**Supporting information**

# Selective Recognition of RNA Substrates by ADAR Deaminase Domains

Yuru Wang, SeHee Park and Peter A. Beal[1]*

[1]Department of Chemistry, University of California, One Shields Ave, Davis, CA 95616, USA

*Corresponding author, tel: (530) 752-4132. E-mail: pabeal@ucdavis.edu

Contents:

**Table S1**: Candidate editing sites from RNA-seq whose editing levels are <10%.

| Position | Edited region | Editing level | transcript |
|---|---|---|---|
| ChrVIII: 496495 | TCGAATTATTAGCCAACGGCG | 8% | AIM18 |
| ChrXV: 897233 | ATGATGCTTTAGCTGAAGATA | 8% | NOP58 |
| ChrIX: 133389 | CAAAGCTTTTAGAAGAACACG | 8% | QDR2 |
| ChrIV: 380146 | TGAAGCACTTAGGCCTAATGG | 8% | NAT1 |
| ChrXII: 349125 | ATTAAAGATTACCAAATACTT | 7% | MDN1 |
| ChrII: 276448 | ATATTACATTAGAAGTTTTAA | 7% | GAL10 |
| ChrIV: 331828 | ACTTGTCATTAGACGTTCAGT | 7% | BDF2 |
| ChrXVI: 858566 | AGTGGGAATTAGTATTTTCTG | 7% | KRE6 |
| ChrV: 547252 | TTTCACTTTTATGAGAGGTGC | 6% | PDA1 |
| ChrXII: 709636 | ATAGTGCGTTAGCTACCAAGT | 6% | CTS1 |
| ChrVII: 387750 | CCAGAAGTTTATCAACATGAA | 6% | PYC1 |
| ChrXIII: 262095 | CCCGGTCTTTAGAATTCTACC | 5% | GLO1 |
| ChrVIII: 124493 | ATGCTAGAAAAGATGAAGTTT | 5% | YHR009C |
| ChrXII: 708650 | AATGCTCAATATGCGGCTGGT | 5% | CTS1 |
| ChrII: 81786 | TGCAGTTTCTAGAATGCAATC | 5% | ILS1 |

The editing sites are colored in red.

**Table S2**: Primers used in RT-PCR to confirm *in vivo* deamination on the candidate editing sites.

| Editing sites | Primers (5' to 3') |
|---|---|
| HER1 | For: AGAAGAGGAAGGTGAATTACATAAATGG<br>Rev: CGGTAGGAATCTTTGGTAAGATTTTTG |
| PDR15 | For: AACTGCTTGGGACACTTGCC<br>Rev: CAACCTCTAACATCCATTCGGC |
| SPT6 | For: TCAACGTCACGCAGGTGC<br>Rev: TGTTCTCTTTGCACGGGCT |
| SCT1 | For: TTGAAATACCGAAGGAACTAGTCG<br>Rev: TATAGAATTCGTAACCTTTGACCATTCT |
| SCP160 | For: AATTGGTAACAAGGGCTCCAAC<br>Rev: CCAATCAGACTACCATGAACTTAACTG |
| RSM28 | For: GCATCTAAAAGGAGAAAGGCTCC<br>Rev: CCCAAATTTGGCGATCTGTAA |
| RPA190 | For: ACATCTAATGATGTTGCTGCTGTG<br>Rev: ACAGCCTTGGTCAAGAATTGAC |
| GSY1 | For: AAAAGGGTATTGGCGCTGAG<br>Rev: CATAATCCAAACCAAGGATAGGGT |
| BDF2 | For: GCAATGCCACCAAGAGTTTTACCCGC<br>Rev: GCAATCGGATCAACAGGTTGTAAAAAGGG |

