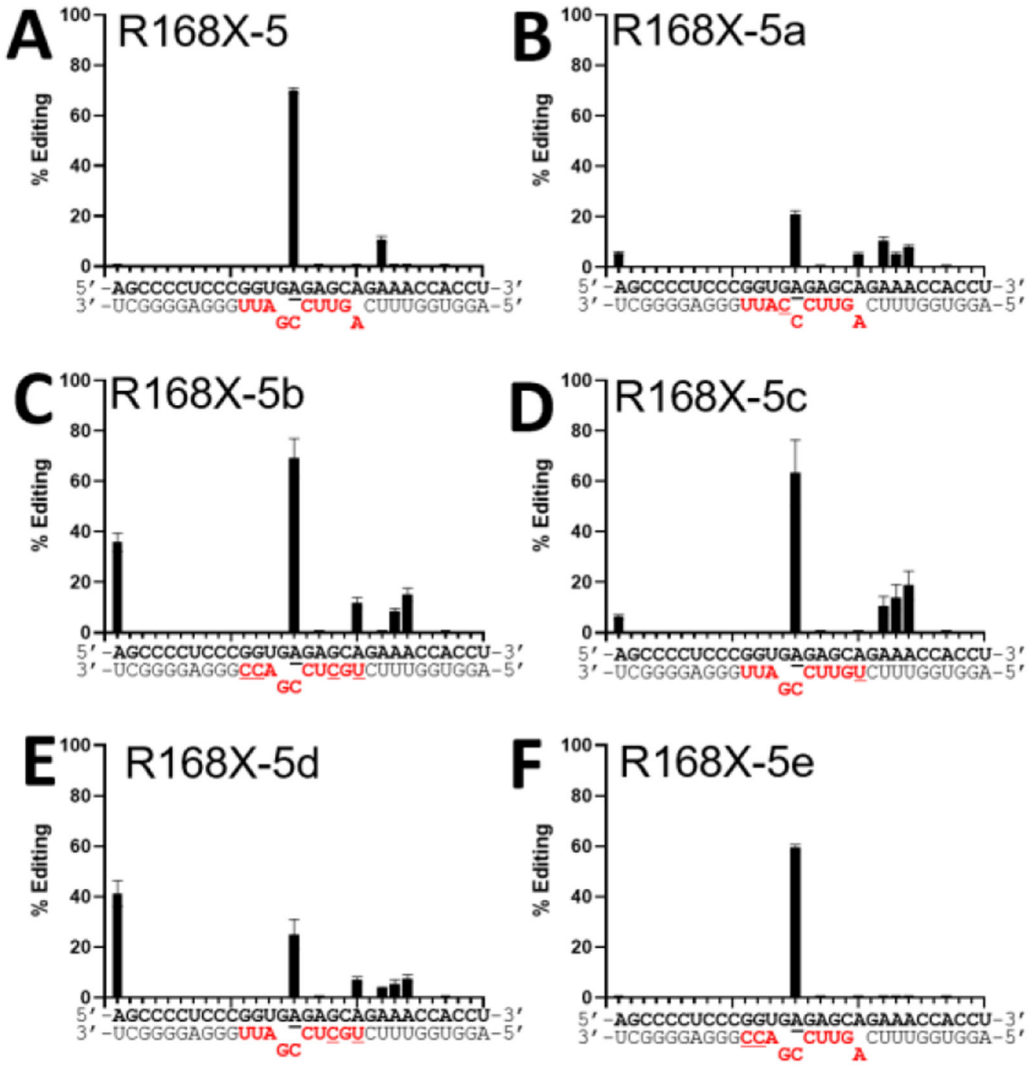


Figure 4:

The effect of mutations in the R168X-5 guide. Sequence changes made to the R168X-5 sequence are underlined in red **A**: R168X-5 guide. **B-F**: R168X-5a-e. On-target editing for R168X-5 is greater than R168X-5e (Welch's t-test, $p < 0.1\%$). Data are plotted as the mean ± s.d. from three independent experiments.

