## Title

Development and Application of Oligonucleotide Assays for Site-Directed RNA Editing



## Author

Jacobsen, Casey Stephen

## Publication Date

2023

Peer reviewed|Thesis/dissertation

Development and Application of Oligonucleotide Assays for Site-Directed RNA Editing

By

CASEY STEPHEN JACOBSEN
DISSERTATION

Submitted in partial satisfaction of the requirements for the degree of

DOCTOR OF PHILOSOPHY

in

Chemistry and Chemical Biology

in the

OFFICE OF GRADUATE STUDIES

of the

UNIVERSITY OF CALIFORNIA

DAVIS

Approved:

Peter A. Beal, Chair

Sheila S. David

Justin B. Siegel

Committee in Charge

2023

Thank you to everyone that has been a mentor, mentee, or friend during this journey.

Thank you to my family for your love and patience.

Dedicated to Ann M. Haworth

# Development and Application of Oligonucleotide Assays for Site-Directed RNA Editing

## 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. Another limitation of ADAR editing is its specificity. ADARs can edit non-target sites due to the presence of good targets elsewhere. To achieve more specific editing, a gRNA may be discovered that allows for a reduction in off-target editing.

In this dissertation, I describe a strategy for the *in vitro* evaluation of very large libraries of ADAR substrates, called En Mass Evaluation of RNA Guides (EMERGe). EMERGe allows for a comprehensive screening of ADAR substrate RNAs that complements current design approaches. I used this approach to identify sequence motifs for gRNAs that enable editing in otherwise difficult-to-edit target sites. 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. Chapter 1 provides a general introduction to ADARs and their application for site-directed RNA editing. This chapter also describes current advancements in gRNA design and

discovery. Chapter 2 outlines the studies used before the EMERGe assay. Additionally, this chapter will also describe the development of the EMERGe assay. The validation of the EMERGe assay is discussed within Chapter 3. Other discoveries made by the EMERGe assays will be tested *in vitro* within this chapter. Finally, Chapter 4 discusses *in cellula* applications of EMERGe-derived results. This includes testing EMERGe-derived gRNAs, *in cellula* results with a previously untested chemical modification, and the limitations of a new yeast-based assay.

# Table of Contents

# List of Figures

# List of Tables

# Abbreviations

| | |
|---|---|
| 10 variable nucleotides | $N_{10}$ |
| 2’-*O*-methyl | OME |
| 2'-5'-oligoadenylate synthase | OAS |
| 3-deaza deoxy A | 3da-dA |
| adeno-associated virus | AAV |
| Adenosine | A |
| Adenosine Deaminase Acting on RNA | ADAR |
| antisense oligonucleotide | ASO |
| artificial intelligence | AI |
| Cytodine | C |
| deoxy ribonucleic acid | DNA |
| double stranded RNA | dsRNA |
| double stranded RNA Binding domain | dsRBD |
| editing site complementary sequence | ECS |
| En Masse Evaluation of RNA Guides | EMERGe |
| Guanosine | G |
| guide RNA | gRNA |
| Human Embryonic Kidney 239T | HEK293T |
| Inosine | I |
| melanoma differentiation-associated protein 5 | MDA5 |
| melting temperature | $T_m$ |
| methyl CpG binding protein 2 | *MECP2* |
| mouse alpha-L-iduronidase | mIDUA |
| Next Generation Sequencing | NGS |
| nuclear export signal | NES |
| nuclear localization signal | NLS |
| Orphan | O |
| phosphate buffered saline | PBS |
| Polymerase Chain reaction | PCR |
| Protein kinase R | PKR |
| reduced abasic | rAb |
| Rett Syndrome | RTT |
| reverse transcriptase | RT |
| reverse transcription PCR | RTPCR |
| Ribonucleic Acid | RNA |
| Saturation mutagenesis, fluorescent-activated cell sorting, and next-generation sequencing | Sat-FACS-Seq |
| single nucleotide variant | SNV |
| single stranded DNA | ssDNA |
| site-directed RNA editing | SDRE |

| | |
|---|---|
| structure activity relationship | SAR |
| Uracil | U |
| yeast enhanced green fluorescent protein | yeGFP |
| β-mercaptoethanol | BME |

# Chapter 1

# A General Introduction to Adenosine Deaminases Acting on RNA and Site Directed RNA Editing

## 1.1: Adenosine Deaminases Acting on RNA (ADARs)

Adenosine deaminases acting on RNA (ADARs) are a family of RNA editing proteins that exists exclusively in metazoans.[1] RNA editing is defined as any structural change in RNA that does not originate directly from DNA.[2] These alterations to the RNA fall under three main categories: insertion, deletion, or base modifications. ADARs catalyze the hydrolytic deamination of Adenosine (A) to Inosine (I) in double stranded RNA (dsRNA), which falls under the base modification category and is the most common form of RNA editing.[3, 4] Inosine is read as Guanosine (G) by cellular processes such as translation and reverse transcription (Figure 1.1). Therefore, ADAR effectively completes an A-to-G edit.

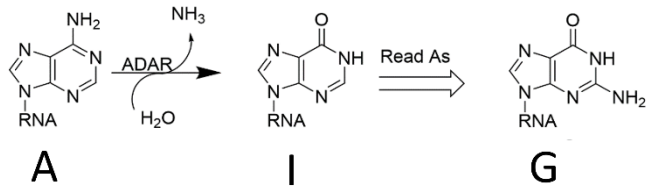


Figure 1.1: ADAR editing Scheme.

Two catalytically active forms of ADARs exist within humans: ADAR1 and ADAR2.[3] All ADARs contain two important features: (1) the catalytically active deaminase domain, and (2) the protein's double stranded RNA binding domains (dsRBDs). These features enable ADARs to detect RNA substrates and catalyze the deamination reaction. Additionally, all human ADARs contain nuclear localization signals (NLS),[5] meaning that they exist mostly in the nucleus (Figure 1.2).

ADAR1 exists in two isoforms, ADAR1 p110 and ADAR1 p150. The main difference comes down to the p150 form of ADAR1 containing an extra Z-DNA binding domain and a nuclear export signal (NES). Hence, ADAR1 p150, can shuttle between the nucleus and the cytoplasm.

ADARs' ability to edit human dsRNAs has the potential to induce changes in secondary structure, codon sequences, and splice sites.[6] ADAR's have a high propensity to edit and thereby change the secondary structure of Alu repeats, which are repetitive, self-folding sequences that account for over 10% of the human genome.[7] Including Alu repeats, ADARs edit at over 100 million different sites in the transcriptome.[8, 9]

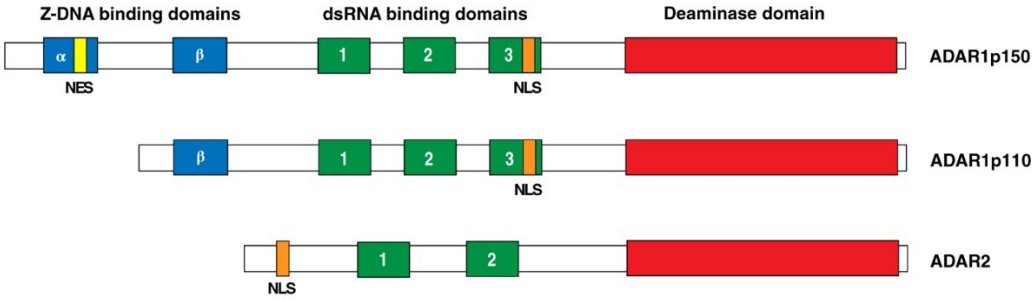


Figure 1.2: Domain map of human ADARs.
Adapted from "The role of RNA editing enzyme ADAR1 in human disease"[5].

Alu editing by ADARs is linked to gene regulation.[10] It has been reported that when ADAR edits Alus, sequences adjacent to these elements are also being edited. Therefore, Alu's are considered as ADAR recruitment sequences.[11] Because Alu's are primate-specific elements and are heavily edited by ADAR, a lack of editing in these sites can be detrimental for cellular function. Alu sequences are involved in innate immunity. Alu dsRNA is a ligand for the protein MDA5, a key protein in an apoptotic pathway.[5] The apoptotic pathway is not triggered when the dsRNAs are edited. Other common ADAR sites also center around innate immunity, such the PKR and OAS

apoptosis pathways.[12] ADAR editing distinguishes dsRNA between "self" and "non-self" for these three immune response pathways. In addition to Alu's needing to be edited by ADARs, ADARs are required for mice and human cell lines, since a knockout variant is embryonically lethal.[13, 14] A key ADAR substrate is the glutamate receptor subunit mRNA. These receptors are key for mouse mobility and reduced ADAR editing has been seen to induce seizures.[15] Therefore, ADAR editing is required for human life, whether this be through Alu editing or through other ADAR editing sites.

### 1.2: Site-directed RNA editing (SDRE)

The most common forms of genetic diseases originate from single point mutations (58%).[16] Of these single point mutations, the majority (47%) can be reversed by an A-to-G edit.[16] ADARs can directly revert 27% of all genetic variants that are associated with diseases on the transcriptomic level. However, there are many other mutations that ADAR has potential to "correct" apart from direct mutation reversions. For example, the disease Rett syndrome (RTT) contains an Arg (CGA) to stop (UGA) mutation (R168X), in the *methyl CpG binding protein 2* (*MECP2*) gene. ADAR has the potential to edit this stop codon to a Trp (UG"G") and cause an overall missense mutation. This missense mutation is functionally superior to a nonsense mutation because it allows for the translation of a full-length protein. This expands ADARs repertoire of therapeutically relevant editable targets in the transcriptome.

When a gene of interest is transcribed, the resulting transcript exists as a single stranded messenger RNA (mRNA). Due to ADAR's requirement for dsRNA substrates, an mRNA transcript containing a desired target can be made into an ADAR substrate by designing a complementary

guide RNA (gRNA). Through the use of gRNAs, ADARs can be recruited to a specific adenosine of interest to achieve site-directed RNA editing. gRNAs typically fall into two categories. The first category is a chemically modified small gRNA. This gRNA can be delivered through lipid nanoparticles or conjugated with tags that target gRNA to specific cell types. [17] The second category is an encodable gRNA. This type of gRNA can be delivered to specific cell types with designed adeno-associated viruses (AAVs). [18]

### 1.3: ADAR mechanism and structure.

ADAR requires a dsRNA for substrate recoginition[19] and recognizes regions of ~30 bp. [20] To facilitate catalysis, ADAR uses a base-flipping mechanism, positioning the target adenosine into the active site. [21-23] Finally, by using an active site coordinated hydroxide, adenosine undergoes a nucleophilic aromatic substitution reaction. This results in a tetrahedral intermediate, [24, 25] (Figure 1.3) that subsequently has an amine functional group leave as ammonia.

ADAR has sequence preferences for edit sites that originate from the nature of the protein-RNA interface. ADAR's sequence preference is strongly influenced by the nucleotides that are directly located 5' and 3' of the edit site, or the nearest neighbor nucleotides. ADARs prefer to edit adenosines in a 5'-UAG sequence context. Additionally, these proteins prefer to edit when the target A is in an A:C mismatch. [23, 26] This is because the target A must be flipped into the active site by the base-flipping loop, which stabilizes the flipped-out RNA conformation. This is due to a glutamate contacting the N3 of cytosine. From the crystal structure of ADAR bound to RNA, three key protein-RNA contacts involving the nearest neighbor nucleotides, and base-flipping loop are observed. First, the G489 residue approaches the base at the 5' position (Figure 1.4A). Second,

the E488 residue contacts the orphan nucleotide, or the nucleotide that is left behind when the target A is flipped into the active site (Figure 1.4B). Lastly, a beneficial hydrogen bond between S486 and the base at the 3' position is observed (Figure 1.4C). These three interactions rationalize the ideal nearest neighbor contexts for ADAR2.

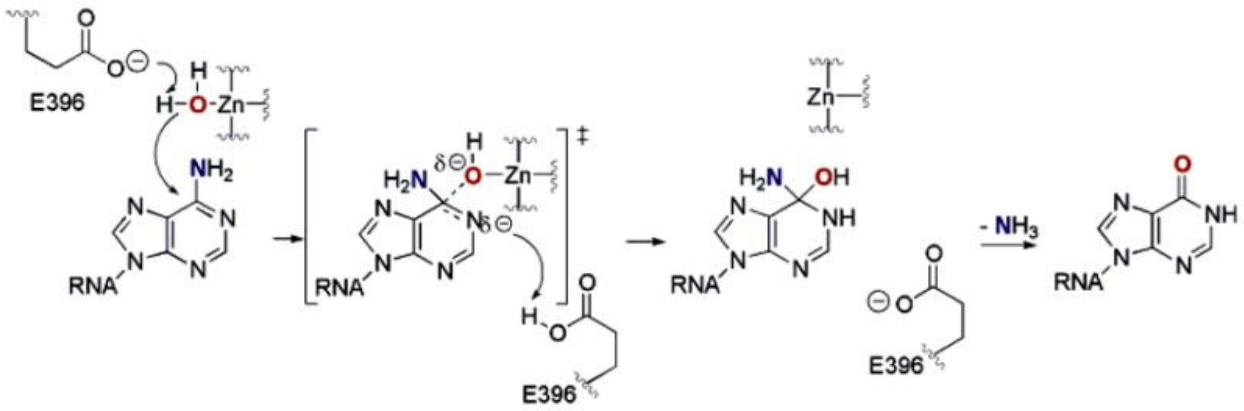


Figure 1.3: ADARs hydrolytic deamination mechanism.
Adapted from "A transition state analogue for an RNA-editing reaction." [25]

In addition to these nearest neighbor contacts, many ADAR residues have been observed to interface with dsRNA. Both 5' and 3' sides of the dsRNA make contact with the deaminase domain of ADAR via the 5' and 3' binding loops. These contacts span over a shorter binding register than the typical ADAR substrate length and can denote a minimum length of duplex RNA for substrate binding[27]. In addition to the deaminase domain, a key dsRBD contact is observed at the +12 position of the RNA (nucleotides that are on the 3' side of the gRNA are referred to as -1, -2, -3, etc., with respect to their distance from the orphan position, while nucleotides on the 5' side of the gRNA are referred to as +1, +2, +3, etc.). At this position, K282 interacts with the backbone of the dsRNA (Figure 1.5). dsRBDs are observed to be transient ADAR domains, allowing

for ADAR to edit targets within many sequence contexts. [20] However, this specific interaction is strong enough to lock the dsRBD into a single position in solved crystal structures.

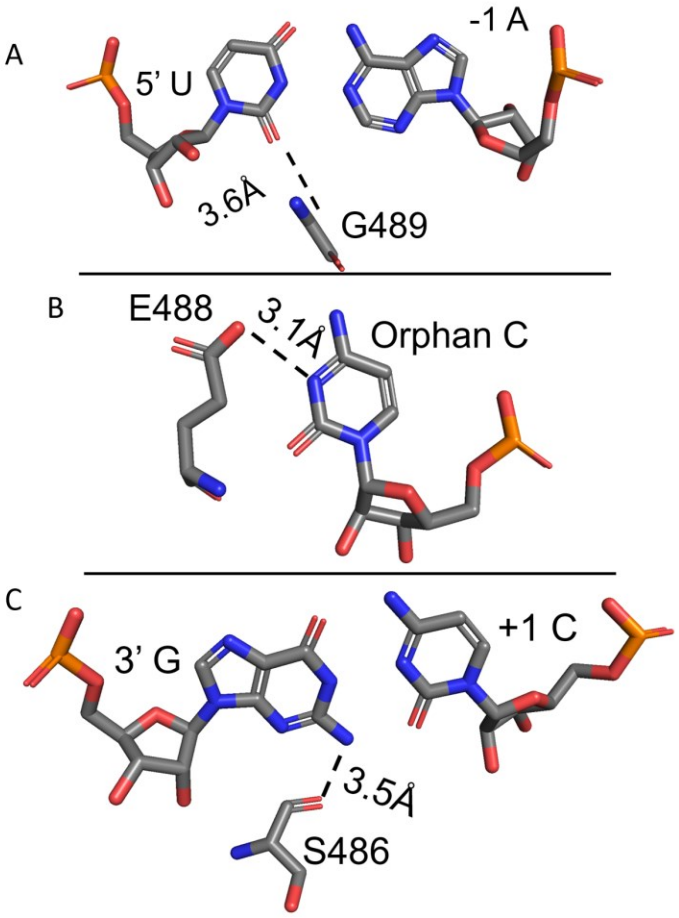


Figure 1.4: ADAR2 nearest neighbor and orphan base contacts.

A: ADAR2's closest contact (G489) to the 5' nearest neighbor. B: ADAR2's closest contact (E488) to the orphan position. C: ADAR2's closest contact (S486) to the 3' nearest neighbor. PDB: 5HP3[23] used for all measurements.

### 1.4: ADAR and gRNA optimization for SDRE

As mentioned previously ADARs have nearest neighbor sequence preferences. It has been observed that ADARs have a dramatic reduction in editing adenosines at positions with a 5'-G. [22, 28] This is observed through the prevalence of endogenous ADAR targets containing a 5'-G

(Table 1.1). Additionally, ADAR's poorest nearest neighbor context has been observed biochemically as a 5'-GAA. [28] This sequence context introduces as steric clash with the G489 position when compared to the ideal 5'-U (Figure 1.6A), completely blocking editing in a G:C pairing[22] (Figure 1.6B). Recently an article has been published providing a solution to the 5'-G, poor nearest neighbor. It was observed that when the nucleotide across from the 5'-G is replaced with a G or a 3-deaza deoxy A (3da-dA), editing is restored (Figure 1.6C and D). This is originates from the new G:G and G:3da-dA pairing interactions, by removing the steric clash through inducing a *syn* conformation in the 5'-G. This context enables editing when the 5' nearest neighbor is a guanosine. There may be other solutions that result in better editing levels.

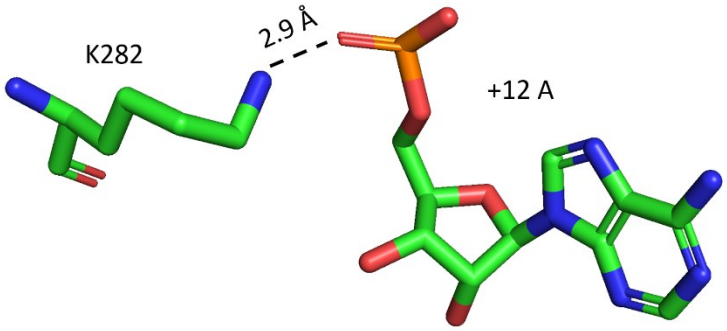


Figure 1.5: A key ADAR2 dsRBD-RNA contact (PDB: 6VFF). [26]

The E488 position has also been a key area of study for enabling ADAR editing. Glu 488 interacts with the orphan position in a pH-dependent manner[29] (Figure 1.7A). In addition to this, there is a hyperactive mutant of ADAR2 (E488Q) that enables editing[29] (Figure 1.7B). In both contexts, low pH and E488Q, there is a hydrogen bridging between the 488 position and the orphan C. This interaction has been mimicked using a modified nucleotide, Benner's base Z[29], which acts as a protonated N3 cytidine.

In addition to enabling editing through mutagenesis and chemical modifications, an orthogonal (bump-hole) system has was developed using the E488 position to enable specific ADAR editing.[30] By using a bulky and typically inactive ADAR mutant (E488X, X= F, Y, W), specific editing can be achieved when a gRNA contains a corresponding hole at the orphan position. The hole used was a reduced abasic (rAb) orphan, a nucleotide without a nucleobase (Figure 1.7C). This combination led to increased specificity but requires a chemically modified guide RNA. Therefore, there is no encodable "hole" for this system. Ideally, the bump-hole system delivers both the mutant ADAR and gRNA simultaneously though the use of an AAV.

Table 1.1: ADAR editing frequencies at different nearest neighbor contexts.

Data derived from the RADAR database.[31] The frequency shown determined by the corresponding genomic loci for each RNA target, independent from RNA expression levels.

| Target Context | Frequency |
|---|---|
| AAA | 6.92% |
| AAC | 5.17% |
| AAG | 9.66% |
| AAU | 7.07% |
| CAA | 6.20% |
| CAC | 7.66% |
| CAG | 15.27% |
| CAU | 3.12% |
| GAA | 1.58% |
| GAC | 1.03% |
| GAG | 5.31% |
| GAU | 1.27% |
| UAA | 6.03% |
| UAC | 7.73% |
| UAG | 12.91% |
| UAU | 3.07% |

Currently, ADARs' therapeutic ability is being investigated. [17, 29, 32-36] These therapeutic attempts are using SDRE by hybridizing gRNAs to specific target RNAs forming a dsRNA ADAR substrate. There are two schools of thought when it comes to promoting ADAR editing. Some laboratories are focused on using chemical modifications within the gRNA to enable editing at therapeutically relevant sites. [17, 29] These modified nucleotides can strengthen contacts within the protein-RNA interface, which will promote higher editing levels. Other laboratories are focused on using sequence mismatches and other dsRNA perturbations to enable ADAR editing. [35] These mismatches are observed in endogenous targets and are encodable. [20] Additionally, it has been observed that mismatches at distal positions within the dsRNA promote editing. [37]

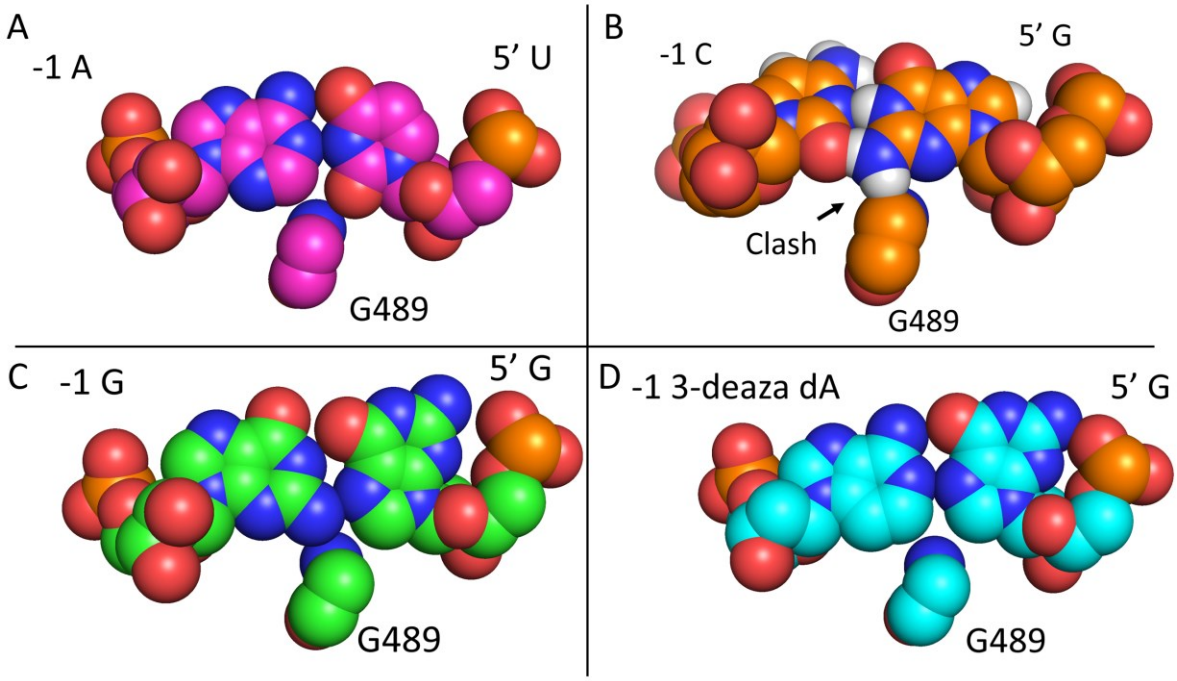


Figure 1.6: The 5' nearest neighbor structures and interactions.

A: The ideal 5' nearest neighbor context of a U:A pair (PDB: 6VFF). [26] B: A computational mutagenesis model of the G:C pair (PDB: 6VFF). [26] A steric clash is observed between the 5'G and the G489 residue. C: The G:G pairing that enables editing in a 5'G context (PDB: 8EOF). [22] D: The G:3deaza-dA pairing that enables editing in a 5'-G context (PDB: 8E4X). [22]

There are many ways to produce gRNAs through rational design, but there aren't many ways to complete a screen on gRNAs. A recent publication coming from the Rosenthal laboratory has shown that there is a way to query large libraries of ADARs guides through a PCR-enabled selection strategy. [35] This publication focuses on enabling editing by altering nucleotides distal to the editing site. Randomized nucleotides were introduced across from the target RNA outside of

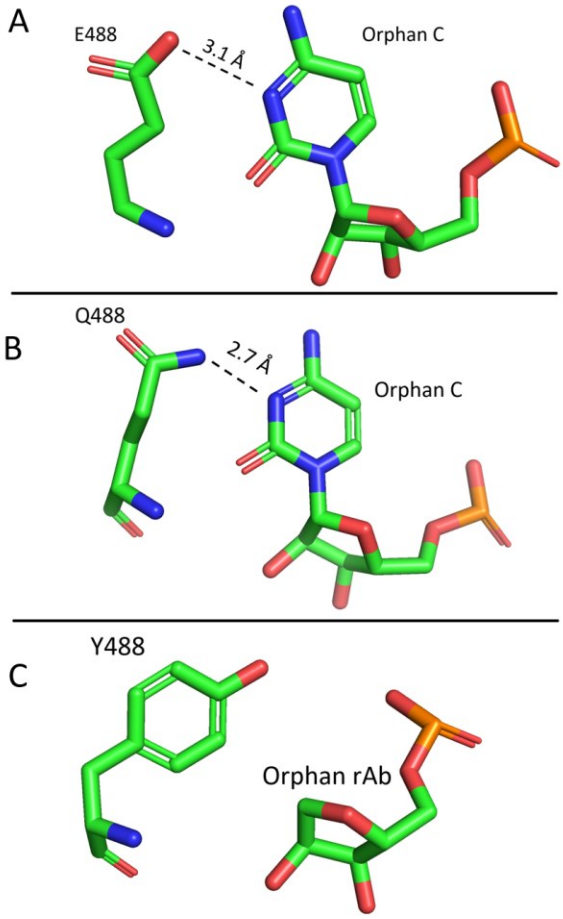


Figure 1.7: Orphan position interactions with multiple amino acids at the 488 position.

A: The standard wild type ADAR2 contact with the orphan position (PDB: 5HP3). [23] B: The hyperactive mutant (E488Q) ADAR2 contact with the orphan position (PDB: 5ED2). [23] C: The bump-hole ADAR2 interaction. The rAb at the orphan position allows enough space for the bulky mutant (E488Y) ADAR2 (PDB: 6D06). [30]

the nearest neighbor region, using a biased dataset. For their randomized regions, 91% of the time the sequence has the complementary base, with each other base being 3%. In the 30 nt gRNA space, this approach queries guides with an average of 7 mismatches. This means that they were not enabling editing through completely new sequence motifs and were only testing a small subset of the possible sequence space. Additionally, the screen was focused on only the ideal context of a 5'-UAG target with a 3'-ACC editing site complementary sequence (ECS) meaning that the distal effect would be beneficial to editing, but not entirely enabling. Through their screen, the Rosenthal laboratory was able to select guides that promote editing using distal mismatches and minor perturbations to the RNA duplex. While this screen is progress towards a large library gRNA assay, their methodology only enabled testing of distal effects.

### 1.5: Goals and Summary

The development of gRNAs enables the ability to recruit ADAR and enables a therapeutic RNA editing system. For ADAR to also be a cellular editor, there needs to be progress in gRNA delivery. One main delivery modality focuses on AAVs and plasmid-based gRNA delivery. This modality requires that ADAR gRNAs be encodable. Designing gRNAs has typically been completed by rational design. Most of the rational design has been limited by chemical modification, meaning many gRNAs are not encodable. While there are ways to rationally design gRNAs, we are void of ways to conduct an ADAR screen completely unbiased by rational design and enable the discovery of encodable gRNAs. Unbiased screens will allow for the discovery of gRNA sequences that are currently unknown and will also provide insight into ADAR-RNA interactions.

The contents of this dissertation will fill this void within ADAR gRNA design by developing a Next Generation Sequencing (NGS) based screen. By creating this screen, I developed an assay that queries gRNAs without rational design biases. This is an advancement for novel gRNA design and enables subsequent experiments that will further our understanding of ADAR-RNA interactions. In Chapter 2, I will discuss development of a yeast-based gRNA screen and its logical successor, En Mass Evaluation of RNA Guides (EMERGe), an NGS-based screen. In Chapter 3, I will describe the validation and *in vitro* application of results stemming from the EMERGe screen. Finally, in Chapter 4, I will describe the cellular applications of EMERGe-derived gRNAs. I will also describe the results from *in cellula* testing of chemical modifications that enable ADAR editing within a 5'-G context, in Chapter 4.

## 1.6: References


(1) Grice, L. F.; Degnan, B. M. The origin of the ADAR gene family and animal RNA editing. *BMC Evol Biol* **2015**, *15* (1), 4. DOI: 10.1186/s12862-015-0279-3 From NLM Medline.
(2) Gott, J. M.; Emeson, R. B. Functions and mechanisms of RNA editing. *Annu Rev Genet* **2000**, *34*, 499-531. DOI: 10.1146/annurev.genet.34.1.499 From NLM Medline.
(3) Nishikura, K. Functions and regulation of RNA editing by ADAR deaminases. *Annu Rev Biochem* **2010**, *79*, 321-349. DOI: 10.1146/annurev-biochem-060208-105251 From NLM Medline.
(4) Bass, B. L. Double-stranded RNA as a template for gene silencing. *Cell* **2000**, *101* (3), 235-238. DOI: 10.1016/s0092-8674(02)71133-1 From NLM Medline.
(5) Song, B.; Shiromoto, Y.; Minakuchi, M.; Nishikura, K. The role of RNA editing enzyme ADAR1 in human disease. *Wiley Interdiscip Rev RNA* **2022**, *13* (1), e1665. DOI: 10.1002/wrna.1665 From NLM Medline.
(6) Erdmann, E. A.; Mahapatra, A.; Mukherjee, P.; Yang, B.; Hundley, H. A. To protect and modify double-stranded RNA–the critical roles of ADARs in development, immunity and oncogenesis. *Critical Reviews in Biochemistry and Molecular Biology* **2021**, *56* (1), 54-87. DOI: 10.1080/10409238.2020.1856768.
(7) Schaffer, A. A.; Levanon, E. Y. ALU A-to-I RNA Editing: Millions of Sites and Many Open Questions. *Methods Mol Biol* **2021**, *2181*, 149-162. DOI: 10.1007/978-1-0716-0787-9_9 From NLM Medline.

