Dissecting the basis for differential substrate specificity of ADAR1 and ADAR2

Marlon S. Zambrano-Mila[1], Monika Witzenberger[1], Anna Uzonyi[1], Ronit Nir[1], Shay Ben-Aroya[2], Erez Y. Levanon[2], Schraga Schwartz[1]

[1] Department of Molecular Genetics, Weizmann Institute of Science, Rehovot 7630031, Israel.
[2] Faculty of Life Sciences, Bar Ilan University, 5290002, Ramat Gan, Israel.

*- Corresponding author:

Schraga Schwartz - schwartz@weizmann.ac.il

Abstract

Millions of adenosines are deaminated throughout the transcriptome by ADAR1 and ADAR2, modulating double-stranded RNA (dsRNA) immunogenicity and recoding mRNA. The high variability in the susceptibility of different adenosines to editing begs the question of what are the determinants of substrate specificity. Here, we systematically monitor how secondary structure modulates ADAR2 vs ADAR1 substrate selectivity, on the basis of systematic probing of thousands of synthetic sequences transfected into ADAR1-deleted cell lines exogenously expressing either ADAR2 or ADAR1. In both cases, structural disruptions gave rise to symmetric, strand-specific induced editing at a fixed offset, but of varying length: -26 nt for ADAR2, and -35 nt for ADAR1. We dissect the basis for the differences in offset between ADAR1 and ADAR2 via diverse mutants, domain-swaps, and ADAR evolutionary homologs, and reveal that it is encoded by the differential RNA binding domain architecture. We demonstrate that this offset-enhanced editing can allow an improved design of ADAR2-recruiting therapeutics, with proof-of-concept experiments suggestive of increased on-target and potentially decreased off-target editing. Our findings provide novel insight into the determinants guiding ADAR2 substrate selectivity and into the roles of the RNA binding domains of ADAR1 and ADAR2 in mediating differential targeting, and should facilitate the design of improved ADAR-recruiting therapeutics.

Introduction

Millions of adenosines are deaminated into inosines transcriptome-wide [1,2], catalyzed by two deaminating enzymes, ADAR1 (ADAR) and ADAR2 (ADARB1). Inosine is perceived as guanosine by the internal cellular machinery, and hence editing can result in protein recoding [3–7], alternative splicing [8–10], and alterations in targeting and maturation of microRNA [11,12]. In parallel, editing can also alter the RNA secondary structure, and in doing so modulate the immunogenicity of self and viral RNAs within cells [13–17]. In accordance with the wide distribution of edited sites, abnormal dysregulation of A-to-I deamination has been associated with a broad spectrum of human diseases [18], and targeting of ADAR enzymes is an emerging therapeutic strategy in cancer [19].

Different adenosines throughout the transcriptome are edited at dramatically different efficiencies (or not at all), begging the question of what governs enzymatic selectivity towards specific targets. Understanding the rules guiding these two enzymes to their diverse targets is of intense interest not only from biological and pathological perspectives, but also from therapeutic ones. In recent years, unraveling the rules dictating deamination via the ADAR enzymes has accrued substantial interest in the context of ongoing efforts to achieve targeted mRNA editing. Targeted editing is emerging as a therapeutic modality that may potentially offer a safer alternative to correct single-nucleotide mutations [20] in comparison to CRISPR-mediated DNA editing. Diverse approaches have been implemented in recent years to recruit ADAR enzymes towards specific substrates [21–25]. Although successful, these attempts often resulted only in partial efficiencies and in some cases also with considerable off-target effects [26]. Improving our understanding of the rules guiding inosine formation and of the factors determining enzyme specificity will pave the path toward the development of both more optimal editors and improved guides.

Studies exploring the targeting efficiencies of ADAR1 and ADAR2 have revealed several general principles. First, the specificity of these two enzymes is only partially overlapping [27,28], suggesting differences in the selectivity of these two enzymes. Second, RNA secondary structure plays a critical role. ADAR1 targets are nearly inevitably within long double-stranded RNAs [29], and hence highly enriched in repetitive elements such as Alu and long interspersed elements [30–32]. ADAR2 targets tend to be in duplex regions interrupted by mismatches or loops [33–36], and *In-vitro* work has shown that distal bulges can, at times, impact editing efficiency [34]. Yet, the structural rules governing editing - which would be of critical importance for predictive models - are not understood. In addition, for both ADAR1 and ADAR2, A-C mismatch harboring targets are particularly prone to undergo editing [37]. Finally, the sequence also plays a role in target selectivity. *In vitro* editing assays with artificial RNA duplexes revealed that ADAR1 and ADAR2 prefer to edit adenosines depleted of G's at the position preceding the target and show some bias for a G downstream of the target edited site [35,38,39].

The factors underlying the differences in specificity between ADAR1 and ADAR2 are understood only to a limited extent. These two paralog proteins, which likely evolved via a gene duplication event roughly 700 million years ago [40], differ in their protein domain architecture. The catalytic domain, present on both ADARs, was shown to play a role in the definition of selectivity [37,41]. In addition to the catalytic domain, human ADAR1 contains either one or two Zα domains (dependent on the isoform) and three RNA binding domains (RBDs) whereas human ADAR2 contains two RBDs but no Zα domains. The RBDs participate in dsRNA substrate recognition and RNA binding [42], and were suggested to partially mediate ADARs selectivity via both sequence-specific and non-

3

specific mechanisms [34,43]. The Z domains, binding left-handed nucleic acids, have been implicated in allowing co-transcriptional binding of ADAR1 to nascent RNAs [44,45] and more recently in preventing Z-RNA dependent activation of pathogenic interferon by Z-DNA binding protein 1 (ZBP1) [46–48]. Whether they have a role in defining substrate specificity is unclear.