**Table S3**: Primers used in saturation mutagenesis to generate hADAR1d 5' binding loop library.

| Locations | Primers (5' to 3') |
|---|---|
| G969X | For: GCACTGCTCCGTGTGGAGACNNSGCCCTCTTTGACAAGTCCTG<br>Rev: CAGGACTTGTCAAAGAGGGCSNNGTCTCCACACGGAGCAGTGC |
| A970X | For: CACTGCTCCGTGTGGAGATGGTNNSCTCTTTGACAAGTCCTGCAGC<br>Rev: GCTGCAGGACTTGTCAAAGAGSNNACCATCTCCACACGGAGCAGTG |
| L971X | For: CTCCGTGTGGAGATGGCGCTNNSTTTGACAAGTCCTGCAGCGAC<br>Rev: GTCGCTGCAGGACTTGTCAAASNNAGCGCCATCTCCACACGGAG |
| F972X | For: GTGTGGAGATGGCGCCCTANNSGACAAGTCCTGCAGCGACC<br>Rev: GGTCGCTGCAGGACTTGTCSNNTAGGGCGCCATCTCCACAC |
| D973X | For: GTGGAGATGGCGCCCTCTTCNNSAAGTCCTGCAGCGACCGTG<br>Rev: CACGGTCGCTGCAGGACTTSNNGAAGAGGGCGCCATCTCCAC |
| K974X | For: GGAGATGGCGCCCTCTTTGATNNSTCCTGCAGCGACCGTGCTA<br>Rev: TAGCACGGTCGCTGCAGGASNNATCAAAGAGGGCGCCATCTCC |
| S975X | For: GATGGCGCCCTCTTTGACAAANNSTGCAGCGACCGTGCTATGGAAAG<br>Rev: CTTTCCATAGCACGGTCGCTGCASNNTTTGTCAAAGAGGGCGCCATC |
| C976X | For: TGGCGCCCTCTTTGACAAGTCTNNSAGCGACCGTGCTATGGAAAGCA<br>Rev: TGCTTTCCATAGCACGGTCGCTSNNAGACTTGTCAAAGAGGGCGCCA |
| S977X | For: CGCCCTCTTTGACAAGTCCTGTNNSGACCGTGCTATGGAAAGCACAG<br>Rev: CTGTGCTTTCCATAGCACGGTCSNNACAGGACTTGTCAAAGAGGGCG |
| H988X | For: ATGGAAAGCACAGAATCCCGTNNSTACCCTGTCTTCGAGAATCCC<br>Rev: GGGATTCTCGAAGACAGGGTASNNACGGGATTCTGTGCTTTCCAT |
| P990X | For: AGCACAGAATCCCGCCACTATNNSGTCTTCGAGAATCCCAAACAAG<br>Rev: CTTGTTTGGGATTCTCGAAGACSNNATAGTGGCGGGATTCTGTGCT |
| F992X | For: CAGAATCCCGCCACTACCCTGTTNNSGAGAATCCCAAACAAGGAAAGC<br>Rev: GCTTTCCTTGTTTGGGATTCTCSNNAACAGGGTAGTGGCGGGATTCTG |
| E993X | For: GAATCCCGCCACTACCCTGTCTTTNNSAATCCCAAACAAGGAAAGCTCC<br>Rev: GGAGCTTTCCTTGTTTGGGATTSNNAAAGACAGGGTAGTGGCGGGATTC |
| N994X | For: CCGCCACTACCCTGTCTTCGAANNSCCCAAACAAGGAAAGCTCCG<br>Rev: CGGAGCTTTCCTTGTTTGGGSNNTTCGAAGACAGGGTAGTGGCGG |
| K996X | For: CTACCCTGTCTTCGAGAATCCANNSCAAGGAAAGCTCCGCACCAAGG<br>Rev: CCTTGGTGCGGAGCTTTCCTTGSNNTGGATTCTCGAAGACAGGGTAG |
| Q997X | For: CCCTGTCTTCGAGAATCCCAAGNNSGGAAAGCTCCGCACCAAGGTG<br>Rev: CACCTTGGTGCGGAGCTTTCCSNNCTTGGGATTCTCGAAGACAGGG |
| G998X | For: CCTGTCTTCGAGAATCCCAAACAGNNSAAGCTCCGCACCAAGGTGGAG<br>Rev: CTCCACCTTGGTGCGGAGCTTSNNCTGTTTGGGATTCTCGAAGACAGG |
| K999X | For: CTTCGAGAATCCCAAACAAGGTNNSCTCCGCACCAAGGTGGAGAACG<br>Rev: CGTTCTCCACCTTGGTGCGGAGSNNACCTTGTTTGGGATTCTCGAAG |