(8) Ulbricht, R. J.; Emeson, R. B. One hundred million adenosine-to-inosine RNA editing sites: hearing through the noise. *Bioessays* **2014**, *36* (8), 730-735. DOI: 10.1002/bies.201400055 From NLM Medline.
(9) Bazak, L.; Haviv, A.; Barak, M.; Jacob-Hirsch, J.; Deng, P.; Zhang, R.; Isaacs, F. J.; Rechavi, G.; Li, J. B.; Eisenberg, E.; Levanon, E. Y. A-to-I RNA editing occurs at over a hundred million genomic sites, located in a majority of human genes. *Genome Res* **2014**, *24* (3), 365-376. DOI: 10.1101/gr.164749.113 From NLM Medline.
(10) Daniel, C.; Lagergren, J.; Ohman, M. RNA editing of non-coding RNA and its role in gene regulation. *Biochimie* **2015**, *117*, 22-27. DOI: 10.1016/j.biochi.2015.05.020 From NLM Medline.
(11) Daniel, C.; Silberberg, G.; Behm, M.; Ohman, M. Alu elements shape the primate transcriptome by cis-regulation of RNA editing. *Genome Biol* **2014**, *15* (2), R28. DOI: 10.1186/gb-2014-15-2-r28 From NLM Medline.
(12) Pujantell, M.; Riveira-Munoz, E.; Badia, R.; Castellvi, M.; Garcia-Vidal, E.; Sirera, G.; Puig, T.; Ramirez, C.; Clotet, B.; Este, J. A.; Ballana, E. RNA editing by ADAR1 regulates innate and antiviral immune functions in primary macrophages. *Sci Rep* **2017**, *7* (1), 13339. DOI: 10.1038/s41598-017-13580-0 From NLM Medline.
(13) Hartner, J. C.; Schmittwolf, C.; Kispert, A.; Muller, A. M.; Higuchi, M.; Seeburg, P. H. Liver disintegration in the mouse embryo caused by deficiency in the RNA-editing enzyme ADAR1. *J Biol Chem* **2004**, *279* (6), 4894-4902. DOI: 10.1074/jbc.M311347200 From NLM Medline.
(14) Gannon, H. S.; Zou, T.; Kiessling, M. K.; Gao, G. F.; Cai, D.; Choi, P. S.; Ivan, A. P.; Buchumenski, I.; Berger, A. C.; Goldstein, J. T.; et al. Identification of ADAR1 adenosine deaminase dependency in a subset of cancer cells. *Nat Commun* **2018**, *9* (1), 5450. DOI: 10.1038/s41467-018-07824-4 From NLM Medline.
(15) Jung, S.; Ballheimer, Y. E.; Brackmann, F.; Zoglauer, D.; Geppert, C. I.; Hartmann, A.; Trollmann, R. Seizure-induced neuronal apoptosis is related to dysregulation of the RNA-edited GluR2 subunit in the developing mouse brain. *Brain Res* **2020**, *1735*, 146760. DOI: 10.1016/j.brainres.2020.146760 From NLM Medline.
(16) Rees, H. A.; Liu, D. R. Base editing: precision chemistry on the genome and transcriptome of living cells. *Nature Reviews Genetics* **2018**, *19* (12), 770-788. DOI: 10.1038/s41576-018-0059-1.
(17) Monian, P.; Shivalila, C.; Lu, G.; Shimizu, M.; Boulay, D.; Bussow, K.; Byrne, M.; Bezigian, A.; Chatterjee, A.; Chew, D.; et al. Endogenous ADAR-mediated RNA editing in non-human primates using stereopure chemically modified oligonucleotides. *Nat Biotechnol* **2022**, *40* (7), 1093-1102. DOI: 10.1038/s41587-022-01225-1 From NLM Medline.
(18) Yi, Z.; Zhao, Y.; Yi, Z.; Zhang, Y.; Tang, G.; Zhang, X.; Tang, H.; Zhang, W.; Zhao, Y.; Xu, H.; et al. Utilizing AAV-mediated LEAPER 2.0 for programmable RNA editing in non-human primates and nonsense mutation correction in humanized Hurler syndrome mice. *Genome Biol* **2023**, *24* (1), 243. DOI: 10.1186/s13059-023-03086-6 From NLM Publisher.
(19) Lehmann, K. A.; Bass, B. L. Double-stranded RNA adenosine deaminases ADAR1 and ADAR2 have overlapping specificities. *Biochemistry* **2000**, *39* (42), 12875-12884. DOI: 10.1021/bi001383g From NLM Medline.
(20) 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. DOI: 10.1038/s41594-020-0398-4 From NLM Medline.

(21) Phelps, K. J.; Tran, K.; Eifler, T.; Erickson, A. I.; Fisher, A. J.; Beal, P. A. Recognition of duplex RNA by the deaminase domain of the RNA editing enzyme ADAR2. *Nucleic Acids Res* **2015**, *43* (2), 1123-1132. DOI: 10.1093/nar/gku1345 From NLM Medline.
(22) Doherty, E. E.; Karki, A.; Wilcox, X. E.; Mendoza, H. G.; Manjunath, A.; Matos, V. J.; Fisher, A. J.; Beal, P. A. ADAR activation by inducing a syn conformation at guanosine adjacent to an editing site. *Nucleic Acids Res* **2022**, *50* (19), 10857-10868. DOI: 10.1093/nar/gkac897 From NLM Medline.
(23) Matthews, M. M.; Thomas, J. M.; Zheng, Y.; Tran, K.; Phelps, K. J.; Scott, A. I.; Havel, J.; Fisher, A. J.; Beal, P. A.; 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. DOI: 10.1038/nsmb.3203.Structures.
(24) Goodman, R. A.; Macbeth, M. R.; Beal, P. A. *Adenosine Deaminases Acting on RNA (ADARs) and A-to-I Editing (Google eBook)*; 2011.
(25) Haudenschild, B. L.; Maydanovych, O.; Veliz, E. A.; Macbeth, M. R.; Bass, B. L.; Beal, P. A. A transition state analogue for an RNA-editing reaction. *J Am Chem Soc* **2004**, *126* (36), 11213-11219. DOI: 10.1021/ja0472073 From NLM Medline.
(26) Thuy-Boun, A. S.; Thomas, J. M.; Grajo, H. L.; Palumbo, C. M.; Park, S.; Nguyen, L. T.; Fisher, A. J.; Beal, P. A. Asymmetric dimerization of adenosine deaminase acting on RNA facilitates substrate recognition. *Nucleic Acids Research* **2020**, *48* (14), 7958-7972. DOI: 10.1093/nar/gkaa532.
(27) Mendoza, H. G.; Matos, V. J.; Park, S.; Pham, K. M.; Beal, P. A. Selective Inhibition of ADAR1 Using 8-Azanebularine-Modified RNA Duplexes. *Biochemistry* **2023**, *62* (8), 1376-1387. DOI: 10.1021/acs.biochem.2c00686 From NLM Medline.
(28) Kuttan, A.; Bass, B. L. Mechanistic insights into editing-site specificity of ADARs. *Proceedings of the National Academy of Sciences of the United States of America* **2012**, *109* (48). DOI: 10.1073/pnas.1212548109.
(29) Doherty, E. E.; Wilcox, X. E.; van Sint Fiet, L.; Kemmel, C.; Turunen, J. J.; Klein, B.; Tantillo, D. J.; Fisher, A. J.; Beal, P. A. Rational Design of RNA Editing Guide Strands: Cytidine Analogs at the Orphan Position. *J Am Chem Soc* **2021**, *143* (18), 6865-6876. DOI: 10.1021/jacs.0c13319 From NLM Medline.
(30) Monteleone, L. R.; Matthews, M. M.; Palumbo, C. M.; Chiang, Y.; Fisher, A. J.; Beal, P. A.; Monteleone, L. R.; Matthews, M. M.; Palumbo, C. M.; Thomas, J. M.; et al. Article A Bump-Hole Approach for Directed RNA Editing Article A Bump-Hole Approach for Directed RNA Editing. *Cell Chemical Biology* **2019**, *26* (2), 269-277.e265. DOI: 10.1016/j.chembiol.2018.10.025.
(31) Ramaswami, G.; Li, J. B. RADAR: a rigorously annotated database of A-to-I RNA editing. *Nucleic Acids Res* **2014**, *42* (Database issue), D109-113. DOI: 10.1093/nar/gkt996 From NLM Medline.
(32) Yi, Z.; Qu, L.; Tang, H.; Liu, Z.; Liu, Y.; Tian, F.; Wang, C.; Zhang, X.; Feng, Z.; Yu, Y.; et al. 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. DOI: 10.1038/s41587-021-01180-3 From NLM Medline.
(33) Sinnamon, J. R.; Jacobson, M. E.; Yung, J. F.; Fisk, J. R.; Jeng, S.; McWeeney, S. K.; Parmelee, L. K.; Chan, C. N.; Yee, S. P.; 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. DOI: 10.1073/pnas.2206053119 From NLM Medline.
(34) Sinnamon, J. R.; Kim, S. Y.; Corson, G. M.; Song, Z.; Nakai, H.; Adelman, J. P.; Mandel, G. Site-directed RNA repair of endogenous *MECP2* RNA in neurons. *Proceedings of the National Academy of Sciences of the United States of America* **2017**, *114* (44), E9395-E9402. DOI: 10.1073/pnas.1715320114.
(35) Diaz Quiroz, J. F.; Ojha, N.; Shayhidin, E. E.; De Silva, D.; Dabney, J.; Lancaster, A.; Coull, J.; Milstein, S.; Fraley, A. W.; Brown, C. R.; Rosenthal, J. J. C. Development of a selection assay for small guide RNAs that drive efficient site-directed RNA editing. *Nucleic Acids Res* **2023**, *51* (7), e41. DOI: 10.1093/nar/gkad098 From NLM Medline.
(36) Merkle, T.; Merz, S.; Reautschnig, P.; Blaha, A.; Li, Q.; Vogel, P.; Wettengel, J.; Li, J. B.; Stafforst, T. Precise RNA editing by recruiting endogenous ADARs with antisense oligonucleotides. *Nat Biotechnol* **2019**, *37* (2), 133-138. DOI: 10.1038/s41587-019-0013-6 From NLM Medline.
(37) Eggington, J. M.; Greene, T.; Bass, B. L. Predicting sites of ADAR editing in double-stranded RNA. *Nat Commun* **2011**, *2*, 319. DOI: 10.1038/ncomms1324 From NLM Medline.

# Chapter 2

# Development of High-throughput Assays for ADAR gRNAs

*The biochemical studies with MECP2 R255X were part of a joint project and were completed by Prince Salvador. This chapter contains excepts from the full manuscript which was published in ACS Chemical Biology in April 2023.* [1]

## 2.1: Introduction

Saturation mutagenesis, fluorescent-activated cell sorting, and next-generation sequencing (Sat-FACS-Seq) has been a powerful tool in the Beal lab for screening amino acids at certain positions in the protein that are vital for ADAR function. [2] Sat-FACS-Seq involves the use of two types of plasmids for a yeast expression system. The first type of plasmid expresses the ADAR protein containing all possible amino acid mutations for screening, while the second one is a reporter plasmid. The reporter plasmid contains a stop (UAG) codon that, when edited by ADAR, is converted to a tryptophan (UGG) codon. This then allows for the translation of a yeast-enhanced green fluorescent protein (yeGFP) reporter that is directly downstream of the UAG edit side. When edited, the resulting cells will fluoresce for subsequent FACS. These fluorescent cells are then pooled, and the most fluorescent cells are selected. These highly active cells are subsequently subjected to next-generation sequencing to determine the mutation responsible for their high activity. By using Sat-FACS-Seq on ADAR1, residues typically seen for metal binding were observed to be highly conserved along the protein's surface. This led to further studies that resulted in the discovery of a second metal-binding site on the surface of ADAR1[3]. These results also enabled improvements in ADAR1 structure modeling and protein purification. When Sat-FACS-Seq was applied to ADAR2, critical RNA binding residues in ADAR2's 5' binding loop were identified, enhancing our understanding of RNA recognition through this binding loop. [2] The

ability to test libraries of ADAR mutants against a single dsRNA substrate has significantly improved our understanding of protein-RNA interactions.

Therefore, if Sat-FACS-Seq could be employed to interrogate gRNAs, we could gain a better understanding of sequence-specific effects that constrain ADARs. This would make Sat-FACS-Seq a foundational approach for assaying variable RNAs in the context of a single protein. Instead of applying the saturation mutagenesis part of Sat-FACS-Seq to an ADAR plasmid, this process can be carried out on a plasmid containing a variable RNA target. This would enable us to simultaneously test thousands of possibilities of RNA targets to determine ADAR editing efficiency. The primary limitation of this methodology is that it can only be applied to hundreds of sequences. To test a library with a thousand possibilities more than 10,000 colonies would have to be obtained hit this 95% confidence in library coverage. Beyond a certain point, testing becomes overly cumbersome since only a few hundred colonies can be prepared at a time.

By building upon the foundations of Sat-FACS-Seq, we created novel methodology called En Masse Evaluation of RNA Guides (EMERGe).[1] This methodology uses close to a million unique gRNAs to query an even larger sequence space. This entirely *in vitro* method eliminates the cellular limitations imposed by Sat-FACS-Seq and enables rapid testing of $10^6$ gRNAs simultaneously. EMERGe's unbiased approach facilitates the discovery of gRNAs that would never be explored using rational design experiments.

## 2.2: Results

### 2.2.1: Developing a cell-based assay to discover novel gRNAs for ADAR.

Our lab has previously employed Sat-FACS-Seq to investigate combinations of mutant ADARs in a high-throughput manner (Figure 2.1).[2] To utilize this methodology, a new reporter was developed. By inserting a hairpin RNA substrate in-frame between the start codon and the yeGFP transcript, an editable reporter can be created (Figure 2.2A). This hairpin can be varied to

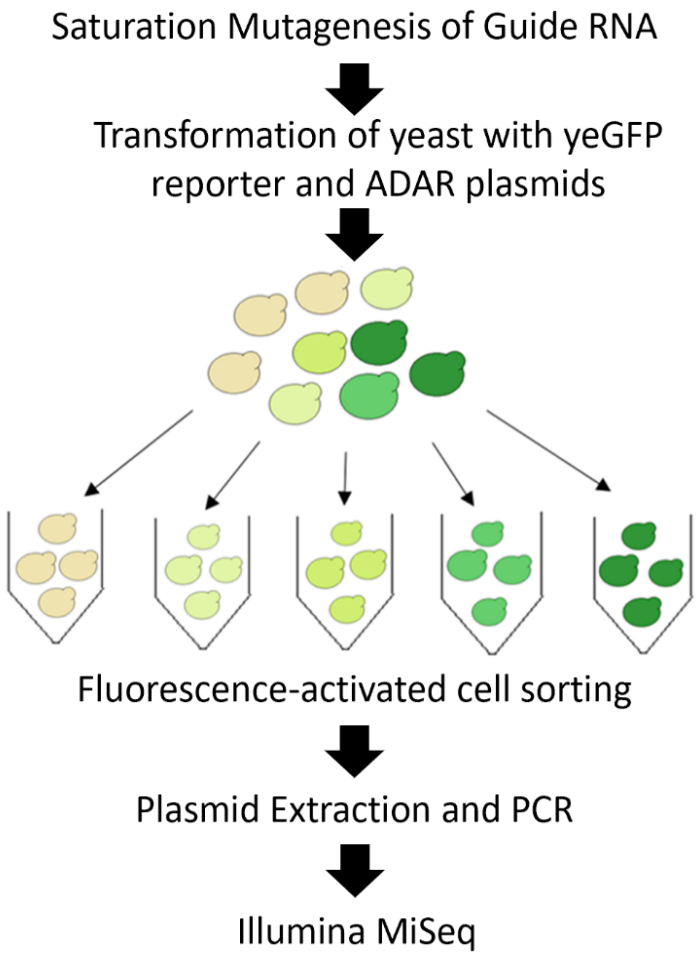


Figure 2.1: A general workflow for Sat-FACS-Seq.

This assay allows for cells to be selected for through FACS if the reporter plasmid gets edited by ADAR.

accommodate different target sequences and corresponding gRNAs (Figure 2.2B & C). Three different ECS motifs were introduced opposite the editing site, denoted as $N_3$, $N_4$, and $N_6$. These three motifs provide a sequence space ranging from 64 to 4096 possible sequences.

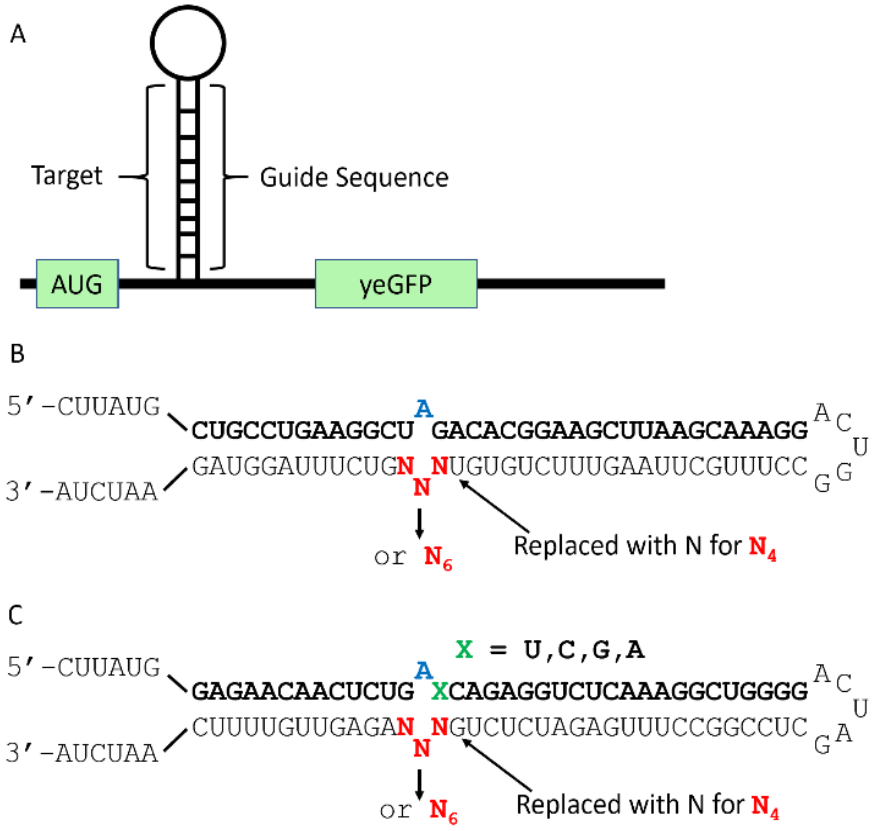


Figure 2.2: Sat-FACS-Seq hairpin designs.

A: The reporter transcript contains a hairpin down stream and in frame of the start codon. B: A reporter library corresponding to the sequence present in the mRNA for human *MECP2* bearing the W104X mutation (bold), ECS randomized regions in red. The target A is shown in blue. C: A reporter library corresponding to the sequence present in mRNA for mIDUA bearing a 5'G modified W392X mutation (bold), ECS randomized regions in red. The target A is shown in blue. Variable 3' nearest neighbor site shown in green. The N6 variable region adds 3 nt to N3 instead of replacing 6 nt within the hairpin.

The primary goal of this assay was to identify an encodable sequence compatible with the bump-hole ADAR system (Figure 2.3A). By employing E488X bulky mutant ADARs (X = F, W, Y), an

orthogonal editing system can be utilized.[4] Typically, these mutants induce a steric clash within ADAR's ideal context of an A:C mismatch, and a chemically modified guide (reduced abasic site) is employed to facilitate editing. This assay was aimed at mitigating this steric clash through the ECS variable regions, ultimately allowing for an encodable version of this reduced abasic guide. A mutation that causes the neurodevelopmental disorder Rett syndrome,[5] *MECP2* W104X, was used as a model sequence context (Figure 2.2B).

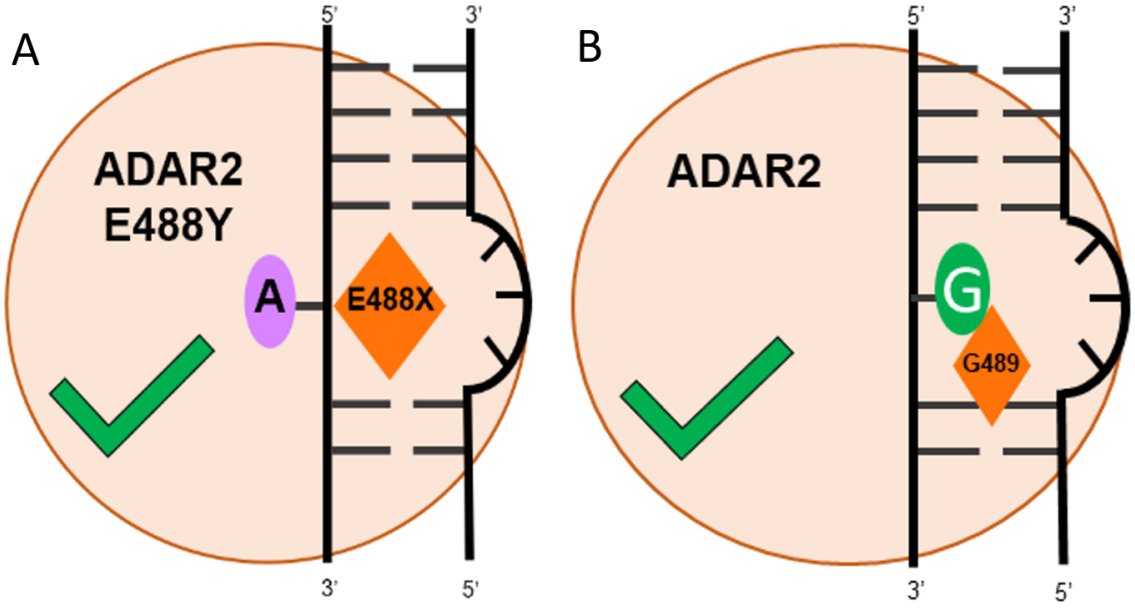


Figure 2.3: Rationale for using encodable gRNAs for promoting ADAR editing at in difficult contexts.

A: A model for enabling editing using a bump-hole ADAR system. The E488X (X=Y) mutant uses a bulky residue to enable specific editing. If a bugle is discovered, that may allow space for the bulky residue (E488Y), enabling editing can occur. B: A model for enabling editing with a poor nearest neighbor context. If a bugle is discovered within the ECS there may be space to alleviate the strained G489 interaction.

An additional objective was to utilize this sequence to discover an encodable guide capable of enabling editing at difficult to edit sites.[6] Typically, a 5'G clashes with the G489 residue. If there is a perturbation across from this site, there may be space for both the 5'G and the G489

residue (Figure 2.3B). For this assay, sequences containing a 5'G could be tested, with variation in the 3' nearest neighbor. This approach would yield an encodable gRNA that facilitates editing through non-standard base-paired secondary structures. A modified version of the mouse alpha-L-iduronidase (mIDUA) transcript was used to represent ADAR's editing with a 5'G nearest neighbor (Figure 2.2C).

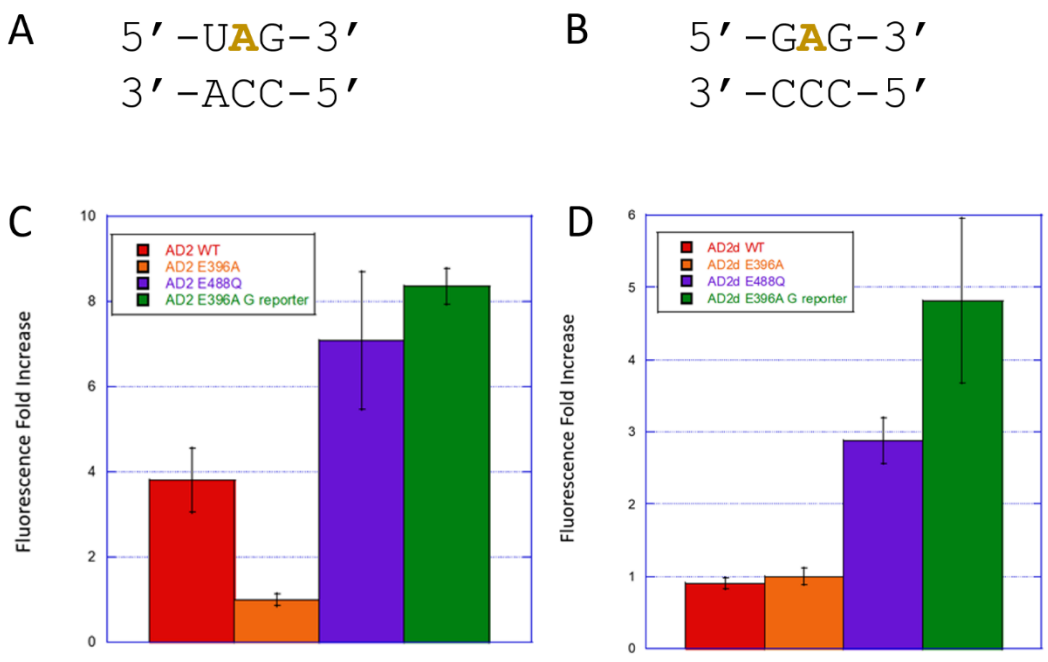


Figure 2.4: yeGFP assays results with the hairpin designs using ideal contexts.

A: The ideal context for the *MECP2* W104X target. Target A bolded in Yellow. B: The ideal context for the 5'G modified mIDUA W392X target. Target A bolded in Yellow. C: The relative fluorescence for yeast cells expressing listed full length ADAR2 conditions with the *MECP2* W104X reporter. D: The relative fluorescence for yeast cells expressing listed deaminase domain ADAR2 conditions with the 5'G modified mIDUA W392X reporter. G reporter contains a G at the edit site to represent a 100% editing condition. C and D: Catalytically dead ADAR2 (E396A) was used for a baseline value of 1. G reporter replaces the target A with a G to represent a 100% editing condition. Data are plotted as the mean ± s.d. from three independent experiments.

Minor alterations were made to these hairpins to verify that these hairpins were functional ADAR substrates. Both hairpins had their $N_3$ region replaced with the ideal ECS[6] (Figure

2.4A & B). These updated substrates and ADAR plasmids were transformed into yeast cells. Using double deficient growth media colonies that contained both plasmids were selected. Those selected for colonies were used to inoculate larger broth growths and the resulting growths were induced for fluorescence yeGFP assays.

It was observed that the growths containing the hyperactive mutant of ADAR (E488Q) were indeed more active than their wild type counterparts (Figure 2.4C & D). In the case of our 5'G reporter, little to no editing was observed with the wild type protein. Additionally, two control groups were included. The first control group contained catalytically dead ADAR (E396A) with the reporter plasmid containing an 'A' at the edit site. The second control group featured catalytically dead ADAR with a reporter that had a 'G' at the edit site, called the G reporter. This second control group served as a 100% editing baseline. These results align with the current understanding of ADAR2. [6]

However, the main limitation of this methodology was the number of colonies required to proceed. Previous completions of Sat-FACS-Seq necessitated a 10-fold excess of colonies compared to plasmids. [2] This means that if there are 64 possibilities within the $N_3$ dataset, 640 colonies would need to be collected. This became a daunting barrier to overcome for the $N_6$ library where the coverage would require 40,960 colonies to be collected.

### 2.2.2: Development of an *in vitro* model for studying features of dsRNA for editing by ADAR

Given that transformation efficiency was the bottleneck in Sat-FACS-Seq, I developed a completely *in vitro* methodology. [1] I created a new screening assay, EMERGe, to identify gRNA sequences that facilitate ADAR editing at challenging sites. EMERGe uses a hairpin RNA substrate

with the target editing site covalently linked to a sequence-randomized guide region (Figure 2.5). 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 containing the edit site and 10 variable nucleotides ($N_{10}$) opposite the edit site; and (3) a fully complementary 10 bp region 3' to the edit site (Figure 2.5). Additionally, a loop linker 3' to the edit site allows for hairpin formation. This substrate library (Figure 2.5, ADAR substrate library) mimics the hairpin design used in Sat-FACS-Seq (Figure 2.2C) for a completely *in vitro* assay.

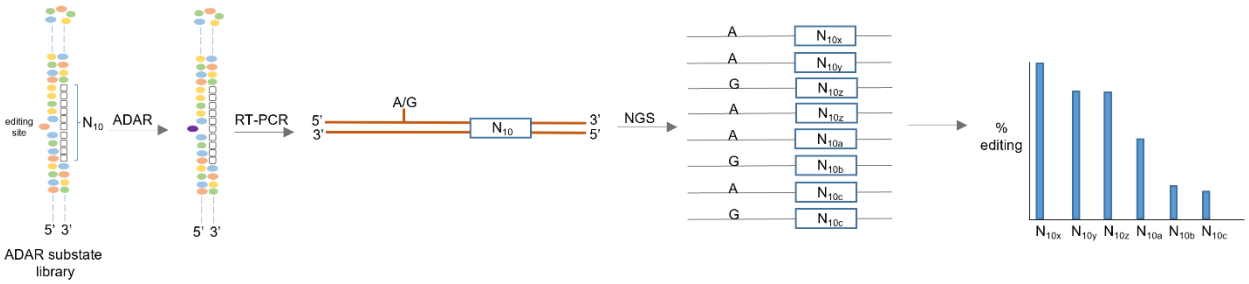


Figure 2.5: En Masse Evaluation of RNA Guides (EMERGe).

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 N10 sequence that has either A or G at the editing site. These values are then used to calculate % editing for each sequence present.

The hairpin sequence of the ADAR substrate can be varied to allow for the discovery of enabling guide strands for different therapeutically-relevant targets, such as another Rett syndrome mutation, *MECP2* R168X[5] (Figure 2.6A). 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.[5, 7] However, the adjacent 5' G makes this a difficult-to-edit site for ADARs.[6] 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 NGS.

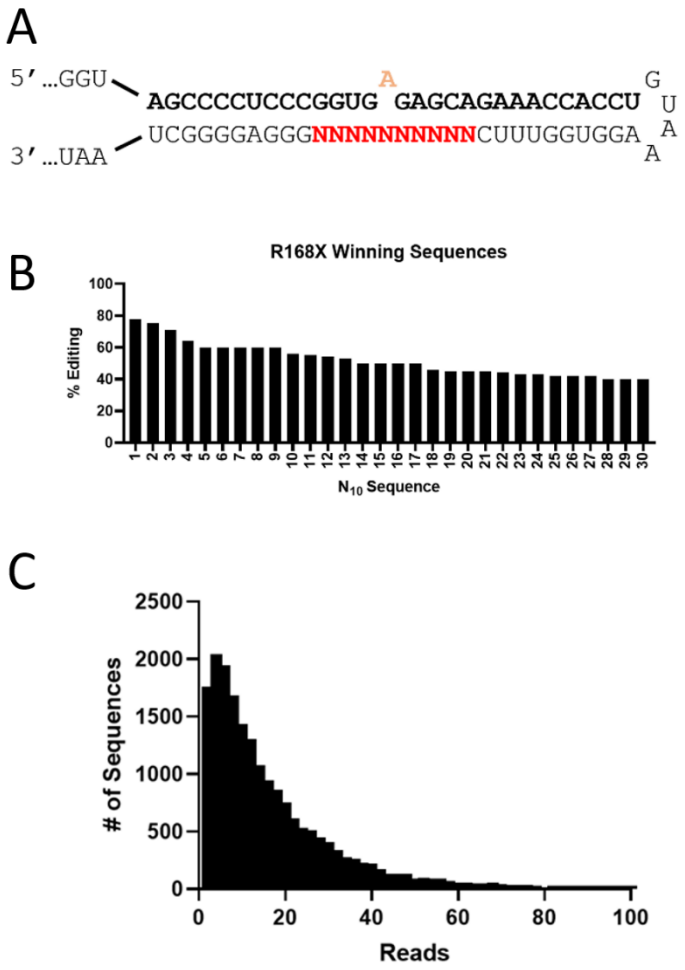


Figure 2.6: Library design and results from data processing the *MECP2* R168X target.

A: 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. B: Calculated % editing by ADAR2 for the top 30 sequences having at least 10 reads from the R168X EMERGe dataset. C: Histogram of the sequence read landscape for sequences containing less than 100 reads.

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 2.6B). In order to minimize false positives, I 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 2.6A, 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 2.1) 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 2.2).

In addition to observing the top sequences, a sequence landscape can be generated (Figure 2.6C). This landscape proves useful for quickly assessing the depth of the datasets. It is observed that the majority of unique sequences have between 1 and 20 reads, which provides a certain level of confidence derived from the NGS run. The presence of sequences with higher read counts enables more extensive data analysis.