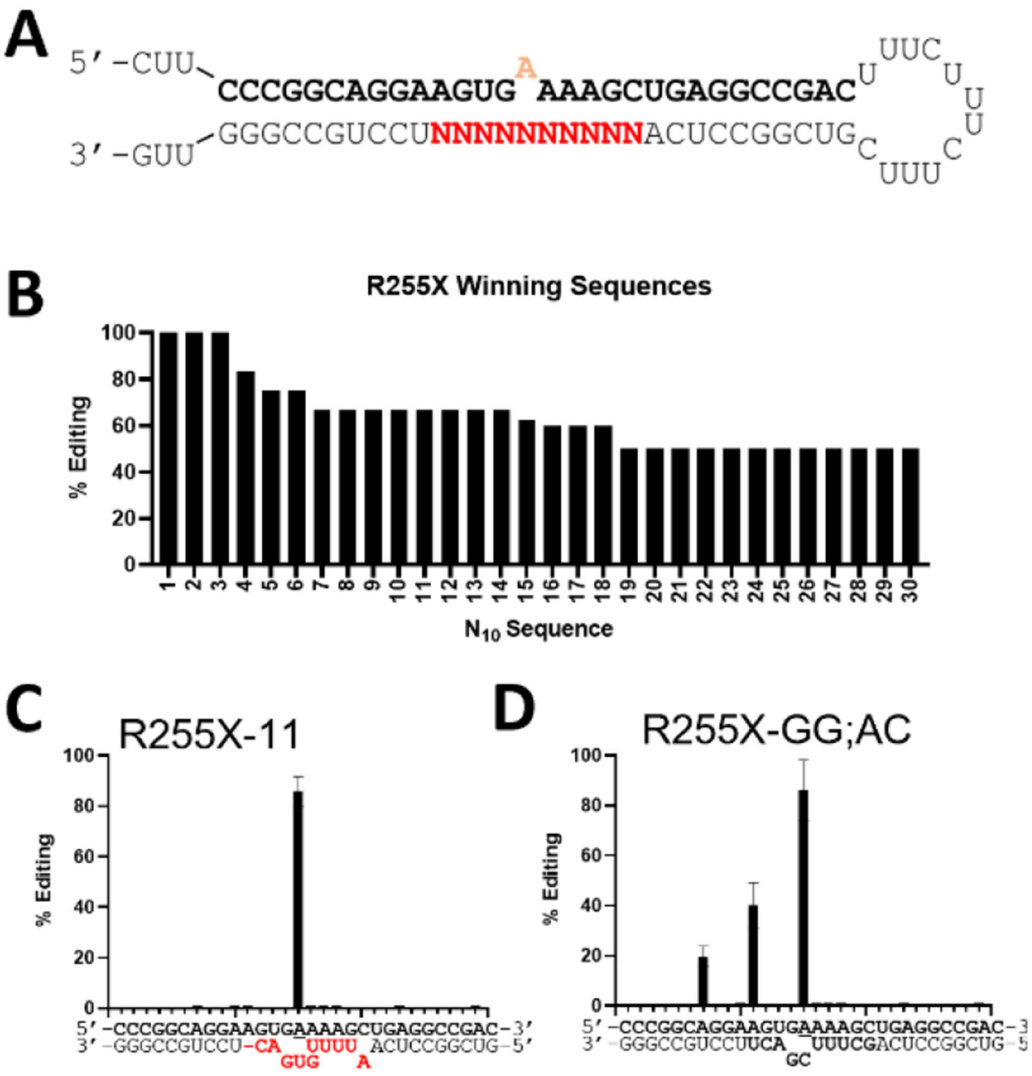


Figure 5:

Results from a screen with hADAR2 and an *MECP2* R255X library. A: R255X library design. B: Percent editing for top 30 sequences from R255X EMERGe dataset. C and D: Percent editing at target site and bystander sites observed with 100 nM hADAR2 in a 1h reaction. C. Complementary guide with A:C mismatch at target site and adjacent G:G mismatch. D: R255X-11 guide. Data are plotted as the mean ± s.d. from three independent experiments.