**Table S4**: Abundance and enrichment levels of R1-R5 gates and corresponding $F_{ave}$ values for amino acids at each position of variation in hADAR1d 5' binding loop library. **Column A-B**: Position under study, randomized codons and the corresponding amino acids introduced to the position. Column C-H: Abundance (the number of reads from Illumina sequencing) of each 20 amino acids and the stop codon in each input library and gate R1 through R5. **Column I-N**: Corrected abundance of each 20 amino acids and the stop codon based on the total reads from Illumina sequencing in each input library and gate R1 through R5. **Column O-S**: The enrichment levels in each gate R1 through R5 that was weighted against their abundance in the input library. **Column T**: Average fluorescence values as an indication of relative activities calculated using the enrichment levels of amino acids across the all five gates (R1-R5) and the median fluorescence values of populations in each gate.

*Table S4 can be found as a separate excel file.*

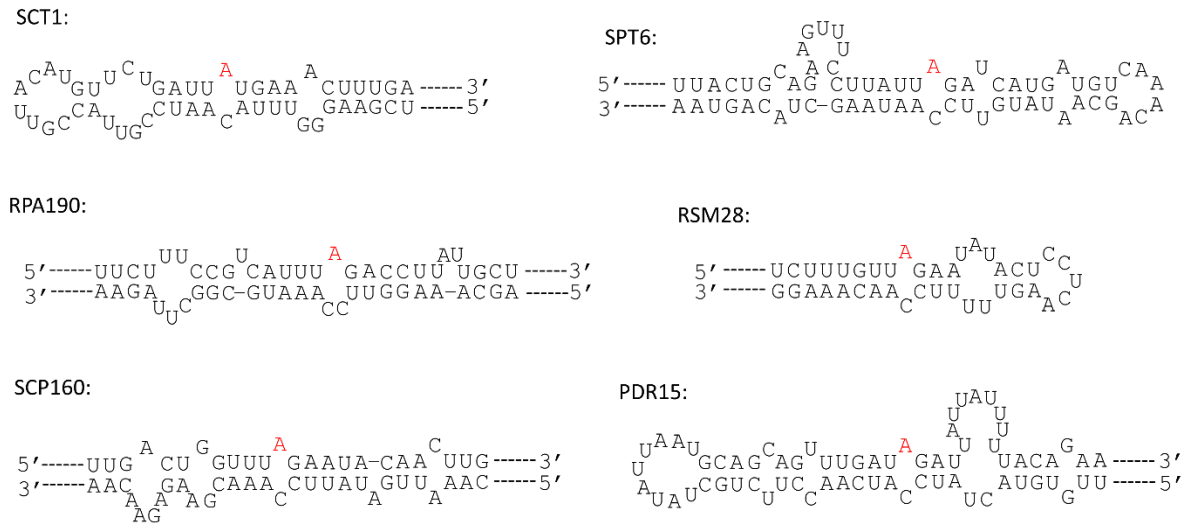


**Figure S1:** Secondary structures surrounding candidate editing sites predicted by Mfold (1). The editing sites are colored in red.