One of the most challenging sites for ADAR editing is when the target A is within a 5'-GAA context.[6] 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. 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, I generated a new hairpin bearing the *MECP2* R255X sequence as the target. For this hairpin, I also replaced the GUAA loop present in the R168X hairpin with a longer and potentially more flexible loop sequence (Figure 2.7A). With the help of Prince Salvador, this new hairpin was subjected to an EMERGe screen and editing of the top 30 sequences was identified (Figure 2.7B). The dataset from this screen had a lower read depth per $N_{10}$, so the cut-off was

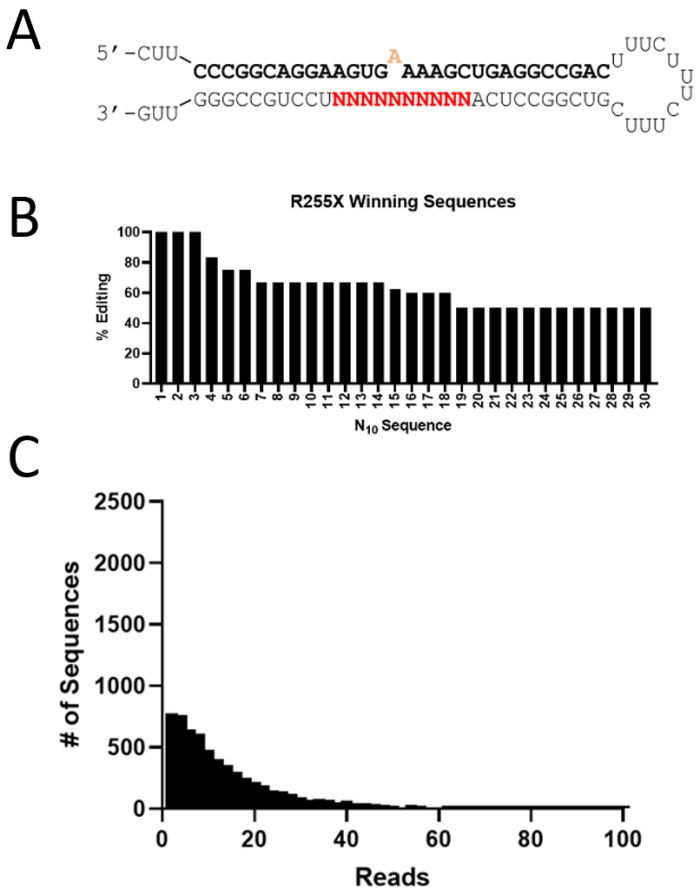


Figure 2.7: Library design and results from data processing the *MECP2* R255X target.

A: An example of an EMERGe library corresponding to the sequence present in the mRNA for human *MECP2* bearing the R255X mutation (bold) showing the $N_{10}$ sequence randomized region in red. The target A is shown in yellow. B: Calculated % editing by ADAR2 for the top 30 sequences having at least 10 reads from the R255X EMERGe dataset. C: Histogram of the R255X EMERGe sequence read landscape for sequences containing less than 100 reads.

lowered to three reads per sequence for consideration (Table 2.3). As mentioned earlier, the significance of having a deep read depth was crucial in identifying enabling ECSs. However, lowering the read cutoff carried the risk of including numerous false positives in the final dataset. The sequence landscape shows that the R255X dataset had significantly fewer sequences than the R168X dataset (Figure 2.7C). Some of these false positives may have come through the dataset as 100% editing but low reads (Table 2.3). This was not present in our previous dataset and may be an artifact of low read depth.

I also carried out an EMERGe screen with a hairpin bearing the *MECP2* R270X site (Figure 2.8A). 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 2.8C). Again, the top 30 sequences from the screen were selected (Table 2.4) and sorted according to editing efficiencies (Figure 2.8B). This increased read NGS methodology had allowed for significantly more sequences existing at high read depths (Figure 2.8C). Additionally, strand invading primers were used to destabilize the hairpin structure, which in turn would promote transcription.

The EMERGe system has the ability to evaluate a large number of ADAR editing substrates. I have developed two more EMERGe screens and they are currently underway by other Beal laboratory members. The first focuses on using the EMERGe screen to query which $N_{10}$'s can allow for editing with bulky ADAR mutants (Figure 2.9A). The second is an EMERGe screen aimed at

discovering enabling sequences that exist at the dsRBD binding site (Figure 2.9B). These screens will play a pivotal role in discovering sequences that facilitate ADAR editing in various unique contexts.

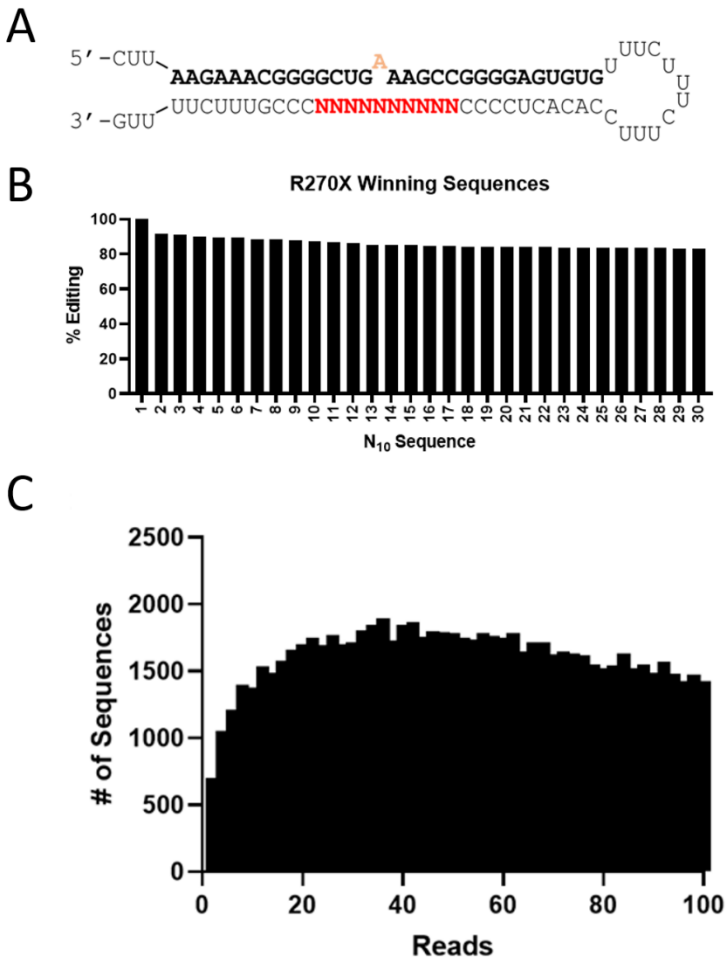


Figure 2.8: Library design and results from data processing the *MECP2* R270X target.

A: An example of an EMERGe library corresponding to the sequence present in the mRNA for human *MECP2* bearing the R255X mutation (bold) showing the $N_{10}$ sequence randomized region in red. The target A is shown in yellow. B: Calculated % editing by ADAR2 for the top 30 sequences having at least 10 reads from the R270X EMERGe dataset. C: Histogram of the R270X EMERGe sequence read landscape for sequences containing less than 100 reads.

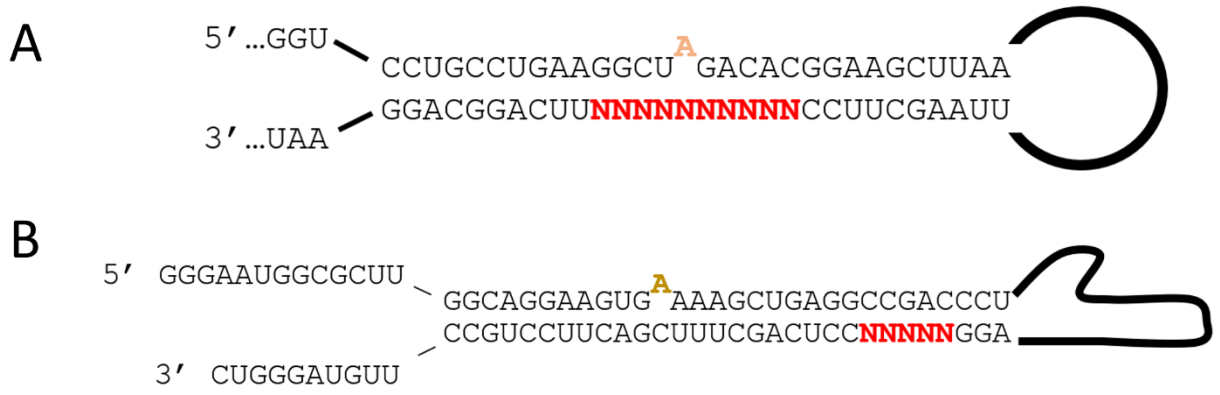


Figure 2.9: Hairpin designs for upcoming EMERGe assays.

A: An example of an EMERGe library corresponding to the sequence present in the mRNA for human *MECP2* bearing the W104X mutation (top strand) showing the $N_{10}$ sequence randomized region in red. The target A is shown in yellow. The loop uses the improved design used for R255X and R270X screens. This hairpin will be deaminated using hADAR2 E488F bulky mutant. B: An example of an EMERGe library corresponding to the sequence present in the mRNA for human *MECP2* bearing the R255X mutation (top strand) showing a new N5 sequence randomized region in red. The target A is shown in yellow. This loop uses a continuing 40 nt of the R255X transcript. This hairpin will be deaminated using hADAR2-RD.

**2.2.3: Using NGS datasets to query single nucleotide variants within select sequences.**

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), I generated an activity heat map that shows how a single nucleotide change in the $N_{10}$ region of R168X-5 affects editing efficiency (Figure 2.10). Through this, I 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 2.10). However, perturbations in the terminal regions of the selected sequence, particularly at the 5’ end, were not as detrimental. This NGS-derived structure activity relationship (SAR) allows for discovery of key positions within the gRNA.

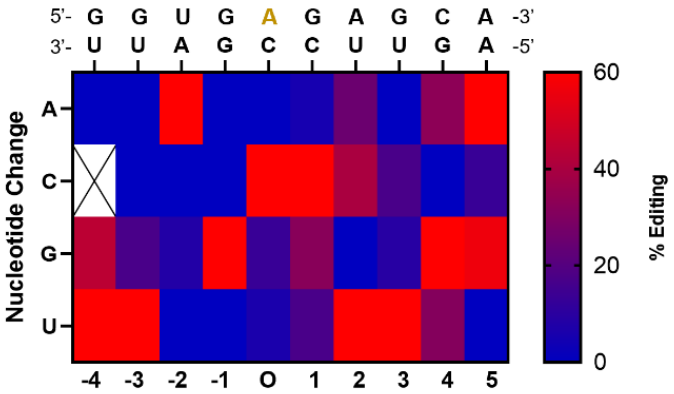


Figure 2.10: Heat Map of R168X-5

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.

NGS-derived SAR also has the potential to uncover sequence motifs. By comparing sequences that are similar yet distinct, I can identify unique motifs that facilitate editing. If we compare editing between R270X-6, R270X-7, and R270X-24, we can observe a distinct motif within the data (Figure 2.11). There is a set of three nucleotides, 3’-CGG, that is conserved across these three datasets at the -2, -1, and orphan positions. This provides a basis for a motif to be tested in subsequent *in vitro* deaminations, R270X-24a. As the first step in logical design after EMERGe, the suffix “a” was added to denote the change from an EMERGe-derived sequence. In addition to NGS-derived SAR, single nucleotide variants (SNVs) can also be employed to establish confidence scores. This is accomplished by calculating the number of SNVs that contain editing for a single $N_{10}$ sequence. For example, if we consider every single nucleotide variant of R168X-5,

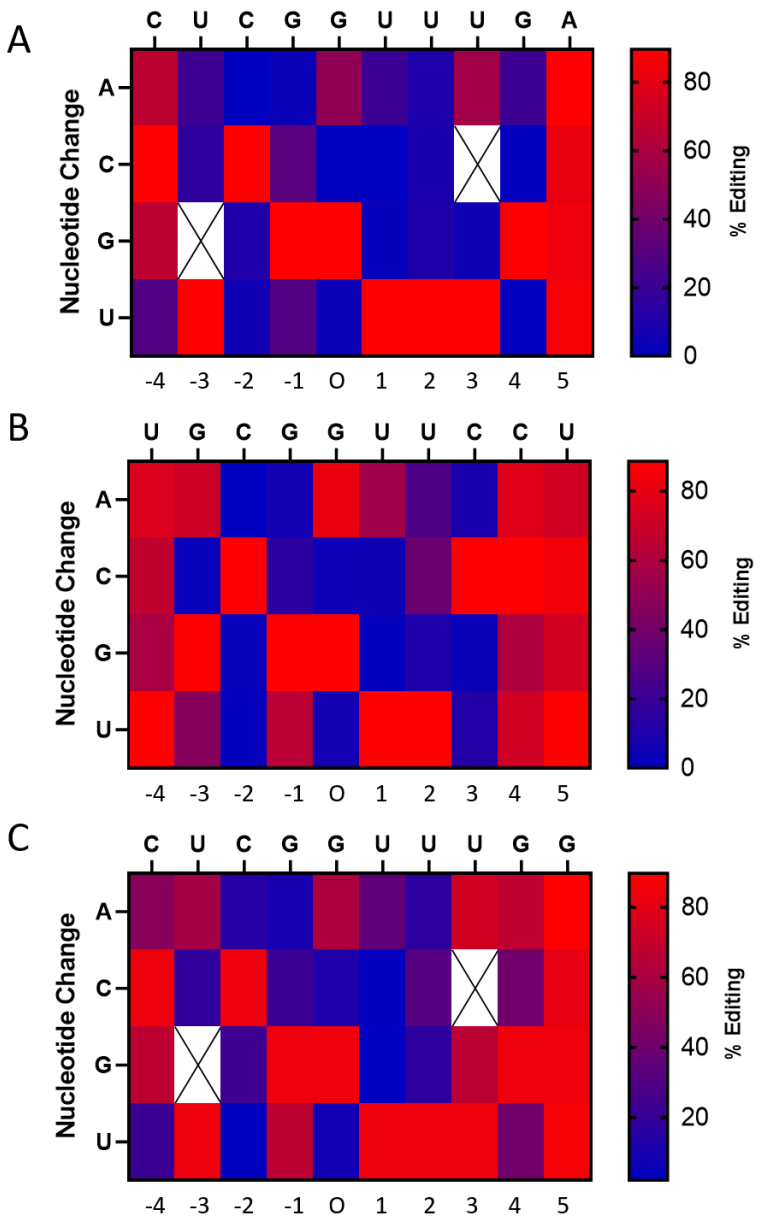


Figure 2.11: Heat Maps of R270X-6, R270X-7, and R270X-24

Percent editing calculated for all the single nucleotide variants of the R270X hairpins present in the EMERGe NGS datasets. A: R270X-6 B: R270X-7 C: R270X-24. Sequences above the heat maps are listed 3' to 5'. Numbers below the heat map correspond to the nucleotide position in the guide strand relative to the orphan (O) position opposite the target A.

we would find that there are 21 SNVs with greater than 10% editing. A confidence score can be

calculated for each of the top N10 sequences. Using this confidence score, each sequence can be

queried to determine if it is a hit sequence. Sequences with high confidence scores indicate the presence of SNVs within the dataset. SNVs are indicative that an editing sequence is not a false positive. Therefore, the higher the confidence score, the more likely it is for a sequence to promote editing. Utilizing confidence scores, editing percentages, and read counts allows for a more informed decision regarding which EMERGe $N_{10}$ sequences to advance to *in vitro* verification.

## 2.3: Discussion

It is clear that ADARs have 5', 3', and opposite base preferences for editing adenosines in A-form duplexes. [8, 9] 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. [4, 10-14] 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. [6, 15, 16] Furthermore, many therapeutically relevant target adenosines for directed RNA editing applications do not conform to ADAR's known preferred nearest neighbors. [5, 7] 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. Recently, a methodology was published that also used a hairpin substrate (Diaz Quiroz, J. Et al) discussed in Chapter 1. [17] While similar to EMERGe, their library was biased by previously published ADAR preferences. [6, 16, 18, 19] In contrary to this recent publication, other published methodologies require transfection of plasmid libraries, limiting the size of libraries that can be practically screened. [2, 20, 21] Sat-FACS-Seq for gRNAs ended up being one of those screens. When using large libraries of gRNA hairpins, there are too many limitations

imposed by the required library coverage. This makes completing the assay time-prohibitive, as it would require 100 days of perfect transformations, assuming 400 colonies a day for the $N_6$ dataset. This is dramatically slower and less comprehensive than using EMERGe.

Here, we describe a type of screen that does not require plasmid transformations, 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.

Once the NGS data is deconvoluted, we have close to a million sequences that can be further subjected to computational processing. The processing, involving NGS-derived SAR and confidence scores, narrows down the subset of guides to be tested *in vitro*. The significance of completing data processing on these NGS files cannot be overstated. Aligning the data is essential for extracting meaningful information from NGS. Two key regions, (1) the editing site (A or G) and (2) the $N_{10}$ region, represent the primary points of interest. While not shown here, the alignment can be altered to view different editing sites. For example, we can look at off-target A's to view how an $N_{10}$ will effect editing elsewhere. In addition to alignment, a deep understanding of the

data allows for the development of NGS-derived SAR studies for EMERGe samples. This, in turn, facilitates continuous exploration of the data throughout the experiments.

Many more sequences can be compared simultaneously through NGS-derived SAR. This allows for the discovery of motifs to test, in addition to specific sequences to verify. These datasets are not only useful for verification but may also prove valuable in future experiments. If a motif is discovered within a new screen, the old screens can be queried. This enables the discovery of ECS preferences between targets and could help validate a new motif. Moreover, aside from verifying new motifs, these extensive datasets could be applied to neural networks and deep learning models. If specific datasets are compiled, nucleotide interactions can be compared across datasets which would enable the application of artificial intelligence (AI) to these data and enable predictive gRNA designs.

## 2.4: Methods

**2.4.1: Sat-FACS-Seq Reporter Plasmid Generation.** Plasmids were digested using HindIII (NEB R0104S) and BglII (NEB R0144S). The digested plasmid was then purified by gel extraction using 0.5% agarose (Qiagen 28706). A dsDNA containing the variable regions was inserted using Gibson Assembly (E2611). The resulting ligation plasmid was transformed into XL10-Gold Ultracompetent Cells (Agilent 200314) and purified (Qiagen 27106). Sequences for mutant plasmids were confirmed by Sanger sequencing. Inserted dsDNA was purchased from IDT.

**2.4.2: Validation of yeGFP reporters for Sat-FACS-Seq.** Transformed yeast cells, containing both ADAR2 and reporter plasmids, were used to inoculate 10 mL of complete media (CM) -ura -trp +2% glucose media for overnight growth. The resulting culture (0.3 mL) was used to inoculate 30

mL of CM -ura -tip +3% glycerol +2 lactate media for another growth. Once this growth reached an $OD_{600}$ of 1-2 (~30 h) galactose was added for a final concentration of 3% for induction (10 h). After the induction, $10^8$ cells were collected from the growth and washed twice with phosphate buffered saline (PBS) and fluorescence was detected. Fluorescence detection was completed using an Optiplate-96 black, black opaque 96-well microplate (PerkinElmer) and CLARIOstar plate reader (BMG labtech), with excitation at 482/16 nm and emission at 520/10 $nm^2$.

**2.4.3: General Sat-FACS-Seq Methods.** All other methods not listed were repeated as published[2].

**2.4.4: Protein overexpression and purification.** hADAR2 with N-terminal His10 tag was 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.

**2.4.5: 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 single stranded DNA (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 MgCl2 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.

**2.4.6: 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 $N_{10}$ 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 $N_{10}$ sequences. A subset of sequences, based on overall reads, editing %, and computationally predicted secondary structure, was taken forward to verification.

## 2.5: Tables of processed, aligned, and sorted EMERGe Sequences

Table 2.1: R168X Top 30 Sequences

| Guide Name | Sequence 5' -> 3' | Edit Site: A Reads | Edit Site: G Reads | % Editing |
|---|---|---|---|---|
| R168X-1 | GGUUGGGAGC | 4 | 14 | 77.80% |
| R168X-2 | UGUUGGGAGU | 3 | 9 | 75.00% |
| R168X-3 | CACUGAGAUU | 5 | 12 | 70.60% |
| R168X-4 | GUUCUAUCGG | 9 | 16 | 64.00% |
| R168X-5 | AGUUCCGAUU | 4 | 6 | 60.00% |
| R168X-6 | CACUGAGAGU | 4 | 6 | 60.00% |
| R168X-7 | CACUUAUCGU | 4 | 6 | 60.00% |
| R168X-8 | CGUUGGGGGC | 4 | 6 | 60.00% |
| R168X-9 | UUUUCCGAUU | 4 | 6 | 60.00% |
| R168X-10 | GGUUCCGAUU | 8 | 10 | 55.60% |
| R168X-11 | UGGUUAUCGU | 9 | 11 | 55.00% |
| R168X-12 | UGGUUAACGU | 6 | 7 | 53.80% |
| R168X-13 | AGUUGAGAGU | 9 | 10 | 52.60% |
| R168X-14 | GCUCCGAUCG | 5 | 5 | 50.00% |
| R168X-15 | CGUUGUGGGC | 5 | 5 | 50.00% |
| R168X-16 | GCUCUAUCGG | 6 | 6 | 50.00% |
| R168X-17 | GUUGGUGAUC | 5 | 5 | 50.00% |
| R168X-18 | GGGUUUAAGC | 7 | 6 | 46.20% |
| R168X-19 | CACUGUGAUU | 18 | 15 | 45.50% |
| R168X-20 | GCCCUAUCGG | 12 | 10 | 45.50% |
| R168X-21 | GCUGGGUAUU | 11 | 9 | 45.00% |
| R168X-22 | AGUUGUGGGC | 9 | 7 | 43.80% |
| R168X-23 | ACCCUAUCGG | 8 | 6 | 42.90% |
| R168X-24 | CGUUGGGGUC | 12 | 9 | 42.90% |
| R168X-25 | GCUGGGUGUC | 7 | 5 | 41.70% |
| R168X-26 | GGCUGGGUGC | 7 | 5 | 41.70% |
| R168X-27 | UUCUGUGGGC | 7 | 5 | 41.70% |
| R168X-28 | AGUUGGGGUC | 9 | 6 | 40.00% |
| R168X-29 | CCCUGUGGCU | 9 | 6 | 40.00% |
| R168X-30 | CCCUUAUCUA | 6 | 4 | 40.00% |

Table 2.2: R168X Top 30 Sequences in a no ADAR control

| Guide Name | Sequence 5' -> 3' | Edit Site: A Reads | Edit Site: G Reads | % Editing |
|---|---|---|---|---|
| R168X-1 | GGUUGGGAGC | 736 | 1 | 0.10% |
| R168X-2 | UGUUGGGAGU | 10 | 0 | 0.00% |
| R168X-3 | CACUGAGAUU | 99 | 0 | 0.00% |
| R168X-4 | GUUCUAUCGG | 568 | 0 | 0.00% |
| R168X-5 | AGUUCCGAUU | 179 | 0 | 0.00% |
| R168X-6 | CACUGAGAGU | 10 | 0 | 0.00% |
| R168X-7 | CACUUAUCGU | 10 | 0 | 0.00% |
| R168X-8 | CGUUGGGGGC | 5 | 0 | 0.00% |
| R168X-9 | UUUUCCGAUU | 3 | 0 | 0.00% |
| R168X-10 | GGUUCCGAUU | 12 | 0 | 0.00% |
| R168X-11 | UGGUUAUCGU | 7 | 0 | 0.00% |
| R168X-12 | UGGUUAACGU | 8 | 0 | 0.00% |
| R168X-13 | AGUUGAGAGU | 20 | 0 | 0.00% |
| R168X-14 | GCUCCGAUCG | 3 | 0 | 0.00% |
| R168X-15 | CGUUGUGGGC | 55 | 0 | 0.00% |
| R168X-16 | GCUCUAUCGG | 8 | 0 | 0.00% |
| R168X-17 | GUUGGUGAUC | 3 | 0 | 0.00% |
| R168X-18 | GGGUUUAAGC | 1 | 0 | 0.00% |
| R168X-19 | CACUGUGAUU | 16 | 0 | 0.00% |
| R168X-20 | GCCCUAUCGG | 31 | 0 | 0.00% |
| R168X-21 | GCUGGGUAUU | 607 | 0 | 0.00% |
| R168X-22 | AGUUGUGGGC | 5 | 0 | 0.00% |
| R168X-23 | ACCCUAUCGG | 21 | 0 | 0.00% |
| R168X-24 | CGUUGGGGUC | 2 | 0 | 0.00% |
| R168X-25 | GCUGGGUGUC | 9 | 0 | 0.00% |
| R168X-26 | GGCUGGGUGC | 6 | 0 | 0.00% |
| R168X-27 | UUCUGUGGGC | 3 | 0 | 0.00% |
| R168X-28 | AGUUGGGGUC | 7 | 0 | 0.00% |
| R168X-29 | CCCUGUGGCU | 8 | 0 | 0.00% |
| R168X-30 | CCCUUAUCUA | 7 | 0 | 0.00% |

Table 2.3: R255X Top 30 Sequences

| Guide Name | Sequence 5' -> 3' | Edit Site: A Reads | Edit Site: G Reads | % Editing |
|---|---|---|---|---|
| R255X-1 | UUUUUGUGAC | 0 | 5 | 100.00% |
| R255X-2 | UUUUUAUGAC | 0 | 5 | 100.00% |
| R255X-3 | GAGUUGUGAC | 0 | 4 | 100.00% |
| R255X-4 | UUUUUGUCCA | 1 | 5 | 83.30% |
| R255X-5 | CUUUUAUGUC | 1 | 3 | 75.00% |
| R255X-6 | AUUUUGUCCG | 1 | 3 | 75.00% |
| R255X-7 | UUUUUGUCCG | 1 | 2 | 66.70% |
| R255X-8 | UCCUUGCGAC | 1 | 2 | 66.70% |
| R255X-9 | CUUUUAGGAA | 1 | 2 | 66.70% |
| R255X-10 | CCCUUCGGCC | 1 | 2 | 66.70% |
| R255X-11 | AUUUUGUGAC | 3 | 6 | 66.70% |
| R255X-12 | AUUUACGACC | 1 | 2 | 66.70% |
| R255X-13 | AUUGUCGAUU | 1 | 2 | 66.70% |
| R255X-14 | ACUUUGUGUC | 1 | 2 | 66.70% |
| R255X-15 | UUUUUGUCUU | 3 | 5 | 62.50% |
| R255X-16 | UAAUUGUCUU | 2 | 3 | 60.00% |
| R255X-17 | CCCUUGUGUC | 2 | 3 | 60.00% |
| R255X-18 | AUUUUGUCUC | 2 | 3 | 60.00% |
| R255X-19 | UUUUUGGCCG | 2 | 2 | 50.00% |
| R255X-20 | UUUUUGGCCC | 2 | 2 | 50.00% |
| R255X-21 | UCUUUGCGGC | 2 | 2 | 50.00% |
| R255X-22 | UCCUUGUGGC | 2 | 2 | 50.00% |
| R255X-23 | UCCUUGUGAC | 2 | 2 | 50.00% |
| R255X-24 | UCAAGAUGGC | 2 | 2 | 50.00% |
| R255X-25 | GUUUUGUGUC | 2 | 2 | 50.00% |
| R255X-26 | GAUUGUGACC | 2 | 2 | 50.00% |
| R255X-27 | AUUUUUGACC | 2 | 2 | 50.00% |
| R255X-28 | AUUUUGUCCA | 3 | 3 | 50.00% |
| R255X-29 | AUUUAUGACC | 2 | 2 | 50.00% |
| R255X-30 | AUGAAACUCG | 2 | 2 | 50.00% |

Table 2.4: R270X Top 30 Sequences

| Guide Name | Sequence 5' -> 3' | Edit Site: A Reads | Edit Site: G Reads | % Editing |
|---|---|---|---|---|
| R270X-1 | AGUUAUGCGA | 0 | 21 | 100.00% |
| R270X-2 | AGUUACGCGA | 2 | 22 | 91.70% |
| R270X-3 | UGCCUGGCUC | 16 | 160 | 90.90% |
| R270X-4 | UCCUUUAUCA | 20 | 178 | 89.90% |
| R270X-5 | UUCUUGUCGA | 3 | 26 | 89.70% |
| R270X-6 | AGUUUGGCUC | 17 | 147 | 89.60% |
| R270X-7 | UCCUUGGCGU | 41 | 319 | 88.60% |
| R270X-8 | CCCUUGGCAA | 14 | 106 | 88.30% |
| R270X-9 | CACUUGGCUC | 21 | 154 | 88.00% |
| R270X-10 | CCCUUAGCGA | 3 | 21 | 87.50% |
| R270X-11 | UGUUUGGCUC | 72 | 472 | 86.80% |
| R270X-12 | UCCUUUAUCG | 19 | 121 | 86.40% |
| R270X-13 | CCUUUAUCGA | 4 | 23 | 85.20% |
| R270X-14 | UCUUUAUUAA | 28 | 160 | 85.10% |
| R270X-15 | CGCCUGUCAA | 3 | 17 | 85.00% |
| R270X-16 | UGCCUGGCUG | 15 | 82 | 84.50% |
| R270X-17 | CCUUUAUCUA | 19 | 103 | 84.40% |
| R270X-18 | CACUUGUCGA | 5 | 27 | 84.40% |
| R270X-19 | CCCUUGUCGA | 6 | 32 | 84.20% |
| R270X-20 | UACUUGGCUC | 35 | 184 | 84.00% |
| R270X-21 | CCCUUGGCGU | 44 | 231 | 84.00% |
| R270X-22 | UUCUUGGCGA | 11 | 57 | 83.80% |
| R270X-23 | CGUUACGCGA | 8 | 41 | 83.70% |
| R270X-24 | GGUUUGGCUC | 16 | 81 | 83.50% |
| R270X-25 | ACUUUAUCUA | 18 | 91 | 83.50% |
| R270X-26 | AACUUGGCAA | 6 | 30 | 83.30% |
| R270X-27 | CGCCUUAUCG | 4 | 20 | 83.30% |
| R270X-28 | UCUUUAUCGA | 7 | 35 | 83.30% |
| R270X-29 | UCCUUGGCUC | 18 | 88 | 83.00% |
| R270X-30 | UCCUUGUCUA | 19 | 92 | 82.90% |

## 2.6: Tables of oligonucleotides

Table 2.5**:** DNA inserts to make reporter plasmids (Figure 2.2B and C)

| Name | Sequence 5’ -> 3’ |
|---|---|
| dsDNA insert for *MECP2* W104X Reporter with $N_3$ (coding strand) | CGACTCACTATAGGGAATATTAAGCTTATGCGCCCCGGCAGGAAGTAG AAAGCTGAGGCCGACTACAAGGGTCGGCCTCAGCTTTNNNCTTCCTG CCGGGGCGAAATCGTCTAAAGGTGAAGAATTATTCACTGG |
| dsDNA insert for *MECP2* W104X Reporter with $N_4$ (coding strand) | CGACTCACTATAGGGAATATTAAGCTTATGCGCCCCGGCAGGAAGTAG AAAGCTGAGGCCGACTACAAGGGTCGGCCTCAGCTTNNNNCTTCCT GCCGGGGCGAAATCGTCTAAAGGTGAAGAATTATTCACTGG |
| dsDNA insert for *MECP2* W104X Reporter with $N_6$ (coding strand) | CGACTCACTATAGGGAATATTAAGCTTATGCGCCCCGGCAGGAAGTAG AAAGCTGAGGCCGACTACAAGGGTCGGCCTCAGCTTTNNNNNNCTT CCTGCCGGGGCGAAATCGTCTAAAGGTGAAGAATTATTCACTGG |
| dsDNA insert for modified mIDUA W392X Reporter with $N_3$ (coding strand) | CGACTCACTATAGGGAATATTAAGCTTATGGGAGAACAACTCTGAGCA GAGGTCTCAAAGGCTGGGGACAAGCTCCGGCCTTTGAGATCTCTGNN NAGAGTTGTTTTCAGATCTAAAGGTGAAGAATTATTCACTGG |
| dsDNA insert for modified mIDUA W392X Reporter with $N_4$ (coding strand) | CGACTCACTATAGGGAATATTAAGCTTATGGGAGAACAACTCTGAGCA GAGGTCTCAAAGGCTGGGGACAAGCTCCGGCCTTTGAGATCTCTNN NNAGAGTTGTTTTCAGATCTAAAGGTGAAGAATTATTCACTGG |
| dsDNA insert for modified mIDUA W392X Reporter with $N_6$ (coding strand) | CGACTCACTATAGGGAATATTAAGCTTATGGGAGAACAACTCTGAGCA GAGGTCTCAAAGGCTGGGGACAAGCTCCGGCCTTTGAGATCTCTGNN NNNNAGAGTTGTTTTCAGATCTAAAGGTGAAGAATTATTCACTGG |