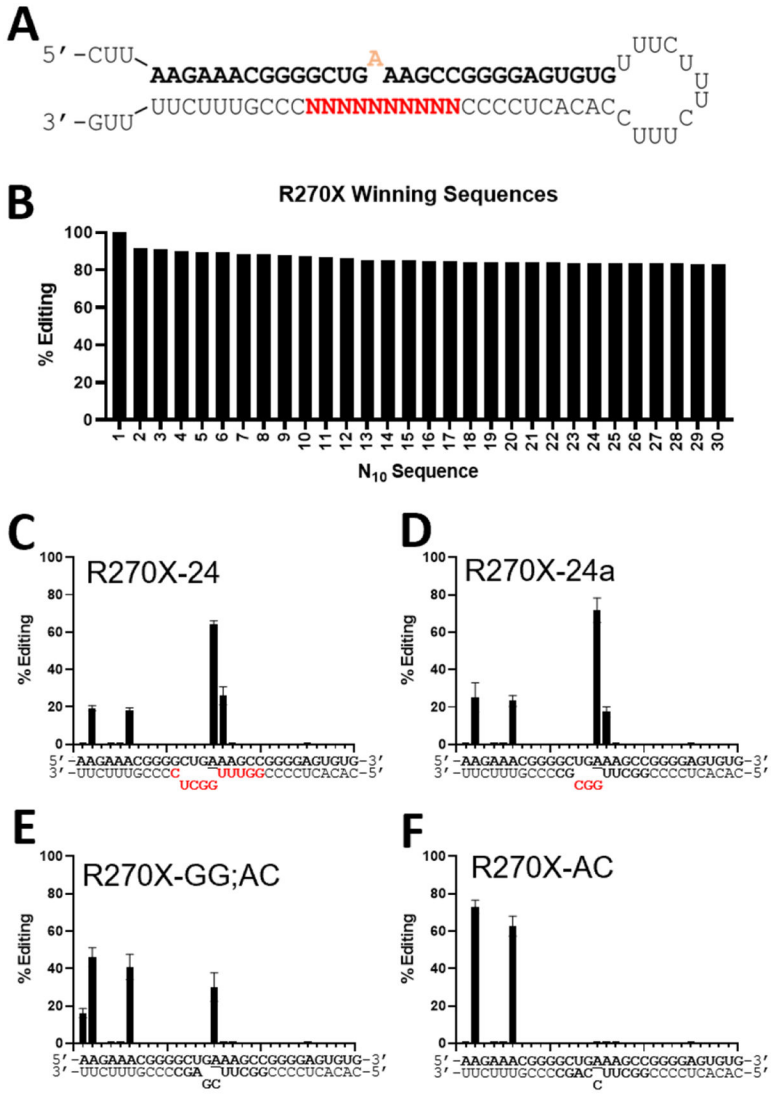


Figure 6:

Results from a screen with hADAR2 and an *MECP2* R270X library. A: *MECP2* R270X substrate design. B: Percent editing for top 30 sequences from R270X EMERGe dataset. C-F: Percent editing at target site and bystander sites observed with 200 nM hADAR2 in a 30 min reaction. C: R270X-24 guide. D: R270X-24a. E: Complementary guide with A:C mismatch at target site and adjacent G:G mismatch. F: Complementary guide with A:C mismatch at target site. Data are plotted as the mean ± s.d. from three independent experiments.