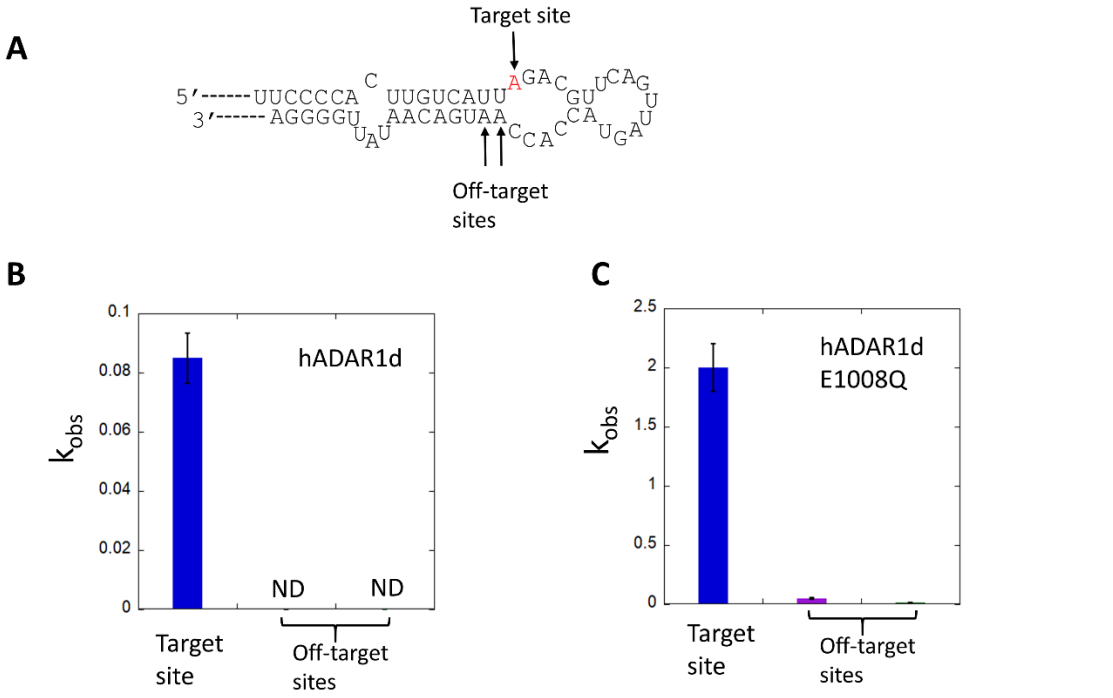


**Figure S2:** Editing of BDF2 target and off-target sites (A) by hADAR1d (B) and hADAR1d E1008Q mutant (C).

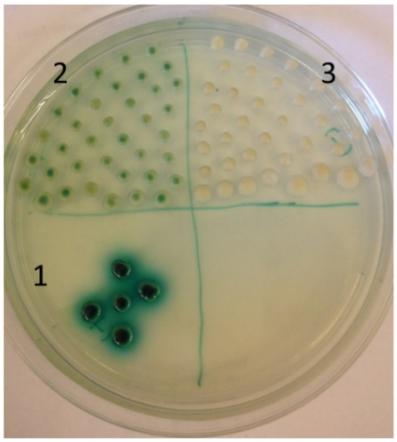


**Figure S3**: Colorimetric assay to test activity of the loop chimera. 1: Wild type hADAR2d; 2: loop chimera; 3: inactive hADAR1 mutant E912A. The picture was taken after nine days of incubation at 30 °C (2).

5'-
CGGATTAGAAGCCGCCGAGCGGGTGACAGCCCTCCGAAGGAAGACTCTCCTCCGTGCGTCCT
CGTCTTCACCGGTCGCGTTCCTGAAACGCAGATGTGCCTCGCGCCGCACTGCTCCGAACAAT
AAAGATTCTACAATACTAGCTTTTATGGTTATGAAGAGGAAAAATTGGCAGTAACCTGGCCC
CACAAACCTTCAAATGAACGAATCAAATTAACAACCATAGGATGATAATGCGATTAGTTTTT
TAGCCTTATTTCTGGGGTAATTAATCAGCGAAGCGATGATTTTTGATCTATTAACAGATATAT
AAATGCAAAAACTGCATAACCACTTTAACTAATACTTTCAACATTTTCGGTTTGTATTACTTC
TTATTCAAATGTAATAAAAGTATCAACAAAAAATTGTTAATATACCTCTATACTTTAACGTC
AAGGAGAAAAAACCCCGGATCGGACTACTAGCAGCTGTAATACGACTCACTATAGGGAATA
TTAAGCTTATGATCCGCCAATTCTTCGAGAATTAGCGGGTATCTAAAGGTGAAGAATTATTC
ACTGGTGTTGTCCCAATTTTGGTTGAATTAGATGGTGATGTTAATGGTCACAAATTTTCTGTC
TCCGGTGAAGGTGAAGGTGATGCTACTTACGGTAAATTGACCTTAAAATTTATTTGTACTAC
TGGTAAATTGCCAGTTCCATGGCCAACCTTAGTCACTACTTTAACTTATGGTGTTCAATGTTT
TTCTAGATACCCAGATCATATGAAACAACATGACTTTTTCAAGTCTGCCATGCCAGAAGGTT
ATGTTCAAGAAAGAACTATTTTTTTCAAAGATGACGGTAACTACAAGACCAGAGCTGAAGTC
AAGTTTGAAGGTGATACCTTAGTTAATAGAATCGAATTAAAAGGTATTGATTTTAAAGAAGA
TGGTAACATTTTAGGTCACAAATTGGAATACAACTATAACTCTCACAATGTTTACATCATGG
CTGACAAACAAAAGAATGGTATCAAAGTTAACTTCAAAATTAGACACAACATTGAAGATGG
TTCTGTTCAATTAGCTGACCATTATCAACAAAATACTCCAATTGGTGATGGTCCAGTCTTGTT
ACCAGACAACCATTACTTATCCACTCAATCTGCCTTATCCAAAGATCCAAACGAAAAGAGAG
ACCACATGGTCTTGTTAGAATTTGTTACTGCTGCTGGTATTACCCATGGTATGGATGAATTGT
ACAAATGA-3'