Table 2.6: DNA inserts to validate reporter plasmids (Figure 2.3)

| Name | Sequence 5’ -> 3’ |
|---|---|
| dsDNA insert for *MECP2* W104X Reporter for validation (coding strand) | CGACTCACTATAGGGAATATTAAGCTTATGCGCCCCGGCAGGAAG TAGAAAGCTGAGGCCGACTACAAGGGTCGGCCTCAGCTTTCCAC TTCCTGCCGGGGCGAAATCGTCTAAAGGTGAAGAATTATTCACT GG |
| dsDNA insert for *MECP2* W104X G Reporter for validation (coding strand) | CGACTCACTATAGGGAATATTAAGCTTATGCGCCCCGGCAGGAAG TGGAAAGCTGAGGCCGACTACAAGGGTCGGCCTCAGCTTTCCA CTTCCTGCCGGGGCGAAATCGTCTAAAGGTGAAGAATTATTCACT GG |
| dsDNA insert for modified mIDUA W392X Reporter for validation (coding strand) | CGACTCACTATAGGGAATATTAAGCTTATGGGAGAACAACTCTGA GCAGAGGTCTCAAAGGCTGGGGACAAGCTCCGGCCTTTGAGAT CTCTGCCCAGAGTTGTTTTCAGATCTAAAGGTGAAGAATTATTCA CTGG |
| dsDNA insert for modified mIDUA W392X G Reporter for validation (coding strand) | CGACTCACTATAGGGAATATTAAGCTTATGGGAGAACAACTCTGG GCAGAGGTCTCAAAGGCTGGGGACAAGCTCCGGCCTTTGAGAT CTCTGCCCAGAGTTGTTTTCAGATCTAAAGGTGAAGAATTATTCA CTGG |

Table 2.7: R168X DNA template for Hairpin RNA library, primers and RNA library (Figure 2.5)

| Name | Sequence 5’ -> 3’ |
|---|---|
| R168X hairpin DNA template | ATACCGTGAGTAATACGACTCACTATAGGGTAATTGTCAACAGACCGT CGAGCCCCTCCCGGTGAGAGCAGAAACCACCTGTAAAGGTGGTTTC NNNNNNNNNNGGGAGGGGCTAAGATGCGTCGGTGTACAAA |
| R168X hairpin Forward Primer | ATACCGTGAGTAATACGACTCACTATA |
| R168X hairpin Reverse | TTTGTACACCGACGCATCTTA |
| R168X hairpin Forward Primer for RT-PCR | GGGTAATTGTCAACAGACCG |
| R168X RNA Hairpin Library | GGGUAAUUGUCAACAGACCGUCGAGCCCCUCCCGGUGAGAGCAG AAACCACCUGUAAAGGUGGUUUCNNNNNNNNNNGGGAGGGGC UAAGAUGCGUCGGUGUACAAA |

Table 2.8: R255X DNA template for hairpin RNA library, primers, and RNA library (Figure 2.6)

| Name | Sequence 5’ -> 3’ |
|---|---|
| R255X DNA template for Hairpin RNA library | GTCCTCAACATAATACGACTCACTATAGGGAATGGCGCTTCCCGGCAGGAAGT GAAAAGCTGAGGCCGACTTTCTTTCTTTCGTCGGCCTCANNNNNNNNNNTCC TGCCGGGTTGTAGGGTC |
| R255X hairpin Forward Primer | GTCCTAACGATAATACGACT |
| R255X hairpin Reverse | GACCCTACAACCCGGC |
| R255X hairpin Forward Primer for RT-PCR | GGGAATGGCGCTTCC |
| R255X RNA Hairpin Library | GGGAAUGGCGCUUCCCGGCAGGAAGUGAAAAGCUGAGGCCGACUUUCUU UCUUUCGUCGGCCUCANNNNNNNNNNUCCUGCCGGGUUGUAGGGUC |

Table 2.9: R270X DNA templates for hairpin RNA library, primers, and RNA library (Figure 2.7)

| Name | Sequence 5’ -> 3’ |
|---|---|
| R270X DNA template for Hairpin RNA library | GTCCTCAACATAATACGACTCACTATAGGGAATGGCGCTTAAGAAACGGGGCT GAAAGCCGGGGAGTGTGTTTCTTTCTTTCCACACTCCCCNNNNNNNNNNCC CGTTTCTTTTGTAGGGTC |
| R270X hairpin Forward Primer | GTCCTCAACATAATACGA |
| R270X hairpin Reverse | GACCCTACAAAAGAAAC |
| R270X hairpin Forward Primer for RT-PCR | GGGAATGGCGCTTAA |
| R270X RNA Hairpin Library | GGGAAUGGCGCUUAAGAAACGGGGCUGAAAGCCGGGGAGUGUGUUUCU UUCUUUCCACACUCCCCNNNNNNNNNNCCCGUUUCUUUUGUAGGGUC |

Table 2.10: W104X DNA templates for hairpin RNA library, primers, and RNA library (Figure 2.8A)

| Name | Sequence 5’ -> 3’ |
|---|---|
| W104X DNA template for Hairpin RNA library | ATACCGTGAGTAATACGACTCACTATAGGGAATGGCGCTTCCTGCCTGAAG GCTAGACACGGAAGCTTAATTTCTTTCTTTCTTAAGCTTCCNNNNNNNNN NTTCAGGCAGGTTGTAGGGTC |
| W104X hairpin Forward Primer | ATACCGTGAGTAATACGACTCACTATA |
| W104X hairpin Reverse | GACCCTACAACCTGCCTGAA |
| W104X hairpin Forward Primer for RT-PCR | GGGAATGGCGCTTCCTG |
| W104X RNA Hairpin Library | GGGAAUGGCGCUUCCUGCCUGAAGGCUAGACACGGAAGCUUAAUUU CUUUCUUUCUUAAGCUUCCNNNNNNNNNNUUCAGGCAGGUUGUAG GGUC |

Table 2.11: R255X dsRBD EMERGe DNA templates for hairpin RNA library, primers, and RNA library (Figure 2.8B)

| Name | Sequence 5’ -> 3’ |
|---|---|
| R255X dsRBD EMERGe for Hairpin RNA library | ATACCGTGAGTAATACGACTCACTATAGGGAATGGCGCTTGGCAGGA AGTGAAAAGCTGAGGCCGACCCTCAGGCCATTCCCAAGAAACGGGG CCGAAAGCCGGGGAGTGTGGTGGCAAGGNNNNNCCTCAGCTTTCG ACTTCCTGCCTTGTAGGGTC |
| R255X dsRBD EMERGe hairpin Forward Primer | ATACCGTGAGTAATACGACTCACTATA |
| R255X dsRBD EMERGe hairpin Reverse | GACCCTACAAGGCAGGAAG |
| R255X dsRBD EMERGe hairpin Forward Primer for RT-PCR | GGGAATGGCGCTTGGC |
| R255X dsRBD EMERGe RNA Hairpin Library | GGGAAUGGCGCUUGGCAGGAAGUGAAAAGCUGAGGCCGACCCUC AGGCCAUUCCCAAGAAACGGGGCCGAAAGCCGGGGAGUGUGGUG GCAGGCCUGAGGGUCGGCCUCAGCUUUCGACUUCCUGCCUUGUA GGGUC |

## 2.7: References

(1) Jacobsen, C. S.; Salvador, P.; Yung, J. F.; Kragness, S.; Mendoza, H. G.; Mandel, G.; Beal, P. A. Library Screening Reveals Sequence Motifs That Enable ADAR2 Editing at Recalcitrant Sites. *ACS Chem Biol* **2023**. DOI: 10.1021/acschembio.3c00107 From NLM Publisher.

(2) Wang, Y.; Beal, P. A. Probing RNA recognition by human ADAR2 using a high-throughput mutagenesis method. *Nucleic Acids Research* **2016**, *44* (20), 9872-9880. DOI: 10.1093/nar/gkw799.
(3) Park, S.; Doherty, E. E.; Xie, Y.; Padyana, A. K.; Fang, F.; Zhang, Y.; Karki, A.; Lebrilla, C. B.; Siegel, J. B.; Beal, P. A. High-throughput mutagenesis reveals unique structural features of human ADAR1. *Nat Commun* **2020**, *11* (1), 5130. DOI: 10.1038/s41467-020-18862-2 From NLM Medline.
(4) Monteleone, L. R.; Matthews, M. M.; Palumbo, C. M.; Chiang, Y.; Fisher, A. J.; Beal, P. A.; Monteleone, L. R.; Matthews, M. M.; Palumbo, C. M.; Thomas, J. M.; et al. Article A Bump-Hole Approach for Directed RNA Editing Article A Bump-Hole Approach for Directed RNA Editing. *Cell Chemical Biology* **2019**, *26* (2), 269-277.e265. DOI: 10.1016/j.chembiol.2018.10.025.
(5) 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. DOI: 10.1002/humu.10194 From NLM Medline.
(6) Kuttan, A.; Bass, B. L. Mechanistic insights into editing-site specificity of ADARs. *Proceedings of the National Academy of Sciences of the United States of America* **2012**, *109* (48). DOI: 10.1073/pnas.1212548109.
(7) Sinnamon, J. R.; Jacobson, M. E.; Yung, J. F.; Fisk, J. R.; Jeng, S.; McWeeney, S. K.; Parmelee, L. K.; Chan, C. N.; Yee, S. P.; 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. DOI: 10.1073/pnas.2206053119 From NLM Medline.
(8) Bass, B. L. RNA editing by adenosine deaminases that act on RNA. *Annu Rev Biochem* **2002**, *71*, 817-846. DOI: 10.1146/annurev.biochem.71.110601.135501 From NLM Medline.
(9) Wang, Y.; Zheng, Y.; Beal, P. A. Adenosine Deaminases That Act on RNA (ADARs). *Enzymes* **2017**, *41*, 215-268. DOI: 10.1016/bs.enz.2017.03.006 From NLM Medline.
(10) Yi, Z.; Qu, L.; Tang, H.; Liu, Z.; Liu, Y.; Tian, F.; Wang, C.; Zhang, X.; Feng, Z.; Yu, Y.; et al. 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. DOI: 10.1038/s41587-021-01180-3 From NLM Medline.
(11) Xiang, Y.; Katrekar, D.; Mali, P. Methods for recruiting endogenous and exogenous ADAR enzymes for site-specific RNA editing. *Methods* **2022**, *205*, 158-166. DOI: 10.1016/j.ymeth.2022.06.011 From NLM Medline.
(12) Schneider, M. F.; Wettengel, J.; Hoffmann, P. C.; Stafforst, T. Optimal guideRNAs for re-directing deaminase activity of hADAR1 and hADAR2 in trans. *Nucleic Acids Res* **2014**, *42* (10), e87. DOI: 10.1093/nar/gku272 From NLM Medline.
(13) Doherty, E. E.; Wilcox, X. E.; van Sint Fiet, L.; Kemmel, C.; Turunen, J. J.; Klein, B.; Tantillo, D. J.; Fisher, A. J.; Beal, P. A. Rational Design of RNA Editing Guide Strands: Cytidine Analogs at the Orphan Position. *J Am Chem Soc* **2021**, *143* (18), 6865-6876. DOI: 10.1021/jacs.0c13319 From NLM Medline.
(14) Monian, P.; Shivalila, C.; Lu, G.; Shimizu, M.; Boulay, D.; Bussow, K.; Byrne, M.; Bezigian, A.; Chatterjee, A.; Chew, D.; et al. Endogenous ADAR-mediated RNA editing in non-human primates using stereopure chemically modified oligonucleotides. *Nat Biotechnol* **2022**, *40* (7), 1093-1102. DOI: 10.1038/s41587-022-01225-1 From NLM Medline.

(15) Polson, A. G.; Bass, B. L. Preferential selection of adenosines for modification by double-stranded RNA adenosine deaminase. *EMBO J* **1994**, *13* (23), 5701-5711. DOI: 10.1002/j.1460-2075.1994.tb06908.x From NLM Medline.
(16) Phelps, K. J.; Tran, K.; Eifler, T.; Erickson, A. I.; Fisher, A. J.; Beal, P. A. Recognition of duplex RNA by the deaminase domain of the RNA editing enzyme ADAR2. *Nucleic Acids Res* **2015**, *43* (2), 1123-1132. DOI: 10.1093/nar/gku1345 From NLM Medline.
(17) Diaz Quiroz, J. F.; Ojha, N.; Shayhidin, E. E.; De Silva, D.; Dabney, J.; Lancaster, A.; Coull, J.; Milstein, S.; Fraley, A. W.; Brown, C. R.; Rosenthal, J. J. C. Development of a selection assay for small guide RNAs that drive efficient site-directed RNA editing. *Nucleic Acids Res* **2023**, *51* (7), e41. DOI: 10.1093/nar/gkad098 From NLM Medline.
(18) Liu, X.; Sun, T.; Shcherbina, A.; Li, Q.; Jarmoskaite, I.; Kappel, K.; Ramaswami, G.; Das, R.; Kundaje, A.; Li, J. B. Learning cis-regulatory principles of ADAR-based RNA editing from CRISPR-mediated mutagenesis. *Nat Commun* **2021**, *12* (1), 2165. DOI: 10.1038/s41467-021-22489-2 From NLM Medline.
(19) Lehmann, K. A.; Bass, B. L. Double-stranded RNA adenosine deaminases ADAR1 and ADAR2 have overlapping specificities. *Biochemistry* **2000**, *39* (42), 12875-12884. DOI: 10.1021/bi001383g From NLM Medline.
(20) 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*. DOI: 10.7554/eLife.75555 From NLM Medline.
(21) Eifler, T.; Chan, D.; Beal, P. A. A screening protocol for identification of functional mutants of RNA editing adenosine deaminases. *Curr Protoc Chem Biol* **2012**, *4* (4), 357-369. DOI: 10.1002/9780470559277.ch120139 From NLM PubMed-not-MEDLINE.

# Chapter 3

## *In Vitro* Verification and Application of EMERGe-derived gRNAs

*The biochemical studies with MECP2 R255X were part of a joint project and were completed by Prince Salvador. This chapter contains excepts from the full manuscript which was published in ACS Chemical Biology in April 2023.* [1]

### 3.1: Introduction

ADAR substrates fall into two main types. The first type is an intramolecular dsRNA substrate, also known as a *cis* substrate. Most of the discovered endogenous ADAR targets fall into this category. [9, 15, 19, 22] The second type is an intermolecular dsRNA substrate, known as a *trans* substrate. An ADAR-based therapeutic, it would likely emerge from this form of intermolecular dsRNA by using a gRNA with an endogenous target. [10, 14, 17, 23-26] This established the need to discover gRNAs that promote ADAR editing. Up until this point, both screens (Sat-FACS-Seq and EMERGe) have been using *cis* substrates. This was the key factor enabling the use of NGS. Since the library and target were linked, it could be sequenced. The main goal of this screen was to discover small gRNAs that will enable editing. Therefore, the linked *cis* library substrate will need to be verified to be used in a *trans* system. To complete this verification, the resulting guides from the NGS screen will need to be adapted to small gRNAs. This will result in an optimal guide from the EMERGe screen.

EMERGe verification proceeded with 24 representatives guides from the completed screens from Chapter 2. These gRNAs were tested with their respective targets. The gRNAs that had activity and contained unique motifs were candidates for SAR to discover the precise nucleotides for each editing-enabling motif. While conducting the verification and SAR studies of the RNAs, I uncovered the potential for artifacts from the EMERGe workflow. These artifacts do

not directly promote editing, but they are features that were selected for through the EMERGe screen. The guides that contained EMERGe artifacts were used as templates to develop EMERGe-derived gRNAs. This allowed for the discovery of ADAR-enabling motifs that would never be revealed through rational design alone.

In addition, the motifs discovered may interact differently with ADAR than the current literature dictates. Since the motifs were localized around the editing site, there may be unique interactions with the E488 residue. To probe these interactions, E488X mutants were tested with the top gRNAs. All 20 common amino acids were tested with these substrates to discover if there are indeed unique interactions. From those, a sterically bulky mutant seemed to be active with a EMERGe derived gRNA. This showed the potential of this screen to select sequences that function with ADARs which are typically considered inactive. [4, 6, 20]

## 3.2: Results

### 3.2.1: Verification and SAR of EMERGe Selected gRNAs.

With the amount of NGS-derived SAR that has been completed it was only logical to move forward to *in vitro* verification. I was particularly interested to know if any of the top sequences across the three datasets in the *MECP2* context (R168X, R255X, and R270X) would support an *in vitro* ADAR2 reaction with a *trans* dsRNA. Between the three data sets, we found 24 representative guide sequences that were promising. These guides were used as representatives of the datasets at large.

The *MECP2* R168X dataset was used for the initial *trans* verification. This was because the R168X context contained a 5'-GAG target. While containing the worst nearest-neighbor of a 5'G

it does contain the optimal 3' nearest neighbor, G.[6] 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 3.1A). 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[6] (Figure 3.1B). Six representative sequences were compared to the control guide. The sequence that showed the most effective editing (70% ±1%) in these studies was R168X-5 (Figure 3.1F). Importantly, this guide contains two features known to enable ADAR editing at 5'-GA sites: an A:C pair at the target adenosine and a G:G pair involving the 5' G[12, 27] (Figure 3.2A). 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 N10 region. These features allowed for the discovery of the G:G preferential mismatch within an unbiased screen. At this time, it was also seen that a G:G mismatch in other sequence contexts was beneficial in a completely independent set of experiments.[27]

To evaluate which features are necessary for editing, SAR studies were completed by designing five new antisense guide strands with key differences from R168X-5 (Figure 3.2B-F). First, the G:G pair at the -1 position was replaced by a G:C pair. This guide, R168X-5a (Figure 3.2B) 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). 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 3.2C). 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,

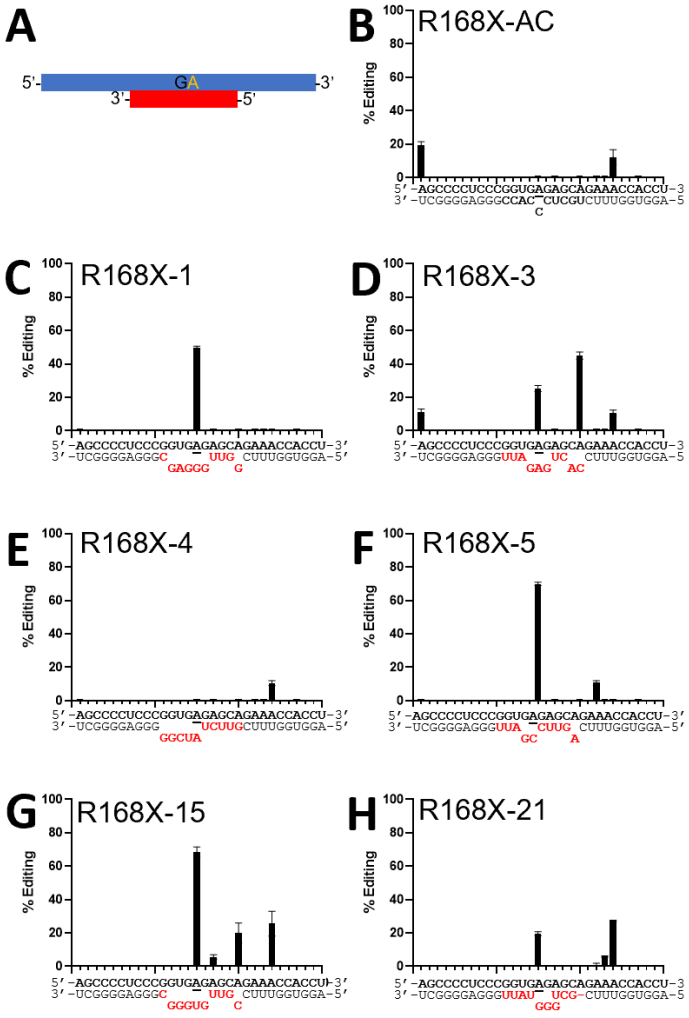


Figure 3.1: Validation of the *MECP2* R168X Screen

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.

guides R168X-5c-e were designed (Figure 3.2D-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 N10 can be beneficial. When we compare this information to that collected within the NGS-derived SAR we see that this interaction is not limited to an A:G mismatch, but an A:G mismatch also works (Figure 2.10). This indicates the presence of a possible purine:purine interaction.

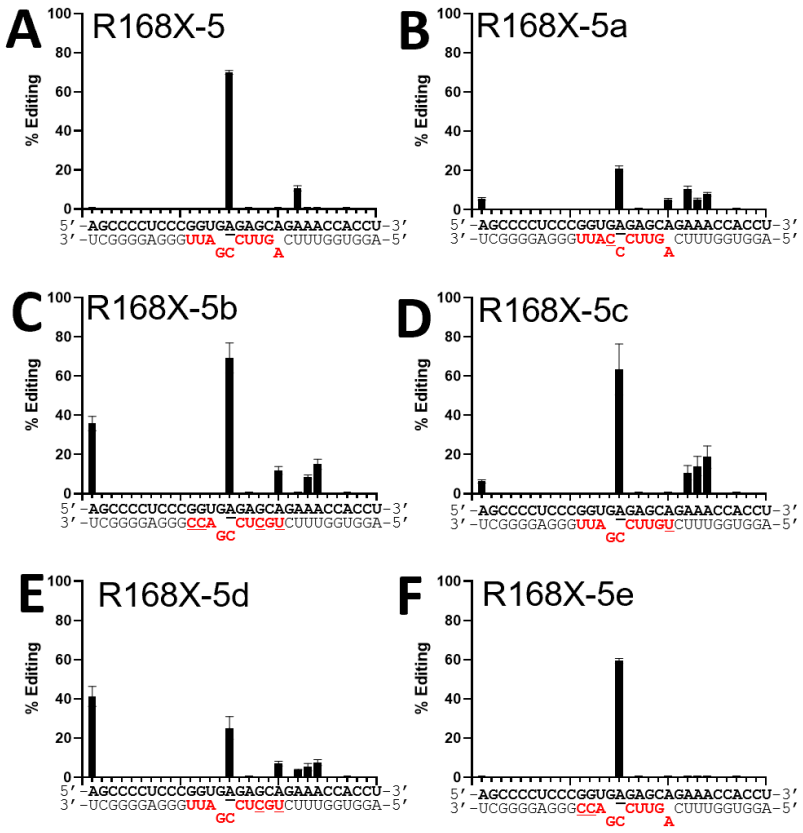


Figure 3.2: The effect of mutations in the R168X-5 guide.

Sequence changes made to the R168X-5 sequence are underlined in red. 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. 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.

Once the data was verified for the suboptimal R168X 5'-GAG context, it was time to move to the hardest ADAR context, 5'-GAA. *MECP2* R255X was the next candidate for verification and these experiments were completed by Prince Salvador. Among the top 30 sequences, a subset was tested independently for on-target and bystander editing in trans (Figure 3.3). In this subset, for example, R255X-11 led to the most specific and efficient on-target editing (Figure 3.3F). R255X-11 (Figure 3.3F) was then compared directly to a guide containing the 5'-GA enabling G:G pair and A:C pair, R255X-GG;AC (Figure 3.3L). R255X-11 supported target site editing to a similar level (86 ± 12% vs 86 ± 6%), and it reduced editing at two bystander sites from 20 ± 4% and 40 ± 9% to undetectable under the conditions of the assay. This reduction in bystander editing is caused by the -4 bystander A getting bulged out of the RNA duplex, and then no longer being a good ADAR edit site. [10] These results again verify EMERGe's ability to identify N10 sequences that enable specific editing. Notable features of the predicted secondary structure for the R255X-11 guide-target RNA duplex include a single bulged A present in the target strand four nucleotides 5' of the target site and a three nucleotide GUG loop opposite the target A and the G on its 5' side. It is hypothesized that the U within the GUG loop is acting as an orphan base for ADAR. This would mean that the third G of the GUG loop is bulged out of the duplex, while the first G is forming a G:G mismatch, as found in R168X-5 as well. While not shown here, Prince Salvador has shown that the key motif of R255X-11 is the 3'-GUG.

As a further expansion of the difficult to edit 5'-GAA context, the R270X data were verified. Representative sequences from this group were used for follow up verification studies as

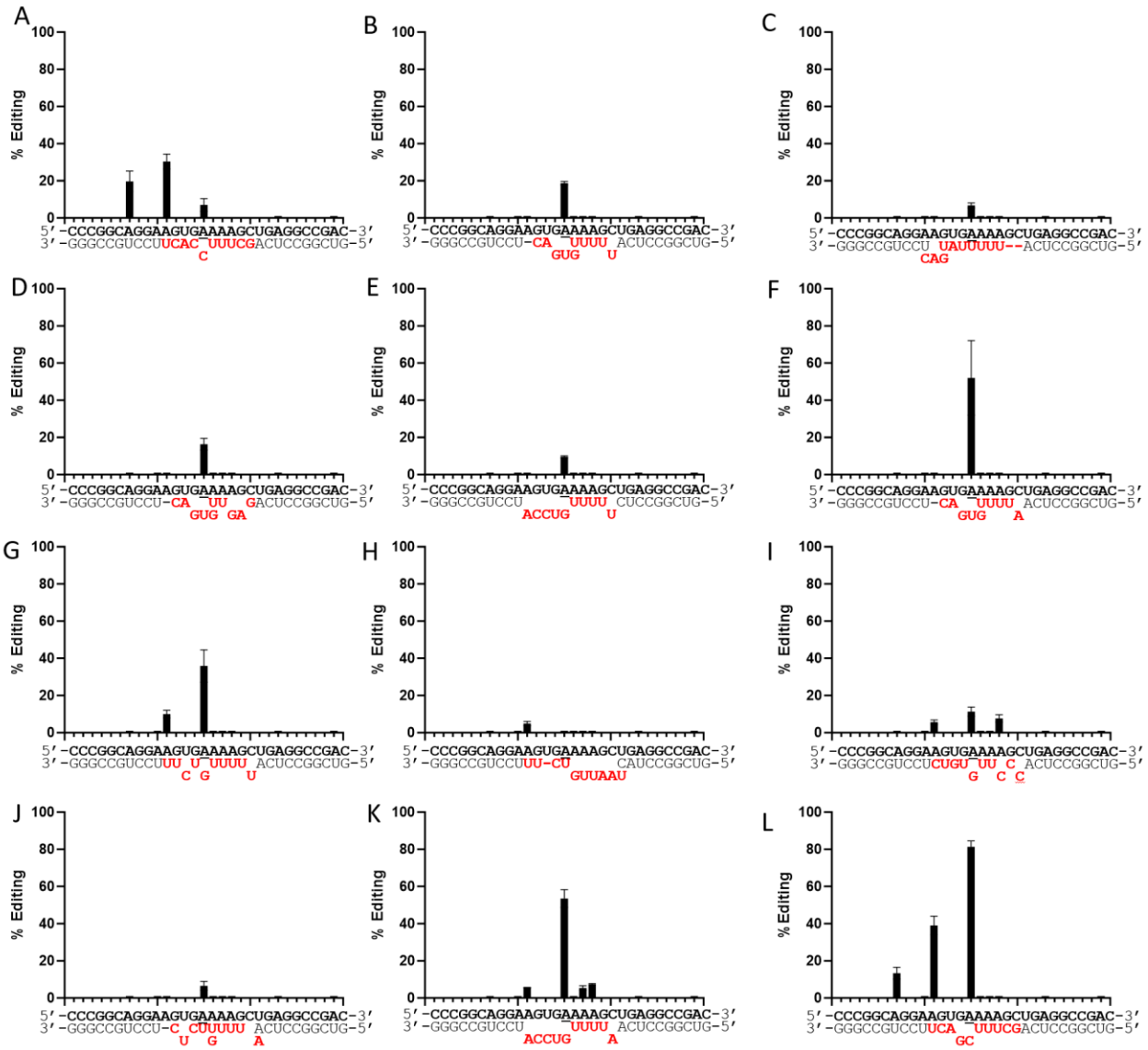


Figure 3.3: Validation of the *MECP2* R255X Screen

On-target and bystander editing observed with hADAR2 and select guide sequences from the *MECP2* R255X screen. Percent editing at target site and bystander sites observed with 200 nM hADAR2 in a 30 min reaction A: Control sequence with an AC mismatch at the target site. B-K: R255X-1, 2, 3, 4, 11, 15, 16, 17, 18. L: R255X-GG;AC. Data are plotted as the mean ± s.d. from three independent experiments.

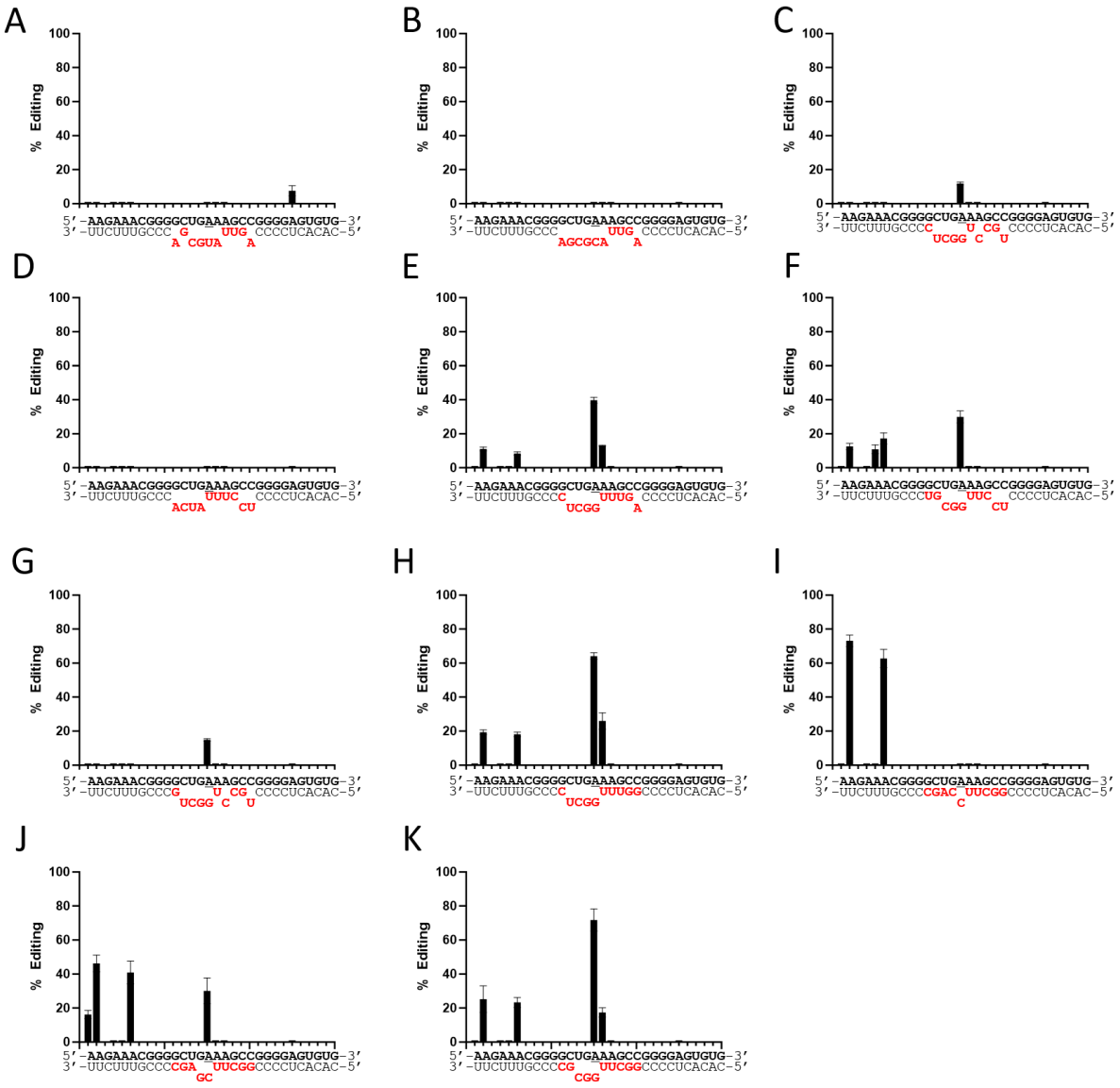


Figure 3.4: Validation of the *MECP2* R270X Screen

On-target and bystander editing observed with hADAR2 and select guide sequences from the *MECP2* R270X screen. Percent editing at target site and bystander sites observed with 200 nM hADAR2 in a 30 min reaction. A-H: R270X-1, 2, 3, 4, 6, 7, 16, 24 I: R270X-AC: control sequence with an AC mismatch at the target site. J: R270X-GG, AC, guide containing an AC mismatch at target and adjacent GG. K. R270X-24a Data are plotted as the mean ± s.d. from three independent experiments.

described above (Figure 3.4). 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 3.4H). Interestingly, analyzing the R270X-24 result along with two closely related hit sequences R270X-6 (Figure 3.4E) and R270X-7 (Figure 3.4D) indicated that a key element enabling editing is the 3'-CGG across from the 5'-UGA editing site. This would agree with the motif discovered in Chapter 2 (Figure 2.11). Therefore, the antisense oligonucleotide guide R270X-24a was synthesized to contain only that 3 nt motif while keeping the remainder of the duplex Watson-Crick base paired (Figure 3.4K). We saw a significant increase in on-target editing, (from 30 ± 7% to 71 ± 7%) and a reduction in bystander editing at two sites (position 2: from 46 ± 5% to 25 ± 8% and position 6: from 41 ± 6% to 23 ± 3%) for the R270X-24a guide, in comparison to the R270X-GG;AC control (Figure 3.4J and Figure 3.5).