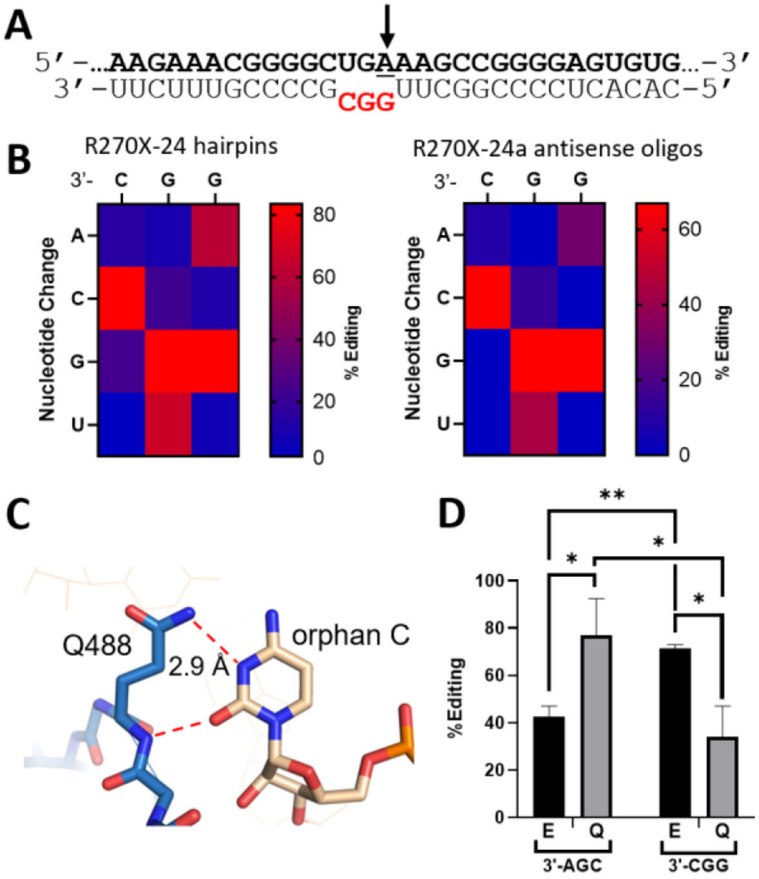


Figure 7:

EMERGe screen identifies a novel motif enabling editing at R270X premature termination codon. A: *MECP2* R270X 300mer RNA substrate and R270X-24a gRNA sequence. The unique 3’-CGG motif is shown in red. B: (Left) Percent editing calculated for all the single nucleotide variants of the R270X-24 hairpin present in the EMERGe NGS dataset. (Right) The effect of mutations in 30 nt antisense guides at the 3-CGG motif in the R270X-24a sequence. C: Interaction of hADAR2 Q488 with orphan C. D: Effect of wild type (E) and E488Q (Q) ADAR2 on reaction of R270X-GG;AC (with 3’-AGC opposite 5’-UGA in target) and R270X-24a (with 3’-CGG opposite 5’-UGA). Data are plotted as the mean ± s.d. from three independent experiments. P values are calculated from Welch’s t-tests (* $p < 0.05$ and ** $p < 0.01$).

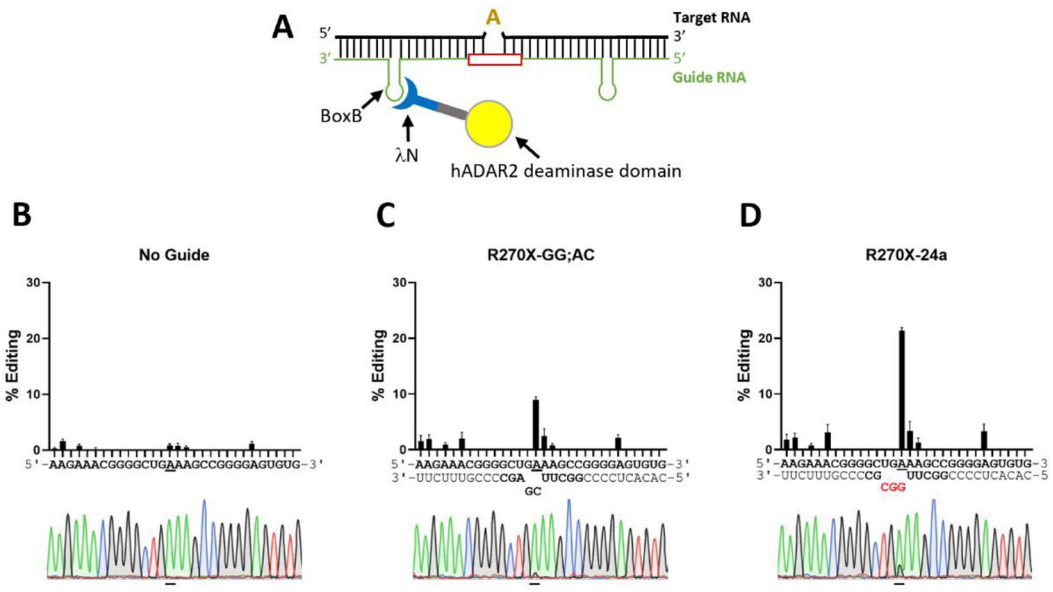


Figure 8:

Directed editing in HEK293T cells using lN-BoxB Editase and R270X guide sequences. A: The general design of the lN-BoxB Editase system. Red box indicates the variable region opposite the target adenosine. B-D: Percent editing at target site and bystander sites observed in transfected HEK293T cells. Representative Sanger sequencing chromatograms span the guide region. Target adenosines are underlined. B: No guide control. C: R270X-GG;AC guide. D: R270X-24a guide. Red nucleotides indicate the 3'-CGG motif identified from the EMERGe screen. Data are plotted as the mean ± s.d. from five independent experiments. *P* value for editing percentages at target adenosines in C versus D used Welch's t test (**** $p<0.0001$).