**Figure S4**: Gene in the HER1 yeGFP reporter plasmid from the GAL1 promoter to the end of the gene. The start codon and stop codon are underlined.

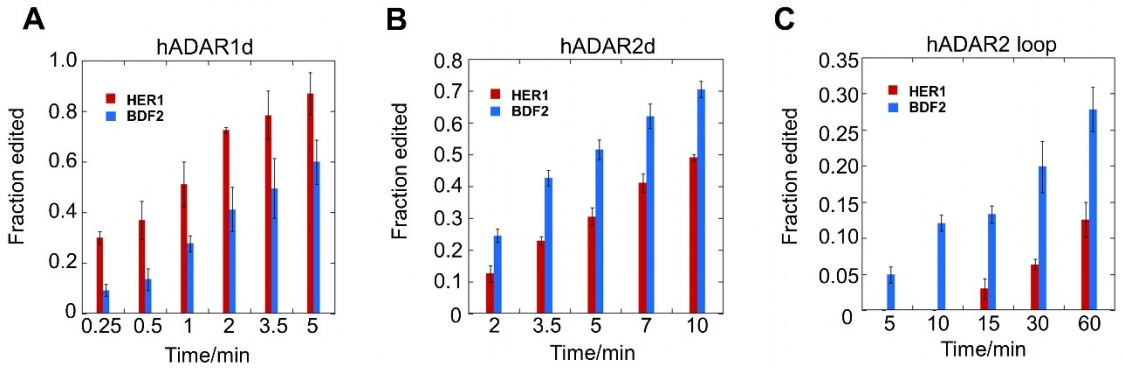


**Figure S5**: A) hADAR1d E1008Q (20 nM) acting on the HER1 and BDF2 substrates (5 nM). B) hADAR2d E488Q (20 nM) acting on the HER1 and BDF2 substrates (5 nM). D) Loop chimera protein (20 nM) acting on the HER1 and BDF2 substrates (5 nM). Error bar indicates SD, n≥3.


**Figure S6**: Western blot analysis of ADAR protein expression in yeast. A) Western blot on lysate of yeast cells expressing hADAR1d (lane 1) and hADAR2d (lane 2). B) Western blot on lysate of yeast cells expressing hADAR1d before (lane 1) and after (lane 2) hADAR1d codon optimization.