To better understand how the 3'-CGG motif enables editing (Figure 3.6A), we analyzed the effect of single nucleotide mutations at each position in two different ways. First, I generated an activity heat map for variants of the R270X-24 hairpin present in the original screening data set (Figure 3.6B). Second, I 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 I generated the corresponding activity heat map (Figure 3.6B). 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 for 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.

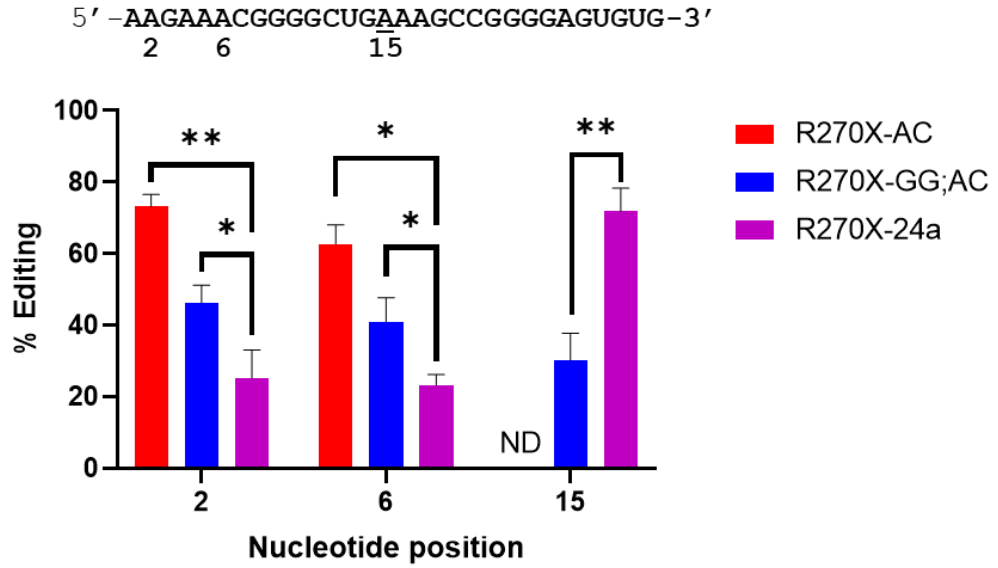


Figure 3.5: On-target and bystander editing between R270X gRNAs

On-target and 2 bystander editing sites observed with hADAR2 and shown guide sequences from the *MECP2* R270X screen. On-target editing site is position 15. Percent editing at target site and bystander sites observed with 200 nM hADAR2 in a 30 min reaction. 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 ** $p < 0.01$).

High resolution structures are available for ADAR2-RNA complexes that reveal details of the interactions between the protein and nucleotides around the editing site.[28, 29] Indeed, our lab recently published a structure of ADAR2 bound to substrate RNA where the editing site is adjacent to a G:G pair.[7] 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 amino acid at position 488 of hADAR2 (Figure 3.6C). There are two hydrogen bonds of interest for this position: the first is an interaction with the backbone; the

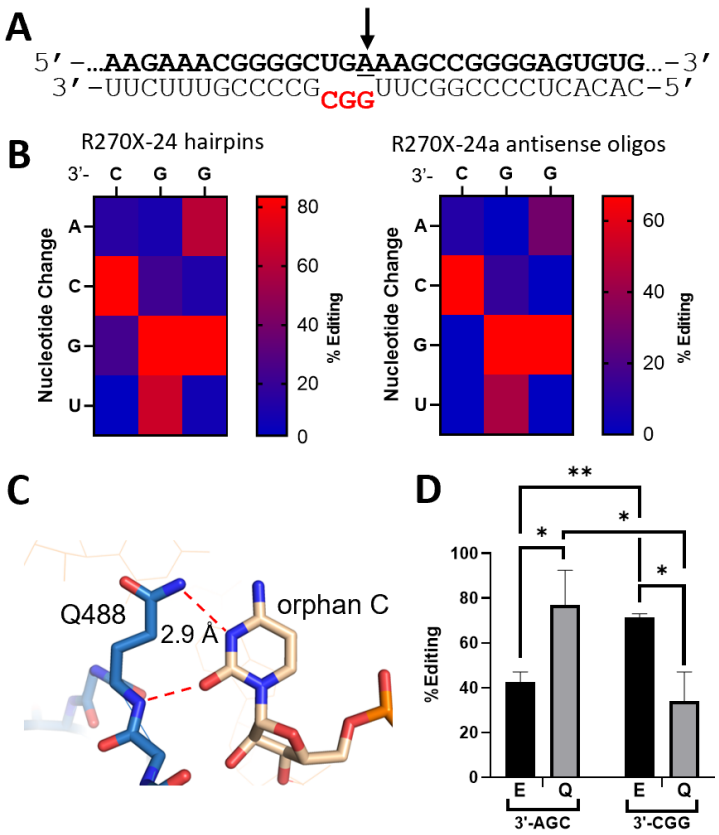


Figure 3.6: EMERGe screen identifies a novel motif enabling editing at R270X premature termination codon.

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 ** $p < 0.01$).

second being an interaction between the side chain and the N1 of cytosine. The latter is prominent for querying the orphan position against wild type and hyperactive mutants of ADAR2.[13] At this point, it is not clear which nucleotide of the 3’-CGG motif of R270X-24a functions as the orphan base. When the orphan base is cytidine, as is the case for the control guide R270X-GG;AC, the ADAR2 mutant E488Q is hyperactive (Figure 3.6D; 3’-AGC). However, with R270X-24a this

mutant is actually less active than the wild type protein (E488) where the 3'-CGG motif is opposite the editing site (Figure 3.6D, 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. [13, 14] It is also possible that mutations other than E488Q may be uniquely beneficial with the 3'-CGG motif targeting the 5'-UGA site. [4, 30]

**3.2.2: ADAR-Orphan base interactions of EMERGe selected gRNAs.**

To determine if editing efficiency can be improved further with a different amino acid at the 488 position, an assay was completed with all other position 488 mutants and an ADAR guide bearing the 3'-CGG motif. The E488X notation denotes a mutation at position 488, which is the substitution of the amino acid glutamic acid (E) with any amino acid (X). Mutagenesis enables changing this position to each of the common 20 amino acids. Each amino acid could exhibit unique interactions with the orphan position, and some mutations may render the protein effectively inactive. [6, 20] Twenty mutants were tested with four distinct substrates. These substrates were based on four unique guides, targeting two different *MECP2* RNA sequences: R270X and W104X. The four specific substrates included R270X-24a, R270X-GG;AC, R270X-GUG, and W104X-AC. The R270X-GUG substrate is a combination of the R255X-11 motif, with adaptation to the R270X target. The sequence contains the 3'-GUG motif, with all other nucleotides complementary. This is similar to how R270X-24a was developed. The W104X target served as a positive control due to the availability of detailed crystal structures, [28, 29] and prior mutagenesis studies in this ideal UAG context. [6, 20] To mitigate variations in the small-scale

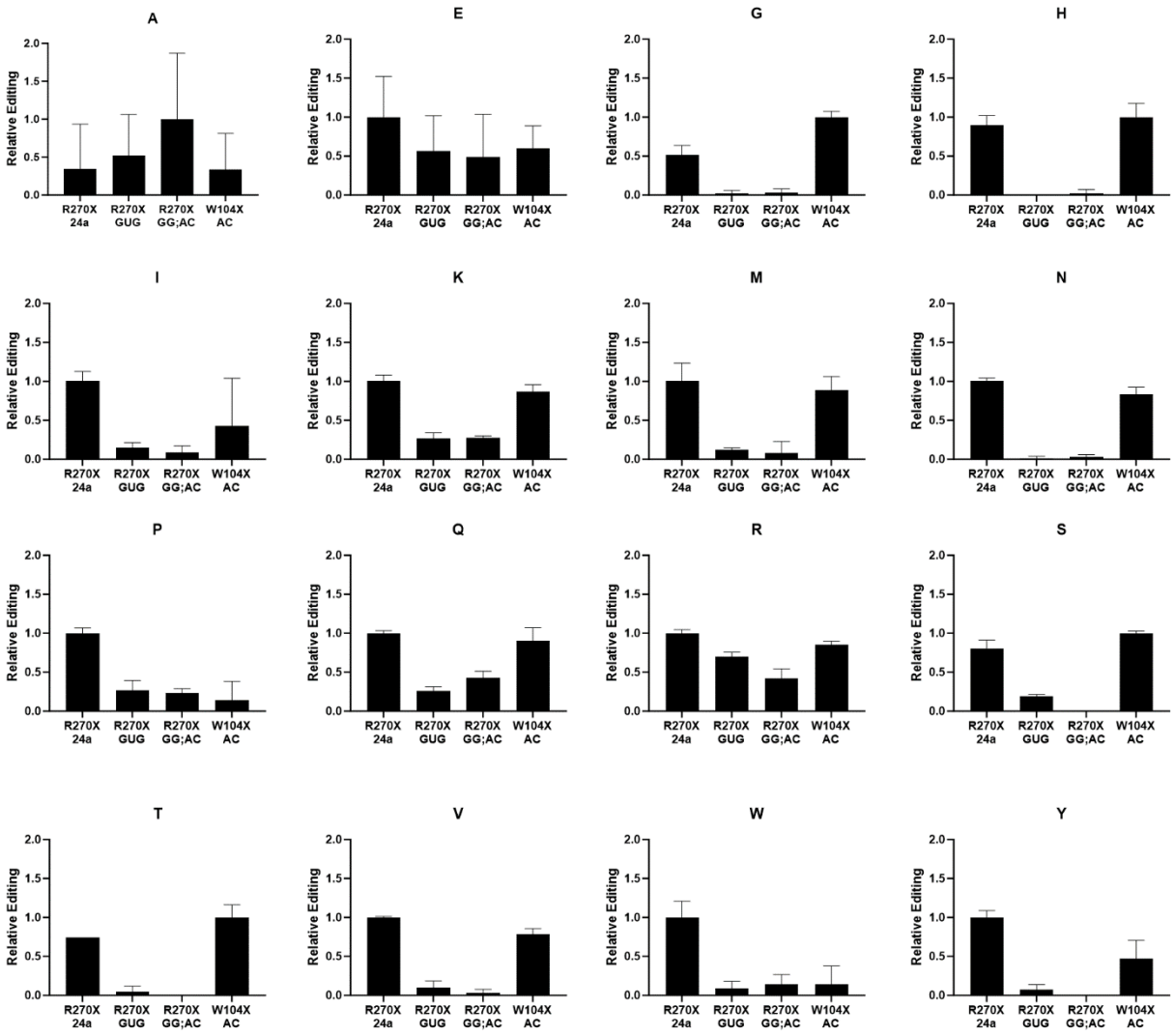


Figure 3.7: R270X-24a, R270X-GUG, R270X-GG;AC and W104X-AC substrates tested with mutant ADAR's.

Figure uses single letter amino acid codes for naming. The mutants used are ADAR2 E488X where X corresponds to the mutant in their respective graph. Concentrated yeast expressed ADAR2 E488X was used with 10 nM target RNA in a 30 minute reaction. Mutants not shown had no detectable editing. The top editing condition for each ADAR2 E488X mutant was set to 1 for their respective graphs. The relative editing for R270X-24a is greater than W104X-AC with the ADAR2 E488W (**) and E488P (*) proteins (Welch's t-test, * $p < 0.05$ and ** $p < 0.01$).

purification used for the 19 mutant proteins and the resulting protein concentrations, relative editing was employed instead of % editing (Figure 3.7).

Through this experiment, it was observed that most mutants were as active with R270X-24a as they were with the ideal context W104X-AC substrate. However, certain mutants were active for R270X-24a, but were inactive in the ideal context of W104X-AC. Bulky ADAR mutants, such as E488W, exhibited significantly higher editing with the 3'-CGG motif, in comparison to the ideal context (Figure 3.7W). Studies have shown that the E488W mutant is detrimental for ADAR editing in a standard context with an orphan C. [4, 20] Therefore, the 3'-CGG motif may possess a secondary structure that allows for the insertion of this bulky residue. This would mean that the 3'-CGG motif has a non-standard ADAR-Orphan interaction. Additionally, it is seen that E488P is more active with R270X-24a than it is with W104X-AC (Figure 3.7P). While not as active as R270X-24a, R270X-GUG was seen to be most active with the E488R mutant (Figure 3.7R). More studies will need to be completed to understand the type of orphan interaction present in R270X-24a and R270X-GUG. Nevertheless, these preliminary studies clearly show that the motifs selected via the EMERGe screening protocol can have unique preferences for the identity of the residue at position 488 of ADAR2. This further supports the hypothesis that novel editing systems can be discovered that are orthogonal to wild type ADARs reacting with duplex RNA containing target adenosines in an A:U or A:C pairs

### 3.3: Discussion

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[7]. 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 Figure 3.1, Figure 3.3, and Figure 3.4). 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. [31]

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.

The use of an E488X mutant assay has begun to shed light on the editing mechanism with the selected guides. The structure adopted by the 3'-CGG motif is currently unknown, but this specific sequence appears to be tolerant of bulky residues. Additionally, the ability to tolerate a bulky mutant can reduce off-target effects due to the bulky mutant's steric characteristics, which typically clash with the orphan base.[4] This capability to tolerate bulky residues may enable this motif to facilitate specific editing at difficult-to-edit sites, while also reducing bystander editing. Here to, further studies are required to gain a better understanding of this interaction.

In addition to R270X-24a being generally active, it was intriguing to observe that R270X-GUG and R270X-GG;AC exhibited similar activities to each other. This may suggest that these two motifs share a structural similarity that is not shared with R270X-24a. We currently hypothesize that the 3'-GUG motif behaves similarly to the previously published G:G pair.[27] Likewise, this E488X result may indicate such a similarity. The next step for these gRNAs would be to subject them to crystallization assays to determine the precise interactions present.

## 3.4: Methods

**3.4.1: Protein overexpression and purification.** hADAR2 WT, hADAR2-RD WT and hADAR2-RD E488Q with N-terminal $His_{10}$ tag were overexpressed in *S. cerevisiae* and purified as described in Chapter 2.

**3.4.2: 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.

**3.4.3: Site-directed mutagenesis for ADAR2 E488X proteins.** Mutagenesis of the human ADAR2 proteins was carried out via Golden Gate Ligation procedures. PCR amplification (NEB M0530L) was completed using primers containing the desired mutations. The DNA was then purified by gel extraction using 0.5% agarose (Qiagen 28706). The purified DNA was ligated using a cocktail of PaqC1 (NEB R0745S), DPN1 (NEB R0176S), and T4 DNA Ligase (NEB M0202) within T4 DNA Ligase Buffer (NEB M0202). The resulting ligated plasmid was transformed into XL10-Gold Ultracompetent Cells (Agilent 200314) and purified (Qiagen 27106). Sequences for mutant plasmids were confirmed by Sanger sequencing. All primers were purchased from IDT.

**3.4.4: Small Scale Protein overexpression and purification of E488X mutants.** hADAR2 E488 mutants with N-terminal His10 tag were overexpressed in S. cerevisiae and purified as follows. The cell growth (20 ml) was pelleted and lysed using glass beads in lysis buffer (1 ml) 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 □-mercaptoethanol (BME). The clarified lysate was then passed over Ni-NTA spin columns (Qiagen 31314), 500 uL each of the following wash buffers were used: (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 with 800 ul of 200 mM imidazole. Protein fractions were pooled, concentrated, then buffer exchanged with a storage buffer containing 20 mM Tris-HCl pH 8.0, 20% (v/v) glycerol, 100 mM NaCl and 1 mM BME. Concentrated protein was stored at -72 °C.

**3.4.5: Detection of Mutant ADAR2 E488X proteins.** Quenched deamination reaction samples were pooled for each specific mutant of ADAR2. These samples were concentrated in vacuo to a total volume of 50 µL. Each sample (10 ul) was fractionated in a sodium dodecyl sulfate-PAGE gel (ThermoFisher NP0322BOX) alongside Page Ruler Prestained Plus Protein Ladder (ThermoFisher 26619). Western blotting was completed using a 6x-His Tag monoclonal antibody (ThermoFisher MA1-21315) as a primary antibody at 1:1000 dilution and anti-mouse IgG with alkaline phosphatase-conjugated secondary antibody (Sigma-Aldrich AP124A) at a 1:1000 dilution. Mutant ADAR proteins were detected using an ECF substrate (Fisher 45000947) on a BioRad GelDock using the Sypro Red protocol.

## 3.5: Table of oligonucleotides

Table 3.1: R168X DNA template for RNA target, primers and RNA target (Figure 3.1 and Figure 3.2)

| Name | Sequence 5’ -> 3’ |
|---|---|
| R168X target dsDNA template. Coding strand shown. | ATACCGTGAGTAATACGACTCACTATAGGGAAAAGCTTTTCGCTCTAAAGTAG AATTGATTGCATACTTTGAAAAGGTGGGAGACACCTCCTTGGACCCTAATGAT TTTGACTTCACGGTAACTGGGAGAGGGAGCCCCTCCAGGTGAGAGCAGAAA CCACCTAAGAAGCCCAAATCTCCCAAAGCTCCAGGAACTGGCAGGGGTCGG GGACGCCCCAAAGGGA |
| R168X Forward Primer | ATACCGTGAGTAATACGACTC |
| R168X Reverse | TCCCTTTGGGGCGTCC |
| R168X Forward Primer for RT-PCR | GGGAAAAGCTTTTCGCTC |
| R168X RNA Target | GGGAAAAGCUUUUCGCUCUAAAGUAGAAUUGAUUGCAUACUUUGAAAA GGUGGGAGACACCUCCUUGGACCCUAAUGAUUUUGACUUCACGGUAAC UGGGAGAGGGAGCCCCUCCAGGUGAGAGCAGAAACCACCUAAGAAGCCC AAAUCUCCCAAAGCUCCAGGAACUGGCAGGGGUCGGGGACGCCCCAAAG GGA |

Table 3.2: DNA templates for representative R168X gRNAs (Figure 3.1)

| Name | Sequence 5’ -> 3’ |
|---|---|
| R168X-AC top for hybridization | CACGATTAATACGACTCACTATAGGGAGGTGGTTTCTGCT CCCACCGGGAGGGGCT |
| R168X-AC bottom for hybridization | AGCCCCTCCCGGTGGGAGCAGAAACCACCTCCCTATAGTG AGTCGTATTAATCGTG |
| R168X-1 top for hybridization | CACGATTAATACGACTCACTATAGGGAGGTGGTTTCGGTT GGGAGCGGGAGGGGCT |
| R168X-1 bottom for hybridization | AGCCCCTCCCGCTCCCAACCGAAACCACCTCCCTATAGTGA GTCGTATTAATCGTG |
| R168X-3 top for hybridization | CACGATTAATACGACTCACTATAGGGAGGTGGTTTCCACT GAGATTGGGAGGGGCT |

| | |
|---|---|
| R168X-3 bottom for hybridization | AGCCCCTCCCAATCTCAGTGGAAACCACCTCCCTATAGTGAGTCGTATTAATCGTG |
| R168X-4 top for hybridization | CACGATTAATACGACTCACTATAGGGAGGTGGTTTCGTTCTATCGGGGGAGGGGCT |
| R168X-4 bottom for hybridization | AGCCCCTCCCCCGATAGAACGAAACCACCTCCCTATAGTGAGTCGTATTAATCGTG |
| R168X-5 top for hybridization | CACGATTAATACGACTCACTATAGGGAGGTGGTTTCAGTTCCGATTGGGAGGGGCT |
| R168X-5 bottom for hybridization | AGCCCCTCCCAATCGGAACTGAAACCACCTCCCTATAGTGAGTCGTATTAATCGTG |
| R168X-15 top for hybridization | CACGATTAATACGACTCACTATAGGGAGGTGGTTTCCGTTGTGGGCGGGAGGGGCT |
| R168X-15 bottom for hybridization | AGCCCCTCCCGCCCACAACGGAAACCACCTCCCTATAGTGAGTCGTATTAATCGTG |
| R168X-21 top for hybridization | CACGATTAATACGACTCACTATAGGGAGGTGGTTTCGCTGGGTATTGGGAGGGGCT |
| R168X-21 bottom for hybridization | AGCCCCTCCCAATACCCAGCGAAACCACCTCCCTATAGTGAGTCGTATTAATCGTG |

Table 3.3: R168X representative gRNAs (Figure 3.1)

| Name | Sequence 5’ -> 3’ |
|---|---|
| R168X-AC | GGGAGGUGGUUUCUGCUCCCACCGGGAGGGGCU |
| R168X-1 | GGGAGGUGGUUUCGGUUGGGAGCGGGAGGGGCU |
| R168X-3 | GGGAGGUGGUUUCCACUGAGAUUGGGAGGGGCU |
| R168X-4 | GGGAGGUGGUUUCGUUCUAUCGGGGGAGGGGCU |
| R168X-5 | GGGAGGUGGUUUCAGUUCCGAUUGGGAGGGGCU |
| R168X-15 | GGGAGGUGGUUUCCGUUGUGGGCGGGAGGGGCU |
| R168X-21 | GGGAGGUGGUUUCGCUGGGUAUUGGGAGGGGCU |

Table 3.4: DNA templates for SAR gRNAs (Figure 3.2)

| Name | Sequence 5’ -> 3’ |
|---|---|
| R168X-5 top for hybridization | CACGATTAATACGACTCACTATAGGGAGGTGGT TTCAGTTCCGATTGGGAGGGGCT |
| R168X-5 bottom for hybridization | AGCCCCTCCCAATCGGAACTGAAACCACCTCCC TATAGTGAGTCGTATTAATCGTG |
| R168X-5a top for hybridization | CACGATTAATACGACTCACTATAGGGAGGTGGT TTCAGTTCCCATTGGGAGGGGCT |
| R168X-5a bottom for hybridization | AGCCCCTCCCAATGGGAACTGAAACCACCTCCC TATAGTGAGTCGTATTAATCGTG |
| R168X-5b top for hybridization | CACGATTAATACGACTCACTATAGGGAGGTGGT TTCTGCTCCGACCGGGAGGGGCT |
| R168X-5b bottom for hybridization | AGCCCCTCCCGGTCGGAGCAGAAACCACCTCCC TATAGTGAGTCGTATTAATCGTG |
| R168X-5c top for hybridization | CACGATTAATACGACTCACTATAGGGAGGTGGT TTCTGTTCCGATTGGGAGGGGCT |
| R168X-5c bottom for hybridization | AGCCCCTCCCAATCGGAACAGAAACCACCTCCC TATAGTGAGTCGTATTAATCGTG |
| R168X-5d top for hybridization | CACGATTAATACGACTCACTATAGGGAGGTGGT TTCTGCTCCGATTGGGAGGGGCT |
| R168X-5d bottom for hybridization | AGCCCCTCCCAATCGGAGCAGAAACCACCTCCC TATAGTGAGTCGTATTAATCGTG |
| R168X-5e top for hybridization | CACGATTAATACGACTCACTATAGGGAGGTGGT TTCAGTTCCGACCGGGAGGGGCT |
| R168X-5e bottom for hybridization | AGCCCCTCCCGGTCGGAACTGAAACCACCTCCC TATAGTGAGTCGTATTAATCGTG |

Table 3.5: SAR gRNAs (Figure 3.2)

| Name | Sequence 5’ -> 3’ |
|---|---|
| R168X-5 | GGGAGGUGGUUUCAGUUCCGAUUGGGAGGGGCU |
| R168X-5a | GGGAGGUGGUUUCAGUUCCCAUUGGGAGGGGCU |
| R168X-5b | GGGAGGUGGUUUCUGCUCCGACCGGGAGGGGCU |
| R168X-5c | GGGAGGUGGUUUCUGUUCCGAUUGGGAGGGGCU |
| R168X-5d | GGGAGGUGGUUUCUGCUCCGAUUGGGAGGGGCU |
| R168X-5e | GGGAGGUGGUUUCAGUUCCGACCGGGAGGGGCU |

Table 3.6: R255X DNA templates for target RNA, primers, and RNA target (Figure 3.3)

| Name | Sequence 5’ -> 3’ |
|---|---|
| R255X target dsDNA template. Coding strand shown. | CACGATTAATACGACTCACTATAGGGGGTGTGCAGGTGAAAAGGGTCCTGGAGA AAAGTCCTGGGAAGCTCCTTGTCAAGATGCCTTTTCAAACTTCGCCAGGGGGCAA GGCTGAGGGGGGTGGGGCCACCACATCCACCCAGGTCATGGTGATCAAACGCCC CGGCAGGAAGTGAAAAGCTGAGGCCGACCCTCAGGCCATTCCCAAGAAACGGG GCCGAAAGCCGGGGAGTGTGGTGGCAGCCGCTGCCGCCGAGGCCAAAAAGAAA GCCGTGAAGGAGTCTTCTATCCGATCTGTGCAGGAGACCGTACTCCCCATCAAGA AG |
| R255X Forward Primer | CACGATTAATACGACTCACT |
| R255X Reverse | CACCACTTCCTTGACCTC |
| R255X Forward Primer for RT-PCR | GGGCTTGTCAAGATGC |
| R255X RNA Target | GGGGGUGUGCAGGUGAAAAGGGUCCUGGAGAAAAGUCCUGGGAAGCUCCU UGUCAAGAUGCCUUUUCAAACUUCGCCAGGGGGCAAGGCUGAGGGGGGUG GGGCCACCACAUCCACCCAGGUCAUGGUGAUCAAACGCCCCGGCAGGAAGUG AAAAGCUGAGGCCGACCCUCAGGCCAUUCCCAAGAAACGGGGCCGAAAGCCG GGGAGUGUGGUGGCAGCCGCUGCCGCCGAGGCCAAAAAGAAAGCCGUGAAG GAGUCUUCUAUCCGAUCUGUGCAGGAGACCGUACUCCCCAUCAAGAAG |

Table 3.7: DNA templates for representative R255X gRNAs (Figure 3.3)

| Name | Sequence 5’ -> 3’ |
|---|---|
| R255X-AC top for hybridization | CACGATTAATACGACTCACTATAGGGGTCGGCC TCAGCTTTCCACTTCCTGCCGGG |
| R255X-AC bottom for hybridization | CCCGGCAGGAAGTGGAAAGCTGAGGCCGACCC CTATAGTGAGTCGTATTAATCGTG |
| R255X-1 top for hybridization | CACGATTAATACGACTCACTATAGGGGTCGGCC TCATTTTTGTGACTCCTGCCGGG |
| R255X-1 bottom for hybridization | CCCGGCAGGAGTCACAAAAATGAGGCCGACCC CTATAGTGAGTCGTATTAATCGTG |
| R255X-2 top for hybridization | CACGATTAATACGACTCACTATAGGGGTCGGCC TCATTTTTATGACTCCTGCCGGG |
| R255X-2 bottom for hybridization | CCCGGCAGGAGTCATAAAAATGAGGCCGACCC CTATAGTGAGTCGTATTAATCGTG |
| R255X-3 top for hybridization | CACGATTAATACGACTCACTATAGGGGTCGGCC TCAGAGTTGTGACTCCTGCCGGG |
| R255X-3 bottom for hybridization | CCCGGCAGGAGTCACAACTCTGAGGCCGACCC CTATAGTGAGTCGTATTAATCGTG |
| R255X-4 top for hybridization | CACGATTAATACGACTCACTATAGGGGTCGGCC TCATTTTTGTCCATCCTGCCGGG |
| R255X-4 bottom for hybridization | CCCGGCAGGATGGACAAAAATGAGGCCGACCC CTATAGTGAGTCGTATTAATCGTG |
| R255X-11 top for hybridization | CACGATTAATACGACTCACTATAGGGGTCGGCC TCAATTTTGTGACTCCTGCCGGG |
| R255X-11 bottom for hybridization | CCCGGCAGGAGTCACAAAATTGAGGCCGACCC CTATAGTGAGTCGTATTAATCGTG |
| R255X-15 top for hybridization | CACGATTAATACGACTCACTATAGGGGTCGGCC TCATTTTTGTCTTTCCTGCCGGG |
| R255X-15 bottom for hybridization | CCCGGCAGGAAAGACAAAAATGAGGCCGACCC CTATAGTGAGTCGTATTAATCGTG |
| R255X-16 top for hybridization | CACGATTAATACGACTCACTATAGGGGTCGGCC TCATAATTGTCTTTCCTGCCGGG |

| | |
|---|---|
| R255X-16 bottom for hybridization | CCCGGCAGGAAAGACAATTATGAGGCCGACCC CTATAGTGAGTCGTATTAATCGTG |
| R255X-17 top for hybridization | CACGATTAATACGACTCACTATAGGGGTCGGCC TCACCCTTGTGTCTCCTGCCGGG |
| R255X-17 bottom for hybridization | CCCGGCAGGAGACACAAGGGTGAGGCCGACCC CTATAGTGAGTCGTATTAATCGTG |
| R255X-18 top for hybridization | CACGATTAATACGACTCACTATAGGGGTCGGCC TCAATTTTGTCTCTCCTGCCGGG |
| R255X-18 bottom for hybridization | CCCGGCAGGAGAGACAAAATTGAGGCCGACCC CTATAGTGAGTCGTATTAATCGTG |
| R255X-28 top for hybridization | CACGATTAATACGACTCACTATAGGGGTCGGCC TCAATTTTGTCCATCCTGCCGGG |
| R255X-28 bottom for hybridization | CCCGGCAGGATGGACAAAATTGAGGCCGACCC CTATAGTGAGTCGTATTAATCGTG |
| R255X-GG;AC top for hybridization | CACGATTAATACGACTCACTATAGGGGTCGGCC TCAGCTTTCGACTTCCTGCCGGG |
| R255X-GG;AC bottom for hybridization | CCCGGCAGGAAGTCGAAAGCTGAGGCCGACCC CTATAGTGAGTCGTATTAATCGTG |