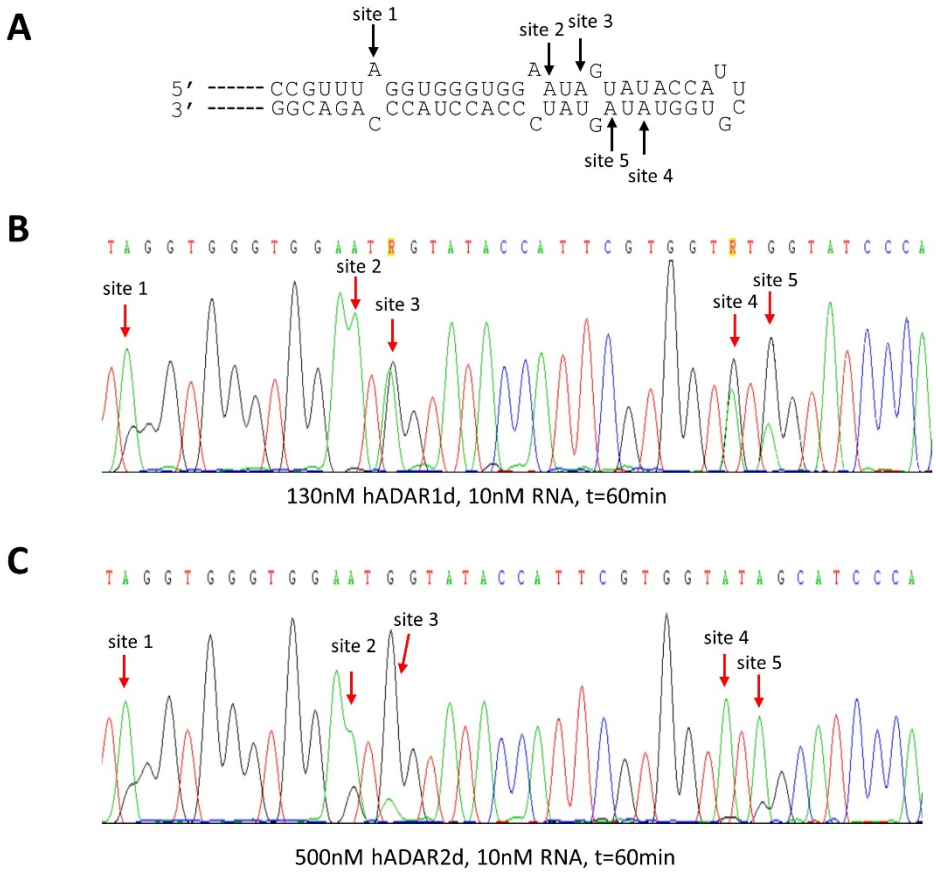


**Figure S7:** A) The editing sites in GRIA2-mimicking RNA. B) Editing of 130 nM hADAR1d on 10 nM GRIA2-derived RNA (WT), t=60 min. C) Editing of 500 nM hADAR2d on 10 nM GRIA2-derived RNA (WT), t=60 min. Editing on this substrate was determined using RT-PCR followed by Sanger sequencing.

# Supplementary methods

### Expressing hADAR1d in yeast and isolation of total RNA for RNA-seq.

Genes encoding hADARds are present in yeast expression plasmid YEpTOP2PGAL1. *S. cerevisiae* INVSc1 strain containing wild type hADAR1d or inactive hADAR1d mutant E912A were generated using a lithium acetate protocol and grown on the complete media (CM) – ura + 2% glucose plates. Single colonies were picked and used to inoculate 5 ml CM – ura + 2% glucose media and incubated for 24 h at 30 °C with shaking at 275 rpm. Cells were pelleted from 1 mL of each culture, washed twice with $ddH_2O$ and used to inoculate 25 mL CM – ura + 2% raffinose + 3% galactose media for induction. The galactose cultures were grown for 48 h at 30 °C with shaking at 275 rpm. To avoid a stress response in yeast, the cultures were frequently diluted to maintain an $OD_{600}$ below 2.0. Total RNA was extracted from INVSc1 strain

expressing either wild type hADAR1d or inactive mutant E912A using RiboPure-Yeast kit (Ambion).

### *In vitro* RNA transcription and purification

A cDNA sequence of 110 nt upstream and 107 nt downstream the HER1 site, and a cDNA sequence of 92 nt upstream and 98 nt downstream BDF2 site were amplified from the total RNA of *S. cerevisiae* INVSc1 strain using RT-PCR. The resulting PCR products were inserted into pYES/CT vector by standard cloning. DNA sequence of 90 nt surrounding Gabra3 site and its reverse complementary strand were ordered from IDT, annealed to form a duplex and inserted into pYES/CT vector. The plasmids were linearized by digestion at a restriction site downstream of the inserted sequences, after which *in vitro* transcription was carried out using MEGAscript kit (Life Technology). Transcribed RNAs were purified using a denaturing polyacrylamide gel as described previously (3). Refolding was carried out by diluting the RNA to the desired concentration in annealing buffer (100 mM NaCl, 10 mM Tris-HCl, pH =8, 1 mM EDTA) and heating at 95 °C for 5 min and then slowly cooled to room temperature. Solutions of refolded RNA were stored at -20 °C.