Table 3.8: Representative R255X gRNAs (Figure 3.3)

| Name | Sequence 5’ -> 3’ |
|---|---|
| R255X-AC | GGGGUCGGCCUCAGCUUUCCACUUCCUGCCGGG |
| R255X-1 | GGGGUCGGCCUCAUUUUUGUGACUCCUGCCGGG |
| R255X-2 | GGGGUCGGCCUCAUUUUUAUGACUCCUGCCGGG |
| R255X-3 | GGGGUCGGCCUCAGAGUUGUGACUCCUGCCGGG |
| R255X-4 | GGGGUCGGCCUCAUUUUUGUCCAUCCUGCCGGG |
| R255X-11 | GGGGUCGGCCUCAAUUUUGUGACUCCUGCCGGG |
| R255X-15 | GGGGUCGGCCUCAUUUUUGUCUUUCCUGCCGGG |
| R255X-16 | GGGGUCGGCCUCAUAAUUGUCUUUCCUGCCGGG |
| R255X-17 | GGGGUCGGCCUCACCCUUGUGUCUCCUGCCGGG |
| R255X-18 | GGGGUCGGCCUCAAUUUUGUCUCUCCUGCCGGG |
| R255X-28 | GGGGUCGGCCUCAAUUUUGUCCAUCCUGCCGGG |
| R255X-GG;AC | GGGGUCGGCCUCAGCUUUCGACUUCCUGCCGGG |

Table 3.9: R270X DNA templates for RNA target, primers, and RNA (Figure 3.4)

| Name | Sequence 5’ -> 3’ |
|---|---|
| R270X dsDNA template for RNA target. Coding strand shown. | CACGATTAATACGACTCACTATAGGGCCTTGTCAAGATGCCTTTTCAAACTTCGC CAGGGGGCAAGGCTGAGGGGGGTGGGGCCACCACATCCACCCAGGTCATGGT GATCAAACGCCCCGGCAGGAAGCGAAAAGCTGAGGCCGACCCTCAGGCCATT CCCAAGAAACGGGGCTGAAAGCCGGGGAGTGTGGTGGCAGCCGCTGCCGCC GAGGCCAAAAAGAAAGCCGTGAAGGAGTCTTCTATCCGATCTGTGCAGGAGAC CGTACTCCCCATCAAGAAGCGCAAGACCCGGGAGACGGTCAGCATCGAGGTCA AGGAAGTGGT |
| R270X Forward Primer | CACGATTAATACGACTCACTATAGG |
| R270X Reverse | ACCACTTCCTTGACCTCG |
| R270X Forward Primer for RT-PCR | GGGCCTTGTCAAGATGC |
| R270X RNA Target | GGGCCUUGUCAAGAUGCCUUUUCAAACUUCGCCAGGGGGCAAGGCUGAG GGGGGUGGGGCCACCACAUCCACCCAGGUCAUGGUGAUCAAACGCCCCGG CAGGAAGCGAAAAGCUGAGGCCGACCCUCAGGCCAUUCCCAAGAAACGGGG CUGAAAGCCGGGGAGUGUGGUGGCAGCCGCUGCCGCCGAGGCCAAAAAGA AAGCCGUGAAGGAGUCUUCUAUCCGAUCUGUGCAGGAGACCGUACUCCCC AUCAAGAAGCGCAAGACCCGGGAGACGGUCAGCAUCGAGGUCAAGGAAGU GGU |

Table 3.10: DNA templates for representative gRNAs (Figure 3.4)

| Name | Sequence 5’ -> 3’ |
|---|---|
| R270X-1 top for hybridization | CACGATTAATACGACTCACTATAGGGCACACTCCCCAGTTATGCGACCCGTTTCTT |
| R270X-1 bottom for hybridization | AAGAAACGGGTCGCATAACTGGGGAGTGTGCCCTATAGTGAGTCGTATTAATCGTG |
| R270X-2 top for hybridization | CACGATTAATACGACTCACTATAGGGCACACTCCCCAGTTACGCGACCCGTTTCTT |
| R270X-2 bottom for hybridization | AAGAAACGGGTCGCGTAACTGGGGAGTGTGCCCTATAGTGAGTCGTATTAATCGTG |
| R270X-3 top for hybridization | CACGATTAATACGACTCACTATAGGGCACACTCCCCTGCCTGGCTCCCCGTTTCTT |
| R270X-3 bottom for hybridization | AAGAAACGGGGAGCCAGGCAGGGGAGTGTGCCCTATAGTGAGTCGTATTAATCGTG |
| R270X-4 top for hybridization | CACGATTAATACGACTCACTATAGGGCACACTCCCCTCCTTTATCACCCGTTTCTT |
| R270X-4 bottom for hybridization | AAGAAACGGGTGATAAAGGAGGGGAGTGTGCCCTATAGTGAGTCGTATTAATCGTG |
| R270X-6 top for hybridization | CACGATTAATACGACTCACTATAGGGCACACTCCCCAGTTTGGCTCCCCGTTTCTT |
| R270X-6 bottom for hybridization | AAGAAACGGGGAGCCAAACTGGGGAGTGTGCCCTATAGTGAGTCGTATTAATCGTG |
| R270X-7 top for hybridization | CACGATTAATACGACTCACTATAGGGCACACTCCCCTCCTTGGCGTCCCGTTTCTT |
| R270X-7 bottom for hybridization | AAGAAACGGGACGCCAAGGAGGGGAGTGTGCCCTATAGTGAGTCGTATTAATCGTG |
| R270X-16 top for hybridization | CACGATTAATACGACTCACTATAGGGCACACTCCCCTGCCTGGCTGCCCGTTTCTT |
| R270X-16 bottom for hybridization | AAGAAACGGGCAGCCAGGCAGGGGAGTGTGCCCTATAGTGAGTCGTATTAATCGTG |
| R270X-24 top for hybridization | CACGATTAATACGACTCACTATAGGGCACACTCCCCGGTTTGGCTCCCCGTTTCTT |

| | |
|---|---|
| R270X-24 bottom for hybridization | AAGAAACGGGGAGCCAAACCGGGGAGTGTGCCCTATAGTGAGTCGTATTAATCGTG |
| R270X-AC top for hybridization | CACGATTAATACGACTCACTATAGGGCACACTCCCCGGCTTCCAGCCCCGTTTCTT |
| R270X-AC bottom for hybridization | AAGAAACGGGGCTGGAAGCCGGGGAGTGTGCCCTATAGTGAGTCGTATTAATCGTG |
| R270X-GG;AC top for hybridization | CACGATTAATACGACTCACTATAGGGCACACTCCCCGGCTTCGAGCCCCGTTTCTT |
| R270X-GG;AC bottom for hybridization | AAGAAACGGGGCTCGAAGCCGGGGAGTGTGCCCTATAGTGAGTCGTATTAATCGTG |
| R270X-CGG top for hybridization | CACGATTAATACGACTCACTATAGGGCACACTCCCCGGCTTGGCGCCCCGTTTCTT |
| R270X-CGG bottom for hybridization | AAGAAACGGGGCGCCAAGCCGGGGAGTGTGCCCTATAGTGAGTCGTATTAATCGTG |

Table 3.11: R270X representative gRNAs (Figure 3.4)

| Name | Sequence 5’ -> 3’ |
|---|---|
| R270X-1 | GGGCACACUCCCCAGUUAUGCGACCCGUUUCUU |
| R270X-2 | GGGCACACUCCCCAGUUACGCGACCCGUUUCUU |
| R270X-3 | GGGCACACUCCCCUGCCUGGCUCCCCGUUUCUU |
| R270X-4 | GGGCACACUCCCCUCCUUUAUCACCCGUUUCUU |
| R270X-6 | GGGCACACUCCCCAGUUUGGCUCCCCGUUUCUU |
| R270X-7 | GGGCACACUCCCCUCCUUGGCGUCCCGUUUCUU |
| R270X-16 | GGGCACACUCCCCUGCCUGGCUGCCCGUUUCUU |
| R270X-24 | GGGCACACUCCCCGGUUUGGCUCCCCGUUUCUU |
| R270X-AC | GGGCACACUCCCCGGCUUCCAGCCCCGUUUCUU |
| R270X-GG;AC | GGGCACACUCCCCGGCUUCGAGCCCCGUUUCUU |
| R270X-CGG | GGGCACACUCCCCGGCUUGGCGCCCCGUUUCUU |

Table 3.12: DNA templates for R270X-CGG single nucleotide variant gRNAs (Figure 3.6B: Antisense Oligos)

| Name | Sequence 5’ -> 3’ |
|---|---|
| R270X 3’-AGG top for hybridization | CACGATTAATACGACTCACTATAGGGCACACTCC CCGGCTTGGAGCCCCGTTTCT |
| R270X 3’-AGG bottom for hybridization | AAGAAACGGGGCTCCAAGCCGGGGAGTGTGCC CTATAGTGAGTCGTATTAATCGTG |
| R270X 3’-TGG top for hybridization | CACGATTAATACGACTCACTATAGGGCACACTCC CCGGCTTGGTGCCCCGTTTCTT |
| R270X 3’-TGG bottom for hybridization | AAGAAACGGGGCACCAAGCCGGGGAGTGTGCC CTATAGTGAGTCGTATTAATCGTG |
| R270X 3’-GGG top for hybridization | CACGATTAATACGACTCACTATAGGGCACACTCC CCGGCTTGGTGCCCCGTTTCTT |
| R270X 3’-GGG bottom for hybridization | AAGAAACGGGGCACCAAGCCGGGGAGTGTGCC CTATAGTGAGTCGTATTAATCGTG |
| R270X 3’-CAG top for hybridization | CACGATTAATACGACTCACTATAGGGCACACTCC CCGGCTTGACGCCCCGTTTCTT |
| R270X 3’-CAG bottom for hybridization | AAGAAACGGGGCGTCAAGCCGGGGAGTGTGCC CTATAGTGAGTCGTATTAATCGTG |
| R270X 3’-CTG top for hybridization | CACGATTAATACGACTCACTATAGGGCACACTCC CCGGCTTGTCGCCCCGTTTCTT |
| R270X 3’-CTG bottom for hybridization | AAGAAACGGGGCGACAAGCCGGGGAGTGTGCC CTATAGTGAGTCGTATTAATCGTG |
| R270X 3’-CCG top for hybridization | CACGATTAATACGACTCACTATAGGGCACACTCC CCGGCTTGCCGCCCCGTTTCTT |
| R270X 3’-CCG bottom for hybridization | AAGAAACGGGGCGGCAAGCCGGGGAGTGTGCC CTATAGTGAGTCGTATTAATCGTG |
| R270X 3’-CGA top for hybridization | CACGATTAATACGACTCACTATAGGGCACACTCC CCGGCTTAGCGCCCCGTTTCTT |
| R270X 3’-CGA bottom for hybridization | AAGAAACGGGGCGCTAAGCCGGGGAGTGTGCC CTATAGTGAGTCGTATTAATCGTG |

| | |
|---|---|
| R270X 3’-CGT top for hybridization | CACGATTAATACGACTCACTATAGGGCACACTCC CCGGCTTTGCGCCCCGTTTCTT |
| R270X 3’-CGT bottom for hybridization | AAGAAACGGGGCGCAAAGCCGGGGAGTGTGCC CTATAGTGAGTCGTATTAATCGTG |
| R270X 3’-CGC top for hybridization | CACGATTAATACGACTCACTATAGGGCACACTCC CCGGCTTCGCGCCCCGTTTCTT |
| R270X 3’-CGC bottom for hybridization | AAGAAACGGGGCGCGAAGCCGGGGAGTGTGCC CTATAGTGAGTCGTATTAATCGTG |
| R270X 3’-CGG top for hybridization | CACGATTAATACGACTCACTATAGGGCACACTCC CCGGCTTGGCGCCCCGTTTCTT |
| R270X 3’-CGG bottom for hybridization | AAGAAACGGGGCGCCAAGCCGGGGAGTGTGCC CTATAGTGAGTCGTATTAATCGTG |

Table 3.13: R270X-CGG single nucleotide variant gRNAs (Figure 3.6B: Antisense Oligos)

| Name | Sequence 5’ -> 3’ |
|---|---|
| R270X 3’-AGG | GGGCACACUCCCCGGCUUGGAGCCCCGUUUCUU |
| R270X 3’-UGG | GGGCACACUCCCCGGCUUGGUGCCCCGUUUCUU |
| R270X 3’-GGG | GGGCACACUCCCCGGCUUGGUGCCCCGUUUCUU |
| R270X 3’-CAG | GGGCACACUCCCCGGCUUGACGCCCCGUUUCUU |
| R270X 3’-CUG | GGGCACACUCCCCGGCUUGUCGCCCCGUUUCUU |
| R270X 3’-CCG | GGGCACACUCCCCGGCUUGCCGCCCCGUUUCUU |
| R270X 3’-CGA | GGGCACACUCCCCGGCUUAGCGCCCCGUUUCUU |
| R270X 3’-CGU | GGGCACACUCCCCGGCUUUGCGCCCCGUUUCUU |
| R270X 3’-CGC | GGGCACACUCCCCGGCUUCGCGCCCCGUUUCUU |
| R270X 3’-CGG | GGGCACACUCCCCGGCUUGGCGCCCCGUUUCUU |

Table 3.14: Mutagenesis Primers for E488X mutants (Figure 3.7).

| Name | Sequence 5’ -> 3’ |
|---|---|
| Forward Mutagenesis primer for ADAR2 E488R | GAGACACCTGCGAGAAAAATAGAGTCTGGTCGGGGGACGATT |
| Forward Mutagenesis primer for ADAR2 E488H | GAGACACCTGCGAGAAAAATAGAGTCTGGTCATGGGACGATT |
| Forward Mutagenesis primer for ADAR2 E488K | GAGACACCTGCGAGAAAAATAGAGTCTGGTAAAGGGACGATT |
| Forward Mutagenesis primer for ADAR2 E488D | GAGACACCTGCGAGAAAAATAGAGTCTGGTGATGGGACGATT |
| Forward Mutagenesis primer for ADAR2 E488S | GAGACACCTGCGAGAAAAATAGAGTCTGGTTCCGGGACGATT |
| Forward Mutagenesis primer for ADAR2 E488T | GAGACACCTGCGAGAAAAATAGAGTCTGGTACAGGGACGATT |
| Forward Mutagenesis primer for ADAR2 E488N | GAGACACCTGCGAGAAAAATAGAGTCTGGTAATGGGACGATT |
| Forward Mutagenesis primer for ADAR2 E488C | GAGACACCTGCGAGAAAAATAGAGTCTGGTTGTGGGACGATT |
| Forward Mutagenesis primer for ADAR2 E488G | GAGACACCTGCGAGAAAAATAGAGTCTGGTGGAGGGACGATT |
| Forward Mutagenesis primer for ADAR2 E488P | GAGACACCTGCGAGAAAAATAGAGTCTGGTCCCGGGACGATT |
| Forward Mutagenesis primer for ADAR2 E488A | GAGACACCTGCGAGAAAAATAGAGTCTGGTGCTGGGACGATT |
| Forward Mutagenesis primer for ADAR2 E488V | GAGACACCTGCGAGAAAAATAGAGTCTGGTGTCGGGACGATT |
| Forward Mutagenesis primer for ADAR2 E488I | GAGACACCTGCGAGAAAAATAGAGTCTGGTATTGGGACGATT |
| Forward Mutagenesis primer for ADAR2 E488L | GAGACACCTGCGAGAAAAATAGAGTCTGGTCTTGGGACGATT |
| Forward Mutagenesis primer for ADAR2 E488M | GAGACACCTGCGAGAAAAATAGAGTCTGGTATGGGGACGATT |

| | |
|---|---|
| Forward Mutagenesis primer for ADAR2 E488F | GAGACACCTGCGAGAAAAATAGAGTCTGGTTTTGGGACGATT |
| Forward Mutagenesis primer for ADAR2 E488Y | GAGACACCTGCGAGAAAAATAGAGTCTGGTTATGGGACGATT |
| Forward Mutagenesis primer for ADAR2 E488W | GAGACACCTGCGAGAAAAATAGAGTCTGGTTGGGGGACGATT |
| Forward Mutagenesis primer for ADAR2 E488Q | GAGACACCTGCGAGAAAAATAGAGTCTGGTCAGGGGACGATT |
| Reverse Mutagenesis primer for ADAR2 E488X | GAGACACCTGCGAGATTTTGGTCCGTAGCTGTCCTCTTG |

## 3.6: References


(1) Jacobsen, C. S.; Salvador, P.; Yung, J. F.; Kragness, S.; Mendoza, H. G.; Mandel, G.; Beal, P. A. Library Screening Reveals Sequence Motifs That Enable ADAR2 Editing at Recalcitrant Sites. *ACS Chem Biol* **2023**. DOI: 10.1021/acschembio.3c00107 From NLM Publisher.

(2) Wang, Y.; Beal, P. A. Probing RNA recognition by human ADAR2 using a high-throughput mutagenesis method. *Nucleic Acids Research* **2016**, *44* (20), 9872-9880. DOI: 10.1093/nar/gkw799.

(3) Park, S.; Doherty, E. E.; Xie, Y.; Padyana, A. K.; Fang, F.; Zhang, Y.; Karki, A.; Lebrilla, C. B.; Siegel, J. B.; Beal, P. A. High-throughput mutagenesis reveals unique structural features of human ADAR1. *Nat Commun* **2020**, *11* (1), 5130. DOI: 10.1038/s41467-020-18862-2 From NLM Medline.

(4) Monteleone, L. R.; Matthews, M. M.; Palumbo, C. M.; Chiang, Y.; Fisher, A. J.; Beal, P. A.; Monteleone, L. R.; Matthews, M. M.; Palumbo, C. M.; Thomas, J. M.; et al. Article A Bump-Hole Approach for Directed RNA Editing Article A Bump-Hole Approach for Directed RNA Editing. *Cell Chemical Biology* **2019**, *26* (2), 269-277.e265. DOI: 10.1016/j.chembiol.2018.10.025.

(5) 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. DOI: 10.1002/humu.10194 From NLM Medline.

(6) Kuttan, A.; Bass, B. L. Mechanistic insights into editing-site specificity of ADARs. *Proceedings of the National Academy of Sciences of the United States of America* **2012**, *109* (48). DOI: 10.1073/pnas.1212548109.

(7) Sinnamon, J. R.; Jacobson, M. E.; Yung, J. F.; Fisk, J. R.; Jeng, S.; McWeeney, S. K.; Parmelee, L. K.; Chan, C. N.; Yee, S. P.; 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. DOI: 10.1073/pnas.2206053119 From NLM Medline.

(8) Bass, B. L. RNA editing by adenosine deaminases that act on RNA. *Annu Rev Biochem* **2002**, *71*, 817-846. DOI: 10.1146/annurev.biochem.71.110601.135501 From NLM Medline.

(9) Wang, Y.; Zheng, Y.; Beal, P. A. Adenosine Deaminases That Act on RNA (ADARs). *Enzymes* **2017**, *41*, 215-268. DOI: 10.1016/bs.enz.2017.03.006 From NLM Medline.

(10) Yi, Z.; Qu, L.; Tang, H.; Liu, Z.; Liu, Y.; Tian, F.; Wang, C.; Zhang, X.; Feng, Z.; Yu, Y.; et al. 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. DOI: 10.1038/s41587-021-01180-3 From NLM Medline.
(11) Xiang, Y.; Katrekar, D.; Mali, P. Methods for recruiting endogenous and exogenous ADAR enzymes for site-specific RNA editing. *Methods* **2022**, *205*, 158-166. DOI: 10.1016/j.ymeth.2022.06.011 From NLM Medline.
(12) Schneider, M. F.; Wettengel, J.; Hoffmann, P. C.; Stafforst, T. Optimal guideRNAs for re-directing deaminase activity of hADAR1 and hADAR2 in trans. *Nucleic Acids Res* **2014**, *42* (10), e87. DOI: 10.1093/nar/gku272 From NLM Medline.
(13) Doherty, E. E.; Wilcox, X. E.; van Sint Fiet, L.; Kemmel, C.; Turunen, J. J.; Klein, B.; Tantillo, D. J.; Fisher, A. J.; Beal, P. A. Rational Design of RNA Editing Guide Strands: Cytidine Analogs at the Orphan Position. *J Am Chem Soc* **2021**, *143* (18), 6865-6876. DOI: 10.1021/jacs.0c13319 From NLM Medline.
(14) Monian, P.; Shivalila, C.; Lu, G.; Shimizu, M.; Boulay, D.; Bussow, K.; Byrne, M.; Bezigian, A.; Chatterjee, A.; Chew, D.; et al. Endogenous ADAR-mediated RNA editing in non-human primates using stereopure chemically modified oligonucleotides. *Nat Biotechnol* **2022**, *40* (7), 1093-1102. DOI: 10.1038/s41587-022-01225-1 From NLM Medline.
(15) Polson, A. G.; Bass, B. L. Preferential selection of adenosines for modification by double-stranded RNA adenosine deaminase. *EMBO J* **1994**, *13* (23), 5701-5711. DOI: 10.1002/j.1460-2075.1994.tb06908.x From NLM Medline.
(16) Phelps, K. J.; Tran, K.; Eifler, T.; Erickson, A. I.; Fisher, A. J.; Beal, P. A. Recognition of duplex RNA by the deaminase domain of the RNA editing enzyme ADAR2. *Nucleic Acids Res* **2015**, *43* (2), 1123-1132. DOI: 10.1093/nar/gku1345 From NLM Medline.
(17) Diaz Quiroz, J. F.; Ojha, N.; Shayhidin, E. E.; De Silva, D.; Dabney, J.; Lancaster, A.; Coull, J.; Milstein, S.; Fraley, A. W.; Brown, C. R.; Rosenthal, J. J. C. Development of a selection assay for small guide RNAs that drive efficient site-directed RNA editing. *Nucleic Acids Res* **2023**, *51* (7), e41. DOI: 10.1093/nar/gkad098 From NLM Medline.
(18) Liu, X.; Sun, T.; Shcherbina, A.; Li, Q.; Jarmoskaite, I.; Kappel, K.; Ramaswami, G.; Das, R.; Kundaje, A.; Li, J. B. Learning cis-regulatory principles of ADAR-based RNA editing from CRISPR-mediated mutagenesis. *Nat Commun* **2021**, *12* (1), 2165. DOI: 10.1038/s41467-021-22489-2 From NLM Medline.
(19) Lehmann, K. A.; Bass, B. L. Double-stranded RNA adenosine deaminases ADAR1 and ADAR2 have overlapping specificities. *Biochemistry* **2000**, *39* (42), 12875-12884. DOI: 10.1021/bi001383g From NLM Medline.
(20) 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*. DOI: 10.7554/eLife.75555 From NLM Medline.
(21) Eifler, T.; Chan, D.; Beal, P. A. A screening protocol for identification of functional mutants of RNA editing adenosine deaminases. *Curr Protoc Chem Biol* **2012**, *4* (4), 357-369. DOI: 10.1002/9780470559277.ch120139 From NLM PubMed-not-MEDLINE.
(22) 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. DOI: 10.1038/s41594-020-0398-4 From NLM Medline.

(23) Lyst, M. J.; Ekiert, R.; Ebert, D. H.; Merusi, C.; Nowak, J.; Selfridge, J.; Guy, J.; Kastan, N. R.; Robinson, N. D.; de Lima Alves, F.; et al. Rett syndrome mutations abolish the interaction of *MECP2* with the NCoR/SMRT co-repressor. *Nat Neurosci* **2013**, *16* (7), 898-902. DOI: 10.1038/nn.3434 From NLM Medline.
(24) Montiel-Gonzalez, M. F.; Vallecillo-Viejo, I.; Yudowski, G. A.; Rosenthal, J. J. C. Correction of mutations within the cystic fibrosis transmembrane conductance regulator by site-directed RNA editing. *Proceedings of the National Academy of Sciences of the United States of America* **2013**, *110* (45), 18285-18290. DOI: 10.1073/pnas.1306243110.
(25) Montiel-Gonźalez, M. F.; Vallecillo-Viejo, I. C.; Rosenthal, J. J. C. An efficient system for selectively altering genetic information within mRNAs. *Nucleic Acids Research* **2016**, *44* (21), 1-12. DOI: 10.1093/nar/gkw738.
(26) Ojha, N.; Diaz Quiroz, J. F.; Rosenthal, J. J. C. *In vitro* and *in cellula* site-directed RNA editing using the lambdaNDD-BoxB system. *Methods Enzymol* **2021**, *658*, 335-358. DOI: 10.1016/bs.mie.2021.06.009 From NLM Medline.
(27) Doherty, E. E.; Karki, A.; Wilcox, X. E.; Mendoza, H. G.; Manjunath, A.; Matos, V. J.; Fisher, A. J.; Beal, P. A. ADAR activation by inducing a syn conformation at guanosine adjacent to an editing site. *Nucleic Acids Res* **2022**, *50* (19), 10857-10868. DOI: 10.1093/nar/gkac897 From NLM Medline.
(28) Matthews, M. M.; Thomas, J. M.; Zheng, Y.; Tran, K.; Phelps, K. J.; Scott, A. I.; Havel, J.; Fisher, A. J.; Beal, P. A.; 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. DOI: 10.1038/nsmb.3203.Structures.
(29) Thuy-Boun, A. S.; Thomas, J. M.; Grajo, H. L.; Palumbo, C. M.; Park, S.; Nguyen, L. T.; Fisher, A. J.; Beal, P. A. Asymmetric dimerization of adenosine deaminase acting on RNA facilitates substrate recognition. *Nucleic Acids Research* **2020**, *48* (14), 7958-7972. DOI: 10.1093/nar/gkaa532.
(30) Wang, Y.; Havel, J.; Beal, P. A. A Phenotypic Screen for Functional Mutants of Human Adenosine Deaminase Acting on RNA 1. *ACS Chem Biol* **2015**, *10* (11), 2512-2519. DOI: 10.1021/acschembio.5b00711 From NLM Medline.
(31) Sinnamon, J. R.; Kim, S. Y.; Corson, G. M.; Song, Z.; Nakai, H.; Adelman, J. P.; Mandel, G. Site-directed RNA repair of endogenous *MECP2* RNA in neurons. *Proceedings of the National Academy of Sciences of the United States of America* **2017**, *114* (44), E9395-E9402. DOI: 10.1073/pnas.1715320114.

# Chapter 4

# *In Cellula* Application of Antisense Oligonucleotides

*This chapter contains excepts from the full manuscript which was published in ACS Chemical biology in April 2023[1].*

## 4.1: Introduction

Although *in vitro* assays with ADAR are valuable for understanding the ADAR reaction mechanism and obtaining kinetic information, cellular assays involving ADAR provide data that better replicates the editing processes that could take place *in vivo*. EMERGe-derived gRNAs will be tested for their ability to promote editing with*in cellula*r systems. A few of these systems, such as dual luciferase assays[2] and ADAR editases, [3-5] are poised to be the next step in gRNA application. Most cellular assays focus on using endogenous or exogenous ADAR. The ideal conditions for a cellular experiment would be to deliver a gRNA and use endogenous ADAR with an endogenous target. While using endogenous ADAR would be ideal, many cell lines do not express ADAR at the levels needed for cellular assays. [6] To remedy this lack of endogenous ADAR, ADAR is expressed exogenously. When exogenous ADAR is delivered via a plasmid, mutant or chimeric ADARs can be tested. Additionally, targets for the exogenous ADAR may be exogenously expressed because the tested cell lines may not naturally express the RNA targets. Therefore, some assays require exogenous ADAR, exogenous target and a gRNA. While this may not directly mimic an *in vivo* test, this enables a function model system to test gRNAs generated from screens.

EMERGe-derived gRNAs that worked *in vitro* were tested to observe their activity in cells. Promising results *in cellula*r experiments may justify further testing of EMERGe-derived gRNAs in therapeutic contexts. In addition to EMERGe-derived gRNAs, recently published chemical

modifications have also been tested *in cellula.* [7] These modified nucleotides have previously been observed to enable editing *in vitro* and are observed to promote editing within cells using the dual luciferase assay and an NGS based assay. Finally, a screen was designed to test 20 amino acids in a high-throughput fluorescence assay against EMERGe-derived substrates. While not yielding conclusive results, the limitations of a E488X ADAR2 yeast-based screen will be discussed.

## 4.2: Results

### 4.2.1: Testing chemically modified gRNAs in HEK293T cells.

The dual luciferase assay allows for a robust testing of gRNAs in Human Embryonic Kidney 239T (HEK293T) cells. [2] ADAR editing is observed through a restoration of luminescence. The reporter plasmid contains three key features. The first feature is the ADAR target site, which includes 90 nt of the *MECP2* transcript, either R168X or R255X for their respective assays. The second feature is the nano luciferase, [2] positioned downstream and in frame of the edit site. If the site is edited, the nano luciferase will be expressed and act as a reporter (Figure 4.1) since the stop codon is no longer present. The third and final feature is a firefly luciferase which is consistently expressed and serves as an internal control. HEK293T cells are transfected with an ADAR plasmid, a reporter plasmid, and a chemically modified oligonucleotide. The cells are lysed and luminescence values are obtained after a 48 h transfection. Since RNA is easily digested upon transfection, [8] modified oligonucleotides are used (Figure 4.2A). These antisense oligonucleotides (ASOs) contain no RNA and are synthesized to contain DNA, 2'-*O*-methyl (OME), and phosphorothioate modifications (Figure 4.3). This ASO modification pattern has shown *in cellula* activity, [9] and results in improved cellular stability. [10] Additionally, these modified oligonucleotides were used at varying concentrations (3 to 30 nM) in the screen to represent a dose response. The

ratio of nano luciferase to firefly luciferase activities is correlated to ADAR editing while also correcting for transformation efficiency[11].