### Splint ligation

The splint ligation was performed by hybridizing 50 pmol of the unlabeled 24 mer RNA, 40 pmol of the radiolabeled 18 mer RNA/DNA chimera and 40 pmol of the 35 mer DNA strand into 1.6X ligation buffer in the presence of 0.064 U/µl RNAsin. The sample was heated at 80 °C for 5 min and slowly cooled to room temperature. RNAsin, ATP and T4 DNA ligase were then added to the sample to final concentrations of 0.16 U/µl RNAsin, 1 mM ATP, 1X ligation buffer and 20 U/µl T4 DNA ligase. The ligation was allowed to proceed at 30 °C for overnight. The ligation reaction mixture was phenol-chloroform extracted and ethanol precipitated and then subjected to DNAse digestion with 0.1U/µl DNAse at 37 °C for 2 h. The digested product was then purified with a 12% denaturing polyacrylamide gel. The ligation product was isolated from the gel by overnight crush and soak in solution containing 0.5 M $NH_4OAc$, 0.1% SDS, and 0.1 mM EDTA, followed by phenol-chloroform extraction and ethanol precipitation. The recovered RNA was refolded as described above and stored in -20°C.

### Nuclease P1 digestion followed by thin layer chromatography (TLC)

RNAs purified after deamination were re-suspended in 1X TE buffer and then added with nuclease P1 to a final concentration of 10 U/µl. The samples were incubated at 50 °C for 30min. After digestion, the samples were subjected to phenol-chloroform extraction and ether extraction, after which the samples were dried overnight. The dried products were re-suspended in 2 µl $H_2O$ and resolved using TLC. The gel was visualized after approximately 3 days of exposure to storage phosphor imaging plates and band intensities were quantified by volume integration using ImageQuant software. The editing data were also fit to the equation $[P]t = \alpha[1 - e^{(-kobs*t)}]$.

### Colorimetric assay to test RNA editing activity

The procedure of performing colorimetric assay is described previously with the following modification (2). Yeast colonies growing on CM – ura – trp + 2% glucose plates were transferred by manual dotting to plates containing CM – ura – trp + 2% raffinose + 3% galactose.

**Construction of HER1 yeGFP reporter**

A HER1 yeGFP reporter was generated using the sequence composing the minimum local secondary structure surrounding the HER1 editing site. To maintain the open reading frame, a 30 nt sequence surrounding the HER1 editing site was used. A DNA oligonucleotide comprising from the 5' to 3' a HindIII restriction site, the 30 nt HER1 sequence and a sequence mapping to N-terminus sequence of yeast enhanced green fluorescence protein (yeGFP), and another oligonucleotide comprising from 5' to 3' a XhoI restriction site and a sequence mapping to the yeGFP C-terminus sequence were purchased. These oligonucleotides were used as primers to amplify the yeGFP sequence from the BDF2 yeGFP plasmid described previously (4) and resulting amplicon was inserted into pYES3/CT (Invitrogen) by standard cloning procedures. The resulting plasmid has a *GAL1* promoter, a *TRP1* auxotrophic marker for selection of transformed yeast, and an ampicillin resistance marker for selection of transformed *E. coli*. Full sequence of the gene in the HER1 yeGFP plasmid is found in **Figure S4**.

**Western blot on yeast cell lysate**

The same number ($10^8$) of cells expressing ADAR proteins were harvested after induction for protein expression. For testing expression levels of wild type ADAR1d and ADAR2d in yeast, cells were harvested after 48 h of induction. For testing expression levels of ADAR1d alanine mutants (including the parent protein) in yeast, cells were harvested after 10 h of induction. Cells were lysed with glass beads and the cell lysates were used directly for western blot analysis. The ADAR proteins have a 10X His-tag at their N-terminus and were probed with a mouse His-probe antibody (Santa Cruz biotechnology).

# Reference


1. Zuker, M. (2003) Mfold web server for nucleic acid folding and hybridization prediction. *Nucleic Acids Res.*, **31**, 3406-3415.
2. Pokharel, S. and Beal, P.A. (2006) High-throughput screening for functional adenosine to inosine RNA editing systems. *ACS Chem. Biol*, **1**, 761-765.
3. Wang, Y., Havel, J. and Beal, P.A. (2015) A phenotypic screen for functional mutants of human adenosine deaminase acting on RNA 1. *ACS Chem. Biol.*, **10**, 2512-2519.
4. Wang, Y. and Beal, P.A. (2016) Probing RNA recognition by human ADAR2 using a high-throughput mutagenesis method. *Nucleic Acids Res.*, **44**, 9872-9880.