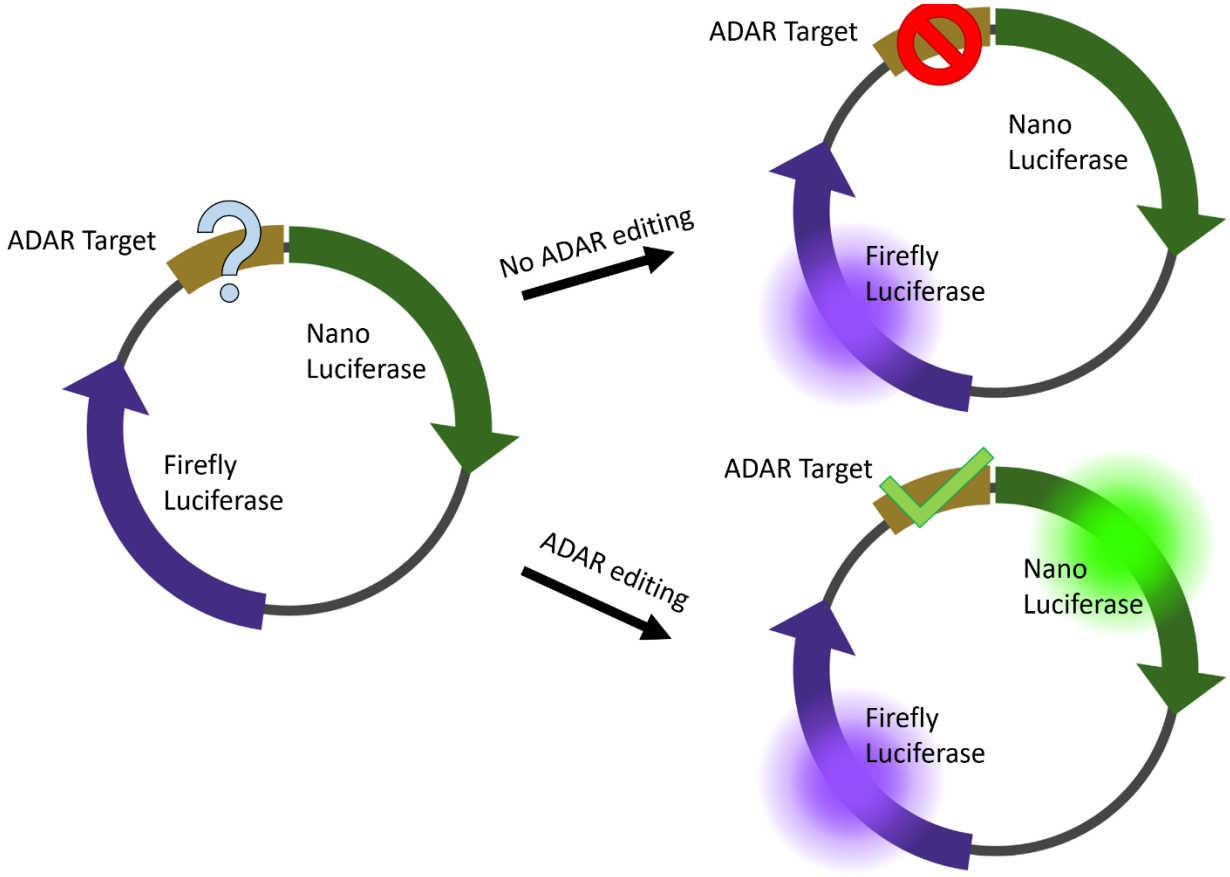


Figure 4.1: A general schematic of the dual luciferase assay.

In the absence of ADAR editing on the target sequence, the nano luciferase will not be translated. However, in the presence of ADAR editing on the target sequence, the nano luciferase will be translated, resulting in luminescence. In both scenarios, whether there is editing or no editing, the firefly luciferase will be expressed, serving as a control signal. Created in BioRender.

For the first experiment, *MECP2* R168X was used to assess how well EMERGe-derived sequences function in HEK293T cells. For this target, four different gRNAs (R168X-5, R168X-GG;AC, R168X-AC, and R168X-G3;AC) were synthesized using the modification pattern shown in Figure 4.3. The R168X-GG;AC guide had been previously reported as R168X-5b, but it was renamed to better represent the key motif. The R168X-G3;AC guide had not been tested before

*in cellula*, and it includes the 3deaza-dA modification at the -1 position. This modified base was described in the introduction, and it had been shown to promote editing when there is a 5'-G[7]. Due to the reporter plasmids having an internal transfection control, the ratio of nano luciferase to firefly luciferase is used to determine cellular editing (nLuc/fLuc). When using ADAR2, it was observed that the EMERGe selected gRNA (R168X-5) was inactive in this cellular experiment (Figure 4.4A).

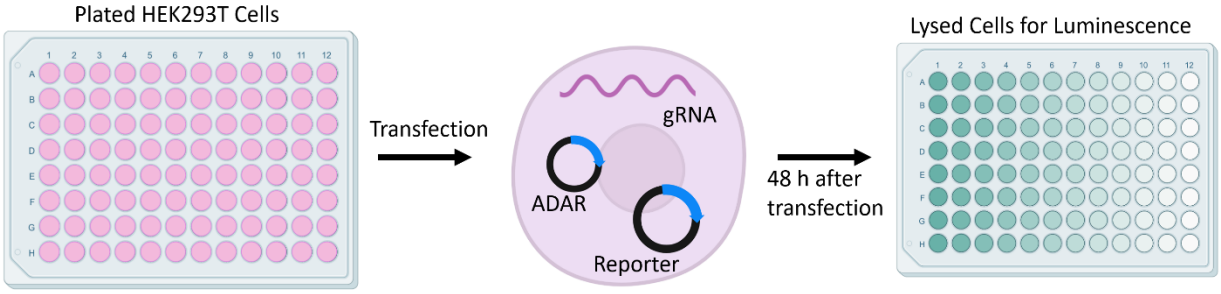


Figure 4.2: A general workflow of the dual luciferase assay and the tested ASOs.

HEK293T cells are transfected with lipid nanoparticles containing gRNA, ADAR plasmid, and Reporter plasmid. After 48 h cells are lysed and luminescence is recorded. Created in BioRender.

Additionally, the R168X-GG;AC and R168X-G3;AC gRNAs were similar at all concentrations. The R168X-AC guide showed no editing, and this served as a negative control. This same experiment was repeated except ADAR2 was replaced with ADAR1 P110 (Figure 4.4B). In this assay R168X-5 was also found to be inactive. Additionally, it was observed that R168X-G3;AC was more active than R168X-GG;AC at 3nM of gRNA. When the guide concentration was increased this preference was reversed. At the 30 nM concentration, the R168X-GG;AC guide was more active than the R168X-G3;AC guide.


Figure 4.3: The tested *MECP2* R168X duplexes for *in cellula* assays.

The ASO (bottom strand) contains 2′-O-Methyl modifications at all nucleotides except those that are bolded. Bolded nucleotides are DNA. Phosphorothioate modifications are installed between underlined nucleotides. A: R168X-5 substrate. B: R168X-GG;AC substrate. C: R168X-G3;AC substrate. D: R168X-AC substrate.

After observing *in cellula* editing with both R168X-GG;AC and R168X-G3;AC gRNA, I was interested in determining if these guides are functional enough to utilize HEK293T's endogenous ADAR levels. We conducted a similar experiment, replacing the dual luciferase assay with RNA isolation followed by NGS. Once the data was deconvoluted and separated based on a barcode sequence, we were able to determine the editing at 5 different sites (Figure 4.5A). Sites -60 and +62 were A's outside of the 90 nt *MECP2* R168X insert. Sites -14 and +5 were bystanders close to the target A. The target A, positioned at 0, is the *MECP2* R168X site. Through this experiment, it became apparent that the endogenous ADAR levels were not substantial enough in these cells to result in a meaningful degree of editing (Figure 4.5B).

The R168X-5 guide showed no editing with either ADAR2 or ADAR1 P110. Consequently, a test was conducted to see if the *in cellula* deamination conditions were operating at concentrations too low for this specific gRNA. An all-RNA *in vitro* experiment was developed to

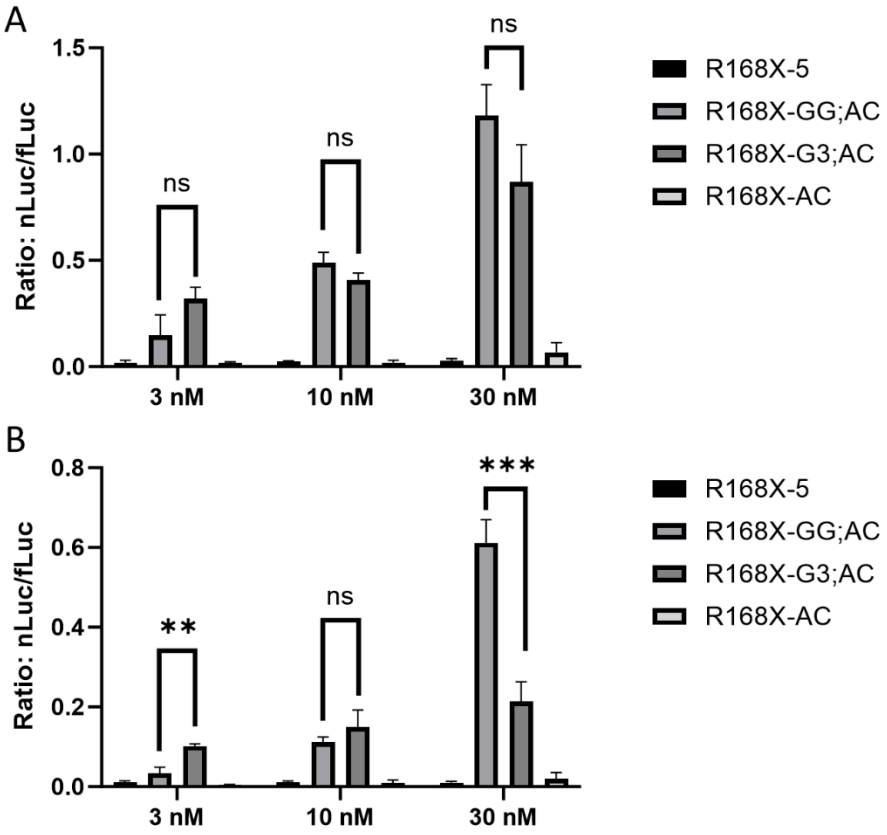


Figure 4.4: HEK293T dual luciferase assay with a *MECP2* R168X reporter

HEK293T cells transfected with variable ASO concentrations were observed after 48 h. A: The dual luciferase assay with ADAR2. B: The dual luciferase assay with ADAR1 P110. Data are plotted as the mean ± s.d. from three independent experiments. P values are calculated from Welch's t-tests (** $p < 0.01$ and *** $p < 0.001$).

view the effect of gRNA concentrations (Figure 4.6). As part of a trial experiment, it was observed that when the gRNA concentration was reduced to 10 nM, minimal to no editing was observed with the R168X-5 gRNA. This lower concentration condition could more closely resemble cellular conditions, when compared to our typical deamination procedure of 100 nM gRNA. This would explain the lack of editing with R168X-5 *in cellula*. It was also noted that the other guides exhibited a similar reduction in editing, although not to the same extent as the drop observed with R168X-5.

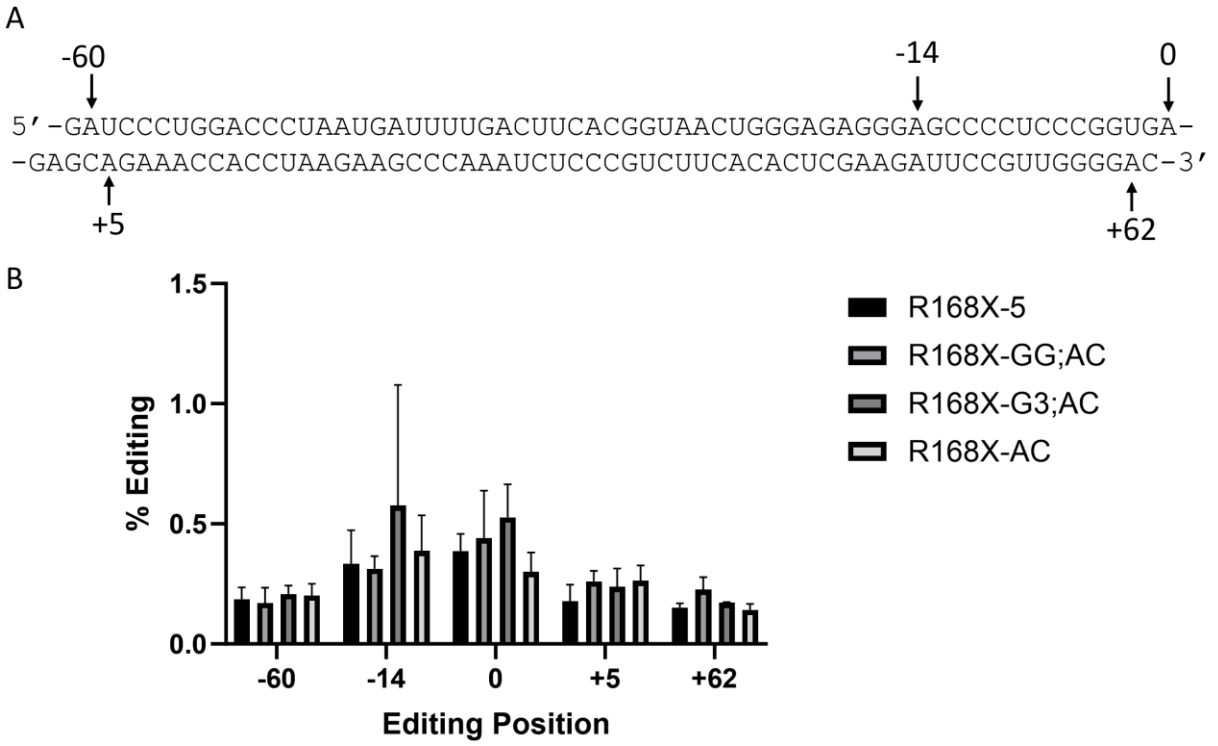


Figure 4.5: NGS determined percent editing from an HEK293T RNA isolation.

HEK293T cells transfected with 30 nM ASOs were observed after 48 h. A: The RNA isolate sent to NGS. Indicated nucleotides have their percent editing shown below. B: Endogenous ADAR was used to deaminate the RNA shown above. The percent editing level was determined through NGS reads. Data are plotted as the mean ± s.d. from three independent experiments.

Though R168X-5 may not have been a successful EMERGe-derived gRNA for cellular experiments, R255X-11 remained to be tested. A new reporter plasmid was created for this assay, containing 90 nt surrounding the *MECP2* R255X mutation. Four new gRNAs (R255X-11, R255X-GG;AC, R255X-AC, and R255X-G3;AC) were synthesized using the modification pattern shown in Figure 4.7. Similar to R168X, EMERGe-derived gRNA (R255X-11) was found to be inactive *in cellula* by a dual luciferase assay (Figure 4.8). When using ADAR2 or ADAR P110, it was observed that

R255X-GG;AC and R255X-G3;AC exhibited similar levels of activity. The R255X-AC guide served as a negative control in all conditions.

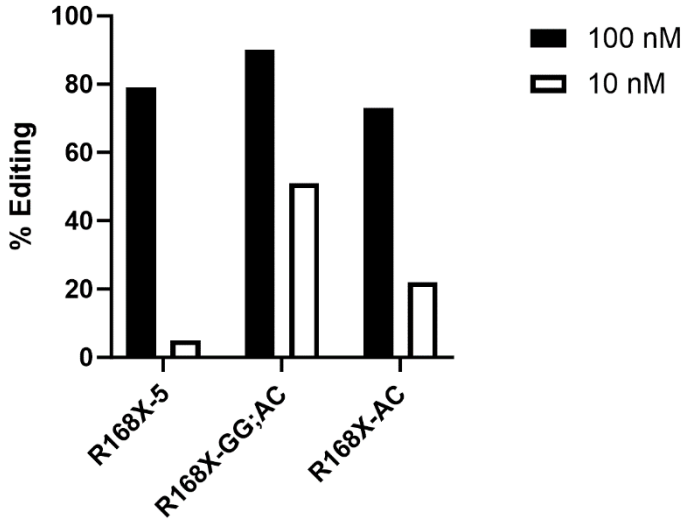


Figure 4.6: *In vitro* deamination to view the effect of gRNA concentration on % editing.

On-target editing observed with hADAR2 and select guide sequences on an *MECP2* R168X target. Repeated with all RNA substrates from Chapter 3. Percent editing at target site observed with 200 nM hADAR2 in a 2 h reaction.

Up to this point, no EMERGe-derived gRNAs have promoted editing in cells. Before proceeding with the *MECP2* R270X target, I wanted to test the ASO modification pattern with the R270X-24a gRNA. It has been observed that this 3'-CGG motif enables editing in an unknown manner. Additionally, there may be crucial interactions with the 2' hydroxyl of RNA's that are disrupted by the modifications. Four oligonucleotides were developed to assess R270X-24a's activity in a chemically modified oligonucleotide. These oligonucleotides contain varying degrees of modifications, ranging from all RNAs to the ASO design without phosphorothioates (Figure 4.9).

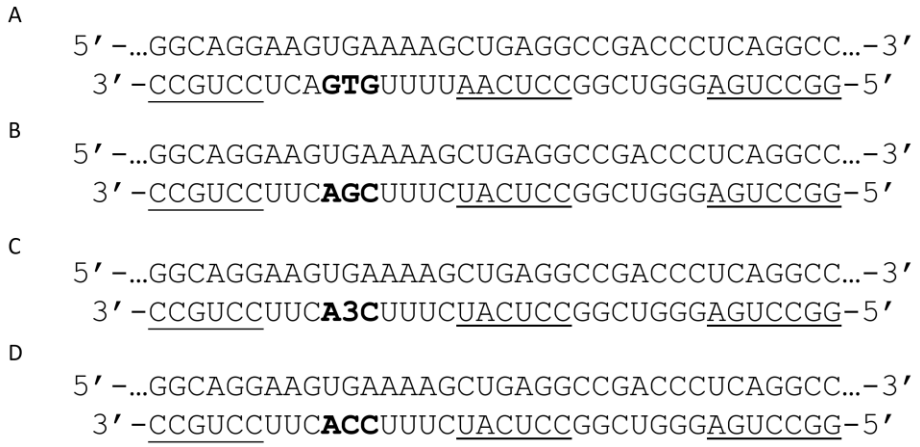


Figure 4.7: The tested *MECP2* R255X duplexes for *in cellula* assays.

The ASO (bottom strand) contains 2'-*O*-Methyl modifications at all nucleotides except those that are bolded. Bolded nucleotides are DNA. Phosphorothioate modifications are installed between underlined nucleotides. A: R255X-5 substrate. dT was used in place of dU due to its commercial availability. B: R255X-GG;AC substrate. C: R255X-G3;AC substrate. D: R255X-AC substrate.

There is a noticeable reduction in editing when the RNA changes from a T7 transcribed RNA to a synthesized all-RNA ASO, decreasing from 82 ± 1% to 47 ± 1% (Figure 4.9E). These two RNA's contain differences in binding registers. The T7 transcribed RNA uses the EMERGe 30 nt symmetric design tested in Chapter 3. The RNA ASO uses a 36 nt binding register that is 3' shifted with respect to the target. Reductions in editing occur with the addition of each modification. When the RNA ASO is replaced with DNA at positions -2, -1, and Orphan, the % editing decreases from 47 ± 1% to 16 ± 1%. It is also observed that when the RNA ASO gains 2'-*O*-Methyl modifications, the editing drops to just above the noise floor for Sanger sequencing (Figure 4.9C and D). Since R270X-24a does not appear to be compatible with our current ASO design, an encodable gRNA methodology was tested while new ASO modification patterns are developed.

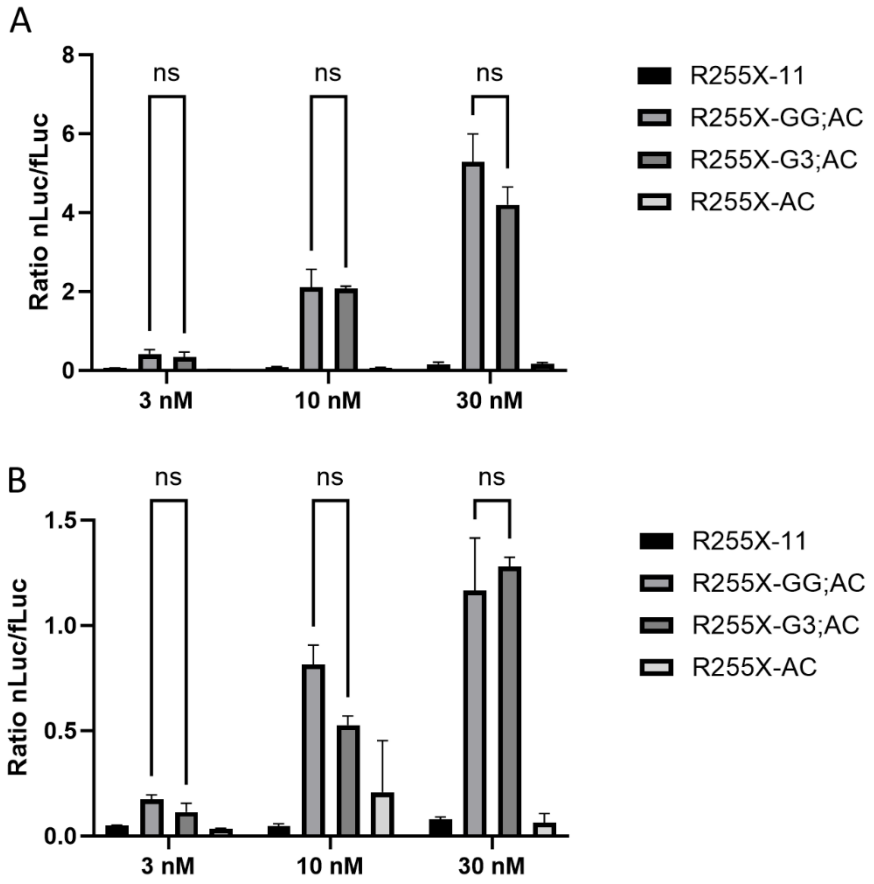


Figure 4.8: HEK293T dual luciferase assay with a *MECP2* R255X reporter

HEK293T cells transfected with variable ASO concentrations were observed after 48 h. A: The dual luciferase assay with ADAR2. B: The dual luciferase assay with ADAR1 P110. Data are plotted as the mean ± s.d. from three independent experiments.

**4.2.2: Adapting EMERGe-Derived gRNAs to an Editase system.**

Due to R270X-24a's inability to tolerate chemical modifications, it seemed like a promising candidate for an exogenously expressed gRNA system. For this purpose, we turned to our collaborators at the Mandel Lab to use an engineered Editase,[28] that has previously been used for editing *MECP2* mutations.[25] An *MECP2* guide RNA recruits the hADAR2 catalytic domain to the

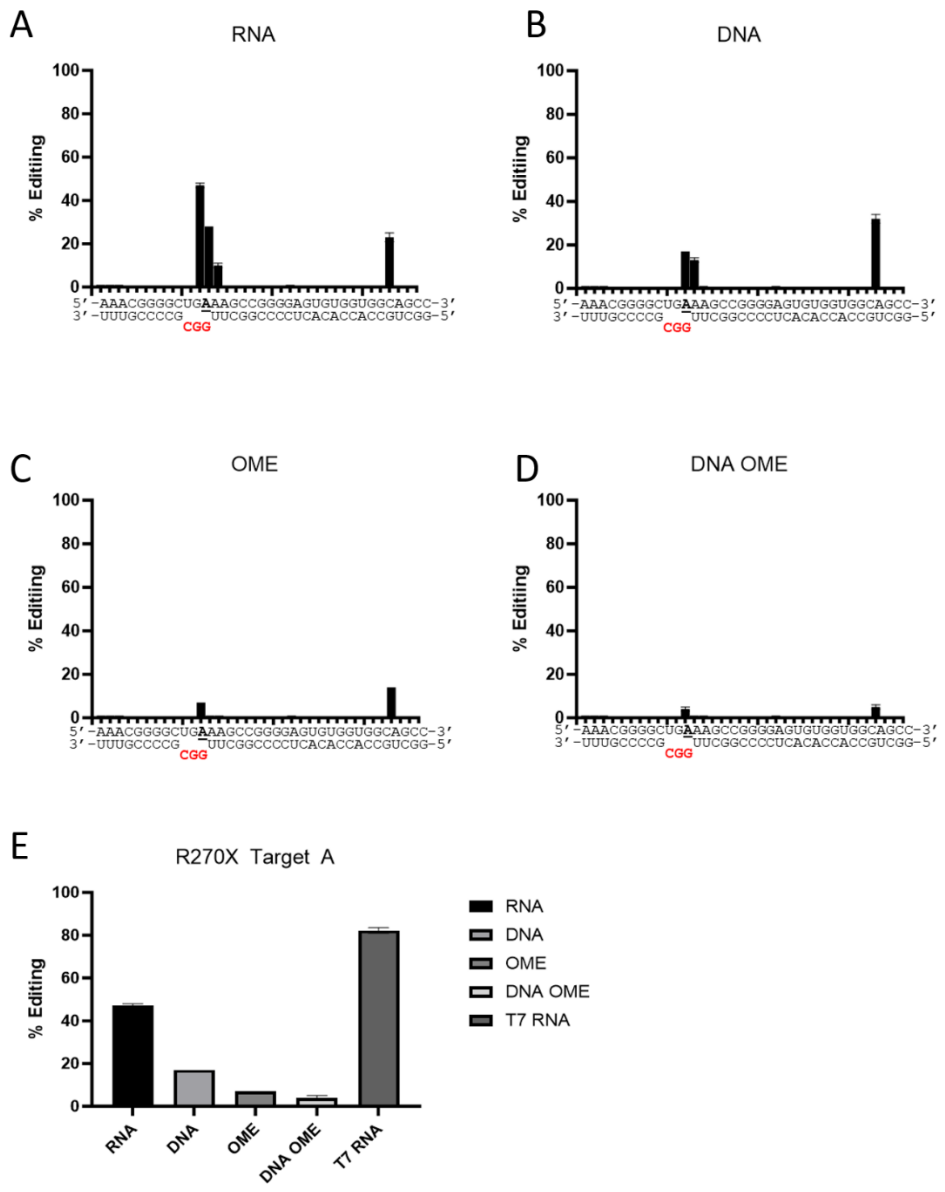


Figure 4.9: On-target and bystander editing observed with hADAR2 and modified R270X-24a ASOs.

Percent editing at target site and bystander sites observed with 200 nM hADAR2 in a 30 min reaction A: An all-RNA variation of the ASO design. B: RNA based ASO design containing DNA at the red nucleotides. C: A 2'-*O*-Methyl ASO design containing all 2'-*O*-Methyl but RNA at the red nucleotides. D: A 2'-*O*-Methyl ASO design containing all 2'-*O*-Methyl but DNA at the red nucleotides. E: On-target editing of the four listed oligonucleotides with a T7 transcription product gRNA as a positive control. Percent editing was with 200 nM hADAR2 in a 30 min reaction.

region in *MECP2* containing the target adenosine. The ADAR2 catalytic domain has been fused to the λN peptide from bacteriophage lambda (Editase) that recognizes two BoxB hairpins flanking the guide region (Figure 4.10A). 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 4.10C-D). The control transfections lacked a guide sequence (Figure 4.10B). The R270X-GG;AC guide bears two previously known enabling

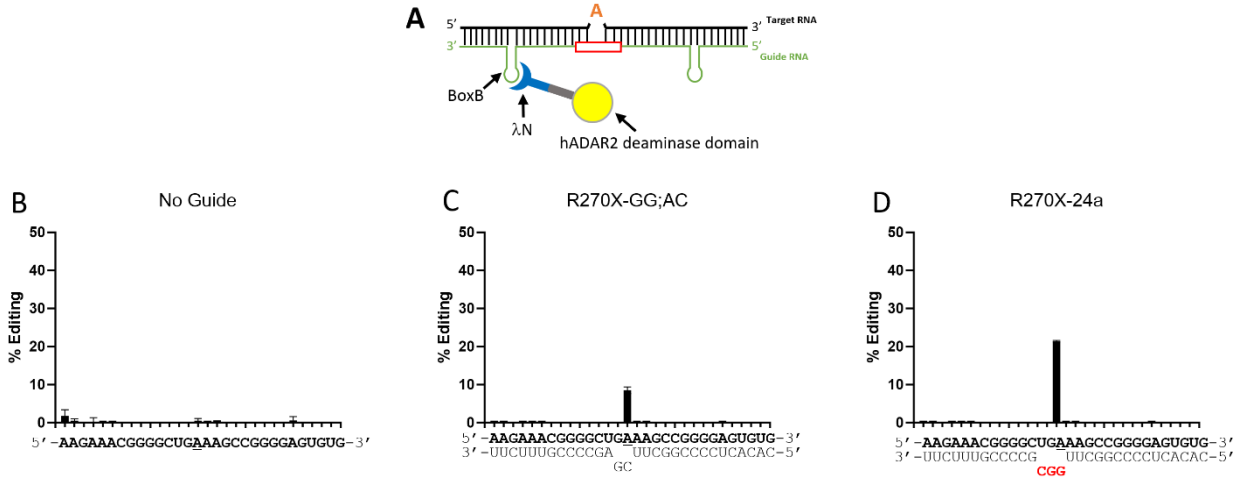


Figure 4.10: Directed editing in HEK293T cells using lN-BoxB Editase and R270X guide sequences.

A: The general design of the λN-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 (**** $p<0.0001$).

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 it is otherwise Watson-Crick complementary to the target site. After transfection of plasmids and after 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 4.10C-D). Thus, enabling sequences discovered in an EMERGe screen can direct ADAR editing in human cells with improved performance.

**4.2.3: Development of an E488X ADAR2 yeast screen.**

While I have previously shown E488X ADAR2 activities with RNA substrates, we originally wanted to test this system in yeast cells. This would remove the requirements of protein purifications and *in vitro* deaminations. For this experiment, a reporter bearing a *MECP2* R270X-24a was installed downstream of a start codon and upstream of a yeGFP (Figure 4.11A and B). This mimics the design used in Sat-FACS-Seq. [12] This reporter plasmid was transformed into yeast cells along with the ADAR2 E488X mutants. Once the cells were induced, fluorescence values were obtained (Figure 4.11C). Fold increase in yeGFP was used to correlate to editing levels. ADAR2 E396A was used as a negative control and baseline value for fluorescence. None of the common 20 amino acids were distinguishable from a catalytically dead E396A ADAR2 control. This may be due to the substrate being a poor ADAR context or a non-standard editing site complementary sequence.

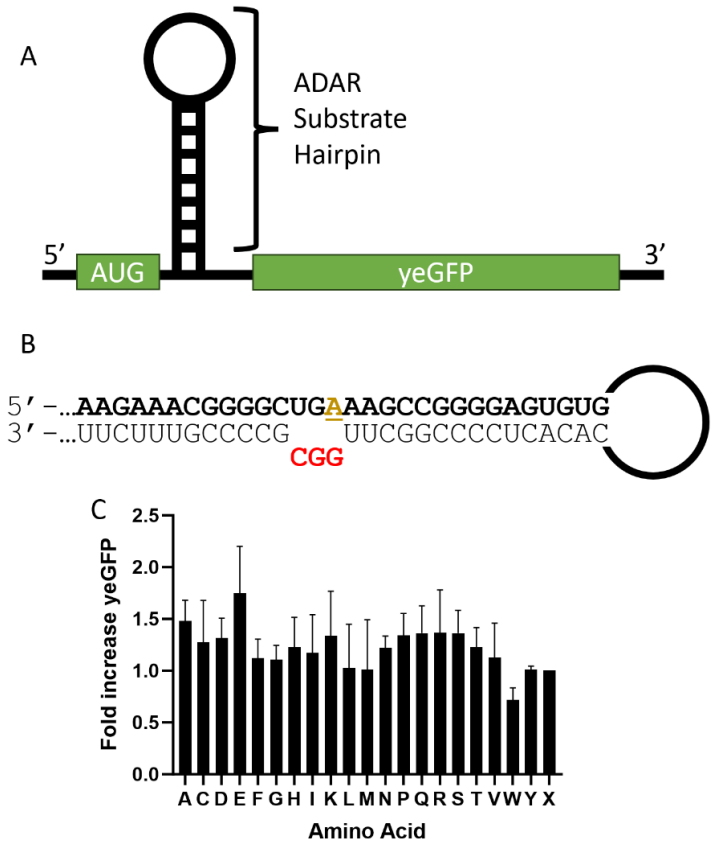


Figure 4.11**:** A yeast-based screen for E488X ADAR2 mutants.

A: The reporter transcript contains a hairpin down stream and in frame of the start codon. B: The reporter hairpin corresponding to the sequence present in the mRNA for human *MECP2* bearing the R270X mutation (bold), and the R270X-24a guide (bottom strand). This hairpin linker is the same as in the R270X library. The target A is shown in yellow. C: Fold increase of yeGFP is determined by a catalytically dead ADAR2 E396A. The amino acid listed as X is this ADAR2 E396A. Amino acids are listed by their single letter code.

A verification of the R270X target sequence was completed with a substrate containing the published -1 G:G pair[7] (Figure 4.12A). This substrate is also known to function well with wild type ADAR2 and the hyperactive E488Q mutant. When ADARs are expressed with the validation reporter in yeast cells, we observe an increase in editing with the hyperactive mutant (Figure 4.12B). This agrees with the results in Chapter 3. There is no editing observed with the wild type protein with this reporter. When this experiment is conducted with only three variants of ADAR,

there is an increase a measurable increase in editing. Therefore, sequence and codon dependencies may impede this screen with the R270X-24a hairpin.

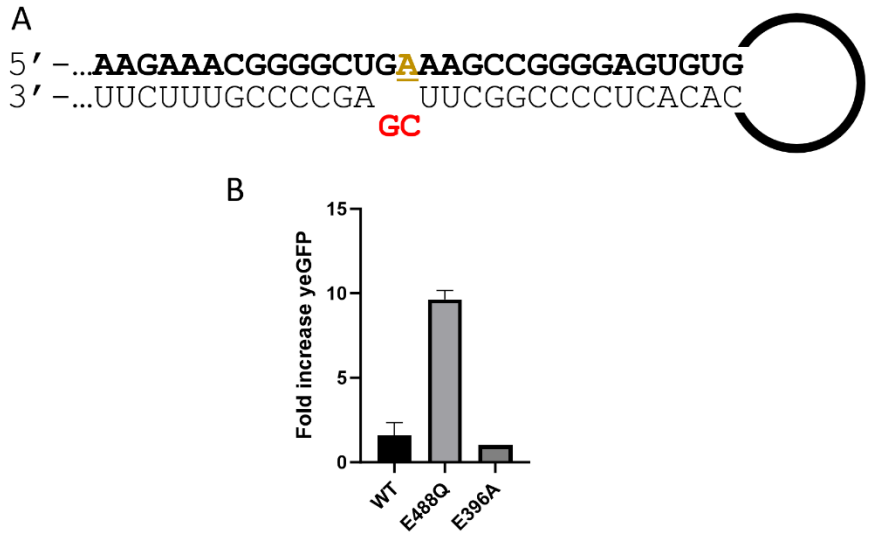


Figure 4.12: Validation of the hairpin design used for the E488X ADAR2 yeast screen.

A: The reporter hairpin corresponding to the sequence present in the mRNA for human *MECP2* bearing the R270X mutation (bold), and the R270X-GG;AC guide (bottom strand). This hairpin linker is the same as in the R270X library. The target A is shown in yellow. B: ADAR2s used for fluorescence values. Fold increase of yeGFP is determined by a catalytically dead ADAR2 E396A

### 4.3: Discussion

EMERGe-derived gRNAs initially appeared to be promising candidates for cellular studies due to their small motifs and appropriate length for in-cell applications. However, upon closer examination, it became evident that these gRNAs do not translate well into such experiments due to certain artifacts arising from EMERGe. The primary limitation in transitioning from *in vitro* to *in cellula* studies is associated with EMERGe's reverse transcriptase (RT) requirement. This stems from the RT step being carried out at 45°C, resulting in artifacts that reduce the overall complementarity of the duplex. We hypothesize that the RT enzyme selects for mismatches that decrease the overall melting temperature ($T_m$) of the RNA. When applied *in cellula*, the overall

concentration of the gRNA and target RNA is low enough that this reduction in $T_m$ reduces the formation of double-stranded RNA (dsRNA). This may explain why R168X-5 and R255X-11 did not promote editing *in cellula* despite their success *in vitro*.

Another explanation for the poor transition from *in vitro* to *in cellula* may be a lack of tolerance for the ASO chemical modification pattern. This intolerance was observed with R270X-24a by an *in vitro* study. R270X-24a's ideal substrate was observed to be a symmetric T7 transcription product that mimics the EMERGe screening product. A binding register shift reduces editing. When chemical modifications were added, there was also a significant reduction in editing. It was observed that the 3'-CGG motif requires ribonucleotides, at each of those 3 nt, to promote editing. The R270X-24a gRNA will need further testing to determine a suitable ASO modification pattern for cell use.

Additionally, I verified that chemical modifications from a recent publication[7] promote editing *in cellula*. The gRNAs that were successful in promoting editing contained a G:G pair and a G:3deaza-dA pair. These guides have been previously used *in vitro* to enhance editing with a 5'-G[7]. It was previously observed that the G:3deaza-dA pair was more effective at promoting editing than the G:G pair. The results from my verification indicate that their function *in cellula* is both sequence dependent and ADAR dependent. While the *MECP2* R168X target was observed to have significant differences in editing levels with ADAR1 p110, this could not be repeated in any other condition. Additionally, this similarity in editing may be due to the ASO's chemical modification pattern reducing the -1 position's influence on the overall activity. To better understand how this -1 position effects editing *in cellula*, different modifications and sequence contexts will have to be tested.

In contrast to EMERGe-derived ASOs not yet working *in cellula*, 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 4.10). 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. [13]

Finally, I aimed to develop a yeast assay capable of quickly testing each of the common 20 amino acids at a specific position within ADAR. To achieve this, I created a yeGFP assay, similar to Sat-FACS-Seq. With this assay 21 transformations could be executed in parallel, the common 20 amino acids and 1 catalytically dead ADAR control. This approach enabled the completion of a full screen with defined substrates and ADARs, as opposed to libraries. However, this assay lacked the sensitivity required to complete a successful screening. While no specific amino acid was observed to enable editing with the *MECP2* R270X-24a reporter, this experiment provided perspective on what is required in a high-throughput ADAR screen.

Protein expression is an unmentioned advantage of Sat-FACS-Seq. While the screen I conducted averaged results across all cells within a growth, FACS allows for the sorting of single cells. This single-cell sorting not only permits the selection of functional ADAR mutants but also inherently biases the data toward mutant proteins that have been well-expressed. I hypothesize that the limiting factor in this E488X ADAR2 screen is the expression of the protein and the

reporter. This unique feature of Sat-FACS-Seq positions it as the gold standard for investigating ADARs in yeast. In contrast, EMERGe is considerably better suited for the study of gRNAs.

## 4.4: Methods

**4.4.1: ASO synthesis and purification.** *MECP2* R168X guides, with the exception of R168X-G3;AC, and *MECP2* R270X guides were purchase and purified from Azenta Life Sciences. All reagents and phosphoramidites were purchased from Glen research. R168X-G3;AC and all *MECP2* R255X guides were synthesized on an solid-phase oligonucleotide synthesizer (ABI 394) using the 0.2 μmol scale methodology. All guides were cleaved from the controlled pore glass (CPG) with 3:1 ammonium hydroxide:ethanol at 60°C overnight. The supernatant was removed from the CPGs and dried in vacuo. Guides were then purified by denaturing urea-PAGE and visualized by UV shadowing. Bands were excised and gel slices were crushed and soaked in 0.5 M $NH_4OAc$ and 1 mM EDTA overnight at 4°C. Gel fragments were removed from the soaking buffer using 0.2 μm cellulose acetate filters. The oligonucleotides were then precipitated out of solution by Ethanol precipitation at -70°C. The supernatant was removed, and the remaining oligonucleotide precipitate was dried in vacuo. The oligonucleotides were resuspended in nuclease-free water, quantified by 260 nm UV absorbance, and confirmed by ESI-MS at Novatia.

**4.4.2: Cloning of dual luciferase reporter plasmids.** Dual luciferase reporter plasmids and dsDNA inserts were digested using BamHI (NEB R0136S) and NotI (R0189S). The digested plasmid was then purified by gel extraction using 0.5% agarose, while the insert was purified by gel extraction using 1.5% agarose (Qiagen 28706). The insert and plasmid were ligated using T4 DNA ligase (NEB M0202S). The resulting ligation plasmid was transformed into XL10-Gold Ultracompetent Cells

(Agilent 200314) and purified (Promega A1223). Sequences for reporter plasmids were confirmed by Sanger sequencing. Inserted dsDNA was purchased from IDT.

**4.4.3: Cellular editing of ADARs in an Overexpressed system.** HEK293T cells were cultured in Dulbecco's modified Eagle's medium (DMEM) (Gibco), 10% fetal bovine serum (FBS) (Gibco) and 1% anti-anti (Gibco) at 37°C and 5% $CO_2$. When the culture reached ~90% confluency, 8 x $10^8$ cells were transferred into opaque walled 96 well plates. After 24 h, cells were transfected. A mixture was made for delivery: 50 ng of dual luciferase plasmid, 150 ng of ADAR plasmid, 0.3 µl of P300 and variable concentrations of gRNAs. In addition to these reagents, every well was delivered 0.5 µl Lipofectamine 3000 and 0.5 µl of Opti-MEM Reduced Serum Media. This solution was vortexed prior to addition of the plasmid mixture. Plasmids and lipofectamine solutions were combined and allowed to equilibrate for 10 min prior to addition into the wells. Once the solution was added, the cells incubated at 37°C and 5% $CO_2$ for 48 h. After the 48 h, luminescence reagents were added to the wells (Nano-Glo Dual-Luciferase Reporter Assay, Promega N1610) and luminescence was measured (BMG Labtech). In the case of NGS editing, the cells were lysed after 48 h and RNA was collected (RNAqueos Total RNA Isolation Kit, ThermoFisher AM1912). The collected RNA was used for RT-PCR (Access RT-PCR, Promega A1260) followed by nested PCR (NEB M0530L). The purified product was subjected to NGS at Genewiz (NGS-EZ). Percent editing was determined by HTStream and RStudio as described in Chapter 2. The Rstudio code was altered to determine barcode regions instead of $N_{10}$ regions for data deconvoluting.

**4.4.4: Directed editing in HEK293T cells using λN-BoxB Editase.** All cloning primers and guide sequences are included in Tables 4.12-13. All plasmids were completely sequenced. Plasmid pGM1090 contains the wild type hADAR2 deaminase domain under control of the CMV promoter

(Editase)[4]. 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[13], 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[4], 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[4]. 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% CO2 humidified incubator. Transfection, processing, and quantification of RNA editing was performed as previously described[4]. Cells were seeded into a 12-well plate at 1.25 x 105 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 4.14). 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[3]. All transfection data is shown as reverse complement.

**4.4.5: Determination of yeGFP fluorescence with Mutant ADAR2 E488X.** Methods were repeated as described in Chapter 2 (Sat-FACS-Seq Reporter Plasmid Generation and Validation of yeGFP reporters for Sat-FACS-Seq) and Chapter 3 (Site-directed mutagenesis).

## 4.5: Tables of oligonucleotides

Table 4.1**:** R168X insert for Dual luciferase Assay (Figure 4.4 and Figure 4.5)

| Name | Sequence 5’ -> 3’ |
|---|---|
| *MECP2* R168X Reporter dsDNA insert with stop | AGGAGGACCTGGAATTCATGGGATACCCCTACGACGTGCCCGACTACGCCGG ATCCCCCTGGACCCTAATGATTTTGACTTCACGGTAACTGGGAGAGGGAGCC CCTCCCGGTGAGAGCAGAAACCACCTAAGAAGCCCAAATCTCGTCTTCACACT CGAAGATTCCGTTGGGGACTGGCGACAGACAGCCGGCTACAACCTGGACCA AGTCCTTGAACAGGGAGGTGTGTCCAGTTTGTTTCAGAATCTCGGGGTGTCC GTAACTCCGATCCAAAGGATTGTCCTGAGCGGTGAAAATGGGCTGAAGATCG ACATCCATGTCATCATCCCGTATGAAGGTCTGAGCGGCGACCAAATGGGCCA GATCGAAAAAATTTTTAAGGTGGTGTACCCTGTGGATGATCATCACTTTAAG GTGATCCTGCACTATGGCACACTGGTAATCGACGGGGTTACGCCGAACATGA TCGACTATTTCGGACGGCCGTATGAAGGCATCGCCGTGTTCGACGGCAAAAA GATCACTGTAACAGGGACCCTGTGGAACGGCAACAAAATTATCGACGAGCG CCTGATCAACCCCGACGGCTCCCTGCTGTTCCGAGTAACCATCAACGGAGTG ACCGGCTGGCGGCTGTGCGAACGCATTCTGGCGTAAGCGGCCGCTGGCCGC AATAAAATATCTTTATTTTCATTACATCTGTGTGTTGG |
| *MECP2* R168X Reporter dsDNA insert with nonstop (G at the edit site) | AGGAGGACCTGGAATTCATGGGATACCCCTACGACGTGCCCGACTACGCCGG ATCCCCCTGGACCCTAATGATTTTGACTTCACGGTAACTGGGAGAGGGAGCC CCTCCCGGTGGGAGCAGAAACCACCTAAGAAGCCCAAATCTCGTCTTCACAC TCGAAGATTCCGTTGGGGACTGGCGACAGACAGCCGGCTACAACCTGGACC AAGTCCTTGAACAGGGAGGTGTGTCCAGTTTGTTTCAGAATCTCGGGGTGTC CGTAACTCCGATCCAAAGGATTGTCCTGAGCGGTGAAAATGGGCTGAAGATC GACATCCATGTCATCATCCCGTATGAAGGTCTGAGCGGCGACCAAATGGGCC AGATCGAAAAAATTTTTAAGGTGGTGTACCCTGTGGATGATCATCACTTTAAG GTGATCCTGCACTATGGCACACTGGTAATCGACGGGGTTACGCCGAACATGA TCGACTATTTCGGACGGCCGTATGAAGGCATCGCCGTGTTCGACGGCAAAAA GATCACTGTAACAGGGACCCTGTGGAACGGCAACAAAATTATCGACGAGCG CCTGATCAACCCCGACGGCTCCCTGCTGTTCCGAGTAACCATCAACGGAGTG ACCGGCTGGCGGCTGTGCGAACGCATTCTGGCGTAAGCGGCCGCTGGCCGC AATAAAATATCTTTATTTTCATTACATCTGTGTGTTGG |
| FWD R168X stop insert | AGGAGGACCTGGAATTCA |
| RVS R168X stop insert | GGACACACCTCCCTGTTCAA |
| RVS R168X stop insert for RT | CCAACACACAGATGTAATGAAA |

Table 4.2: *MECP2* R168X ASO sequences. (Figure 4.3, Figure 4.4, and Figure 4.5)

| Name | Sequence 5’ -> 3’ |
|---|---|
| R168X-5 ASO | GGGCUUCUUAGGUGGUUUCAGUUCCGAUUGGGAGGG |
| R168X-GG;AC ASO | GGGCUUCUUAGGUGGUUUCUGCUCCGACCGGGAGGG |
| R168X-G3;AC ASO | GGGCUUCUUAGGUGGUUUCUGCUCC(3deaza-dA)ACCGGGAGGG |
| R168X-AC ASO | GGGCUUCUUAGGUGGUUUCUGCUCCCACCGGGAGGG |

Table 4.3: *MECP2* R168X ASO masses. (Figure 4.3, Figure 4.4, and Figure 4.5)

| Name | Calculated Mass (g/mol) | Observed Mass (m/z) |
|---|---|---|
| R168X-5 ASO | 10304.9 | 10308.3 |
| R168X-GG;AC ASO | 12277.0 | 12276.7 |
| R168X-G3;AC ASO | 12262.0 | 12261.3 |
| R168X-AC ASO | 10238.8 | 10242.1 |

Table 4.4: Nested PCR Primers with Barcodes (Figure 4.5)

| Name | Sequence 5’ -> 3’ |
|---|---|
| FWD Nested Primer with A barcode | ATCGAAAACTACGACGTGCCCGA |
| FWD Nested Primer with T barcode | ATCGTTTTCTACGACGTGCCCGA |
| FWD Nested Primer with C barcode | ATCGCCCCCTACGACGTGCCCGA |
| FWD Nested Primer with G barcode | ATCGGGGGCTACGACGTGCCCGA |
| RVS Nested Primer | CAGGTTGTAGCCGGC |

Table 4.5: *MECP2* R255X insert for Dual luciferase Assay. (Figure 4.8)

| Name | Sequence 5’ -> 3’ |
|---|---|
| *MECP2* R255X Reporter dsDNA insert with stop | AGGAGGACCTGGAATTCATGGGATACCCCTACGACGTGCCCGACTACG<br>CCGGATCCACCACATCCACCCAGGTCATGGTGATCAAACGCCCCGGCAG<br>GAAGTGAAAAGCTGAGGCCGACCCTCAGGCCATTCCCAAGAAACGGGG<br>CGTCTTCACACTCGAAGATTCCGTTGGGGACTGGCGACAGACAGCCGG<br>CTACAACCTGGACCAAGTCCTTGAACAGGGAGGTGTGTCCAGTTTGTTT<br>CAGAATCTCGGGGTGTCCGTAACTCCGATCCAAAGGATTGTCCTGAGCG<br>GTGAAAATGGGCTGAAGATCGACATCCATGTCATCATCCCGTATGAAGG<br>TCTGAGCGGCGACCAAATGGGCCAGATCGAAAAAATTTTTAAGGTGGT<br>GTACCCTGTGGATGATCATCACTTTAAGGTGATCCTGCACTATGGCACA<br>CTGGTAATCGACGGGGTTACGCCGAACATGATCGACTATTTCGGACGGC<br>CGTATGAAGGCATCGCCGTGTTCGACGGCAAAAAGATCACTGTAACAG<br>GGACCCTGTGGAACGGCAACAAAATTATCGACGAGCGCCTGATCAACC<br>CCGACGGCTCCCTGCTGTTCCGAGTAACCATCAACGGAGTGACCGGCTG<br>GCGGCTGTGCGAACGCATTCTGGCGTAAGCGGCCGCTGGCCGCAATAA<br>AATATCTTTATTTTCATTACATCTGTGTGTTGG |
| *MECP2* R255X Reporter dsDNA insert with nonstop (G at the edit site) | AGGAGGACCTGGAATTCATGGGATACCCCTACGACGTGCCCGACTACG<br>CCGGATCCACCACATCCACCCAGGTCATGGTGATCAAACGCCCCGGCAG<br>GAAGTGGAAAGCTGAGGCCGACCCTCAGGCCATTCCCAAGAAACGGGG<br>CGTCTTCACACTCGAAGATTCCGTTGGGGACTGGCGACAGACAGCCGG<br>CTACAACCTGGACCAAGTCCTTGAACAGGGAGGTGTGTCCAGTTTGTTT<br>CAGAATCTCGGGGTGTCCGTAACTCCGATCCAAAGGATTGTCCTGAGCG<br>GTGAAAATGGGCTGAAGATCGACATCCATGTCATCATCCCGTATGAAGG<br>TCTGAGCGGCGACCAAATGGGCCAGATCGAAAAAATTTTTAAGGTGGT<br>GTACCCTGTGGATGATCATCACTTTAAGGTGATCCTGCACTATGGCACA<br>CTGGTAATCGACGGGGTTACGCCGAACATGATCGACTATTTCGGACGGC<br>CGTATGAAGGCATCGCCGTGTTCGACGGCAAAAAGATCACTGTAACAG<br>GGACCCTGTGGAACGGCAACAAAATTATCGACGAGCGCCTGATCAACC<br>CCGACGGCTCCCTGCTGTTCCGAGTAACCATCAACGGAGTGACCGGCTG<br>GCGGCTGTGCGAACGCATTCTGGCGTAAGCGGCCGCTGGCCGCAATAA<br>AATATCTTTATTTTCATTACATCTGTGTGTTGG |
| FWD R255X stop insert | AGGAGGACCTGGAATTCA |
| RVS R255X stop insert | GGACACACCTCCCTGTTCAA |
| RVS R255X stop insert for RT | CCAACACACAGATGTAATGAAA |

Table 4.6: *MECP2* R255X ASO sequences. (Figure 4.7 and Figure 4.8)

| Name | Sequence 5’ -> 3’ |
|---|---|
| R255X-11 ASO | GGCCUGAGGGUCGGCCUCAAUUUUGTGACUCCUGCC |
| R255X-GG;AC ASO | GGCCUGAGGGUCGGCCUCAGCUUUCGACUUCCUGCC |
| R255X-G3;AC ASO | GGCCUGAGGGUCGGCCUCAGCUUUC(3deaza-dA)ACUUCCUGCC |
| R255X-AC ASO | GGCCUGAGGGUCGGCCUCAGCUUUCCACUUCCUGCC |

Table 4.7: *MECP2* R255X ASO masses. (Figure 4.7 and Figure 4.8)

| Name | Calculated Mass (g/mol) | Observed Mass (m/z) |
|---|---|---|
| R255X-11 ASO | 12172.2 | 12172.4 |
| R255X-GG;AC ASO | 12133.8 | 12133.6 |
| R255X-G3;AC ASO | 12116.8 | 12116.8 |
| R255X-AC ASO | 12093.7 | 12093.8 |

Table 4.8: R270X DNA templates for RNA target, primers, and RNA (Figure 4.9)

| Name | Sequence 5’ -> 3’ |
|---|---|
| R270X dsDNA template for RNA target. Coding strand shown. | CACGATTAATACGACTCACTATAGGGCCTTGTCAAGATGCCTTTTCAAACTTCGCCAGGGGGCAAGGCTGAGGGGGGTGGGGCCACCACATCCACCCAGGTCATGGTGATCAAACGCCCCGGCAGGAAGCGAAAAGCTGAGGCCGACCCTCAGGCCATTCCCAAGAAACGGGGCTGAAAGCCGGGGAGTGTGGTGGCAGCCGCTGCCGCCGAGGCCAAAAAGAAAGCCGTGAAGGAGTCTTCTATCCGATCTGTGCAGGAGACCGTACTCCCCATCAAGAAGCGCAAGACCCGGGAGACGGTCAGCATCGAGGTCAAGGAAGTGGT |
| R270X Forward Primer | CACGATTAATACGACTCACTATAGG |
| R270X Reverse | ACCACTTCCTTGACCTCG |
| R270X Forward Primer for RT-PCR | GGGCCTTGTCAAGATGC |
| R270X RNA Target | GGGCCUUGUCAAGAUGCCUUUUCAAACUUCGCCAGGGGGCAAGGCUGAGGGGGGUGGGGCCACCACAUCCACCCAGGUCAUGGUGAUCAAACGCCCCGGCAGGAAGCGAAAAGCUGAGGCCGACCCUCAGGCCAUUCCCAAGAAACGGGGCUGAAAGCCGGGGAGUGUGGUGGCAGCCGCUGCCGCCGAGGCCAAAAAGAAAGCCGUGAAGGAGUCUUCUAUCCGAUCUGUGCAGGAGACCGUACUCCCCAUCAAGAAGCGCAAGACCCGGGAGACGGUCAGCAUCGAGGUCAAGGAAGUGGU |

Table 4.9: *MECP2* R270X modified ASO sequences with modification pattern. (Figure 4.9)

| Name | Sequence 5’ -> 3’ |
|---|---|
| R270X-24a RNA ASO | rGrGrCrUrGrCrCrArCrCrArCrArCrUrCrCrCrCrGrGrCrUrU rGrGrCrGrCrCrCrCrGrUrUrU |
| R270X-24a DNA ASO | rGrGrCrUrGrCrCrArCrCrArCrArCrUrCrCrCrCrGrGrCrUrU dGdGdCrGrCrCrCrCrGrUrUrU |
| R270X-24a Ome ASO | mGmGmCmUmGmCmCmAmCmCmAmCmAmCmUmC mCmCmCmGmGmCmUmUrGrGrCmGmCmCmCmCmG mUmUmU |
| R270X-24a DNA Ome ASO | mGmGmCmUmGmCmCmAmCmCmAmCmAmCmUmC mCmCmCmGmGmCmUmUdGdGdCmGmCmCmCmCm GmUmUmU |

Table 4.10: *MECP2* R270X modified ASO masses. (Figure 4.7)

| Name | Calculated Mass (g/mol) | Observed Mass (m/z) |
|---|---|---|
| R270X-24a RNA ASO | 11363.8 | 11364.5 |
| R270X-24a DNA ASO | 11315.8 | 11317.0 |
| R270X-24a Ome ASO | 11825.8 | 11827.5 |
| R270X-24a DNA Ome ASO | 11777.8 | 11778.9 |

Table 4.11: Generation of pGM1524 (*MECP2* R270X) (Figure 4.10)

| Name | Sequence 5’ -> 3’ |
|---|---|
| huR270X SDM Top | AAACGGGGCTGAAAGCCGGGGAG |
| huR270X SDM Bottom | CTTGGGAATGGCCTGAGGGTCG |

Table 4.12: DNA templates to be used for gRNA production within HEK293T experiments. (Figure 4.10)

| Name | Sequence 5’ -> 3’ |
|---|---|
| *MECP2* R270X-GG;AC Sense: | caccGCTGCCACCAGGCCCTGAAAAAGGGCCTCCCCGGCTTCGAGCCCCGTTTCTTGGGAGGCCCTGAAAAAGGGCCGCCTGAGGGT |
| *MECP2* R270X-GG;AC Antisense: | aaaaACCCTCAGGCGGCCCTTTTTCAGGGCCTCCCAAGAAACGGGGCTCGAAGCCGGGGAGGCCCTTTTTCAGGGCCTGGTGGCAGC |
| *MECP2* R270X-CGG Sense: | caccGCTGCCACCAGGCCCTGAAAAAGGGCCTCCCCGGCTTGGCGCCCCGTTTCTTGGGAGGCCCTGAAAAAGGGCCGCCTGAGGGT |
| *MECP2* R270X-CGG Antisense: | aaaaACCCTCAGGCGGCCCTTTTTCAGGGCCTCCCAAGAAACGGGGCGCCAAGCCGGGGAGGCCCTTTTTCAGGGCCTGGTGGCAGC |

Table 4.13: *MECP2* cDNA PCR amplification primers (Figure 4.10)

| Name | Sequence 5’ -> 3’ |
|---|---|
| FLAG Seq-B Fwd | ATGGATTACAAGGATGACGATGA |
| *MECP2* Rev | CTTCAGCTAACTCTCTCGG |

Table 4.14: *MECP2* R270X Sanger sequencing primer (Figure 4.10)

| Name | Sequence 5’ -> 3’ |
|---|---|
| *MECP2* E1 1086 Rev | GCTGCTCTCCTTGCTTTTCC |

Table 4.15: *MECP2* R270X insert for yeGFP assay (Figure 4.11)

| Name | Sequence 5’ -> 3’ |
|---|---|
| dsDNA insert for *MECP2* R270X Reporter with CGG (coding strand) | CGACTCACTATAGGGAATATTAAGCTTATGAAGAAACGGGGCTGAAAGCCGGGGAGTGTGTTTCTTTCTTTCCACACTCCCCGGCTTGGCGCCCCGTTTCTTTCTAAAGGTGAAGAATTATTCACTGG |

Table 4.16: *MECP2* R270X insert for yeGFP assay for Validation (Figure 4.12)

| Name | Sequence 5’ -> 3’ |
|---|---|
| dsDNA insert for *MECP2* R270X Reporter with validation sequence (coding strand) | CGACTCACTATAGGGAATATTAAGCTTATGAAGAAACGGGGCTGAAAGCCGGGGAGTGTGTTTCTTTCTTTCCACACTCCCCGGCTTCGAGCCCCGTTTCTTTCTAAAGGTGAAGAATTATTCACTGG |

## 4.6: References

(1) Jacobsen, C. S.; Salvador, P.; Yung, J. F.; Kragness, S.; Mendoza, H. G.; Mandel, G.; Beal, P. A. Library Screening Reveals Sequence Motifs That Enable ADAR2 Editing at Recalcitrant Sites. *ACS Chem Biol* **2023**. DOI: 10.1021/acschembio.3c00107 From NLM Publisher.
(2) Brinkman, H. F.; Jauregui Matos, V.; Mendoza, H. G.; Doherty, E. E.; Beal, P. A. Nucleoside analogs in ADAR guide strands targeting 5'-UA sites. *RSC Chem Biol* **2023**, *4* (1), 74-83. DOI: 10.1039/d2cb00165a From NLM PubMed-not-MEDLINE.
(3) Sinnamon, J. R.; Jacobson, M. E.; Yung, J. F.; Fisk, J. R.; Jeng, S.; McWeeney, S. K.; Parmelee, L. K.; Chan, C. N.; Yee, S. P.; 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. DOI: 10.1073/pnas.2206053119 From NLM Medline.
(4) Sinnamon, J. R.; Kim, S. Y.; Corson, G. M.; Song, Z.; Nakai, H.; Adelman, J. P.; Mandel, G. Site-directed RNA repair of endogenous *MECP2* RNA in neurons. *Proceedings of the National Academy of Sciences of the United States of America* **2017**, *114* (44), E9395-E9402. DOI: 10.1073/pnas.1715320114.
(5) Montiel-Gonzalez, M. F.; Vallecillo-Viejo, I.; Yudowski, G. A.; Rosenthal, J. J. C. Correction of mutations within the cystic fibrosis transmembrane conductance regulator by site-directed RNA editing. *Proceedings of the National Academy of Sciences of the United States of America* **2013**, *110* (45), 18285-18290. DOI: 10.1073/pnas.1306243110.
(6) Gabay, O.; Shoshan, Y.; Kopel, E.; Ben-Zvi, U.; Mann, T. D.; Bressler, N.; Cohen-Fultheim, R.; Schaffer, A. A.; Roth, S. H.; Tzur, Z.; et al. Landscape of adenosine-to-inosine RNA recoding across human tissues. *Nat Commun* **2022**, *13* (1), 1184. DOI: 10.1038/s41467-022-28841-4 From NLM Medline.
(7) Doherty, E. E.; Karki, A.; Wilcox, X. E.; Mendoza, H. G.; Manjunath, A.; Matos, V. J.; Fisher, A. J.; Beal, P. A. ADAR activation by inducing a syn conformation at guanosine adjacent to an editing site. *Nucleic Acids Res* **2022**, *50* (19), 10857-10868. DOI: 10.1093/nar/gkac897 From NLM Medline.
(8) Bennett, C. F.; Baker, B. F.; Pham, N.; Swayze, E.; Geary, R. S. Pharmacology of Antisense Drugs. *Annu Rev Pharmacol Toxicol* **2017**, *57*, 81-105. DOI: 10.1146/annurev-pharmtox-010716-104846 From NLM Medline.
(9) Doherty, E. E.; Wilcox, X. E.; van Sint Fiet, L.; Kemmel, C.; Turunen, J. J.; Klein, B.; Tantillo, D. J.; Fisher, A. J.; Beal, P. A. Rational Design of RNA Editing Guide Strands: Cytidine Analogs at the Orphan Position. *J Am Chem Soc* **2021**, *143* (18), 6865-6876. DOI: 10.1021/jacs.0c13319 From NLM Medline.
(10) Merkle, T.; Merz, S.; Reautschnig, P.; Blaha, A.; Li, Q.; Vogel, P.; Wettengel, J.; Li, J. B.; Stafforst, T. Precise RNA editing by recruiting endogenous ADARs with antisense oligonucleotides. *Nat Biotechnol* **2019**, *37* (2), 133-138. DOI: 10.1038/s41587-019-0013-6 From NLM Medline.
(11) Fritzell, K.; Xu, L. D.; Otrocka, M.; Andreasson, C.; Ohman, M. Sensitive ADAR editing reporter in cancer cells enables high-throughput screening of small molecule libraries. *Nucleic Acids Res* **2019**, *47* (4), e22. DOI: 10.1093/nar/gky1228 From NLM Medline.

(12) Wang, Y.; Beal, P. A. Probing RNA recognition by human ADAR2 using a high-throughput mutagenesis method. *Nucleic Acids Research* **2016**, *44* (20), 9872-9880. DOI: 10.1093/nar/gkw799.
(13) Lyst, M. J.; Ekiert, R.; Ebert, D. H.; Merusi, C.; Nowak, J.; Selfridge, J.; Guy, J.; Kastan, N. R.; Robinson, N. D.; de Lima Alves, F.; et al. Rett syndrome mutations abolish the interaction of *MECP2* with the NCoR/SMRT co-repressor. *Nat Neurosci* **2013**, *16* (7), 898-902. DOI: 10.1038/nn.3434 From NLM Medline.