To systematically dissect how substrate selectivity by ADAR1 is governed by secondary structure, we previously screened ADAR1-mediated editing across thousands of sequence variants, which had been designed to systematically perturb the secondary structure along two highly double-stranded backbones. We discovered that introduction of structural disruptions within an otherwise perfect double-stranded RNA structure gives rise to robust and predictable ADAR1-mediated editing at a fixed offset of 35 bp upstream from the disruption [49]. Whether structural disruptions of ADAR2 targets lead to editing at a fixed offset, and what the mechanistic basis for this offset is, remains unknown.

Here, we systematically monitor how secondary structure modulates ADAR2 substrate selectivity, on the basis of systematic probing of thousands of synthetic sequences transfected into ADAR1-deleted cell lines exogenously expressing ADAR2. We find that similarly to ADAR1, structural disruptions give rise to symmetric, strand-specific induced editing at a fixed offset. However, in contrast to ADAR1 acting at a -35 bp offset, in the case of ADAR2, structural disruptions give rise to induced editing at an offset of -26 bp. We dissect the basis for the differences in offset between ADAR1 and ADAR2 via diverse mutants, domain-swaps and ADAR evolutionary homologs. We uncover that the difference in the offset is encoded by the differential RNA binding domain architecture of the two ADARs, yet that it is not determined by the number of RBDs. We demonstrate that this new understanding of ADAR2 specificity can allow an improved design of ADAR2-recruiting therapeutics, yielding increased on-target editing, with some evidence also for reduced off-target editing. Our findings provide novel insight into the features determining ADAR2 substrate selectivity and into the roles of the RNA binding domains of ADAR1 and ADAR2 in mediating differential targeting, and should facilitate the design of improved ADAR2-recruiting therapeutics.

Results

Screening of ADAR2 substrates

We sought to systematically compare the targeting specificity of ADAR2 to its ADAR1 counterpart. Toward this goal, we employed a pool of thousands of sequence variants that we had previously designed to probe the specificity of ADAR1, described in [49]. In brief, these sequence variants are based on two distinct backbones folding into a perfect hairpin structure: the endogenous mouse B2 element, serving as a more 'natural' editing target, and a sequence complementary to the 3' UTR of the fluorescent reporter mNeonGreen (mNG) transcript, serving as a completely synthetic target. In both cases, this hairpin consists of a 146-nt long stem and a 46-nt long loop (**Figure 1A**). For each of these two backbones, we previously designed and synthesized roughly two thousand sequence variants systematically perturbing the hairpin structure via random structural disruptions, systematic incorporation of single, double, or random mismatches, the introduction of pyrimidine-rich bulges, and systematic shortening or elongation of the stem (**Figure 1B**). These perturbations were all designed to take place in the 'lower' arm of the stem structure, whereas the 'upper' arm remained constant. We transfected each oligo library

into ADAR1-knockout HEK293T cells (in which ADAR2 is not expressed)[50], alongside a plasmid expressing either ADAR2 or ADAR1 or neither of the two ('No-ADAR'), as a negative control. Subsequently, RNA was extracted, and the constant upper arm of each construct was reverse transcribed, PCR-amplified, and sequenced **(Figure 1C)**.

All B2 and mNG constructs were detected across all treatments with a mean coverage of ~4000 reads per barcode per sample across all conditions (ADAR1, ADAR2, and No-ADAR). No editing was observed in ADAR1 KO cells transfected with the No-ADAR vector, corroborating that all deamination activity is triggered by the two exogenously overexpressed ADAR enzymes **(Figure 1D)**. In ADAR-expressing cells, editing percentages between technical duplicates were highly reproducible (r > 0.99, *P*< 2.2e-16 for all treatments) **(Figure 1E & Figure S1A)**. In addition, editing measurements were independent of barcode identity, as was assessed by comparing editing levels at a subset of identical sequences with distinct barcodes **(Figure 1F & Figure S1B).**

The editing patterns in the B2 and mNG constructs following ADAR1 overexpression were well correlated with ones observed in WT HEK293T cells in which ADAR1 is expressed at endogenous levels **(Figure 1D)**, suggesting that ADAR overexpression is a valid approach for interrogating the rules defining the substrate specificities of ADAR enzymes. The editing patterns following overexpression of ADAR1 and ADAR2 were substantially less correlated, in line with previous reports indicating their only partially overlapping target specificity [27,28] **(Figure 1D)**. We also note that ADAR2 overexpression gave rise to higher levels of editing in comparison to ADAR1, in line with previous reports [51]. In mNG constructs, a median of ~22 out of 44 adenosines per molecule was edited in ADAR2-overexpressing cells, in comparison to ~18 in ADAR1-overexpressing counterparts **(Figure 1G).** This trend was even more pronounced, with ~3 and ~12 out of 41 edited sites per molecule in B2 constructs in ADAR1- and ADAR2-overexpressing cells, respectively.

As an additional quality control, we assessed editing levels across a series of constructs in which the double-stranded stem was randomly disrupted to varying levels. Consistent with our expectations, we found that editing by both ADAR1 and ADAR2 was continuously disrupted with progressive disruption of the secondary structure **(Figure 1H & Figure S1C)**. ADAR2 was slightly more resilient to the introduction of structural disruptions, consistent with previous studies showing that ADAR2 can efficiently edit shorter double-stranded substrates than ADAR1 [38,52]. Collectively, these analyses establish that the two synthetic constructs and their perturbed counterparts are edited by both ADAR1 and ADAR2, yet these two enzymes are associated with both varying levels and different patterns of editing.
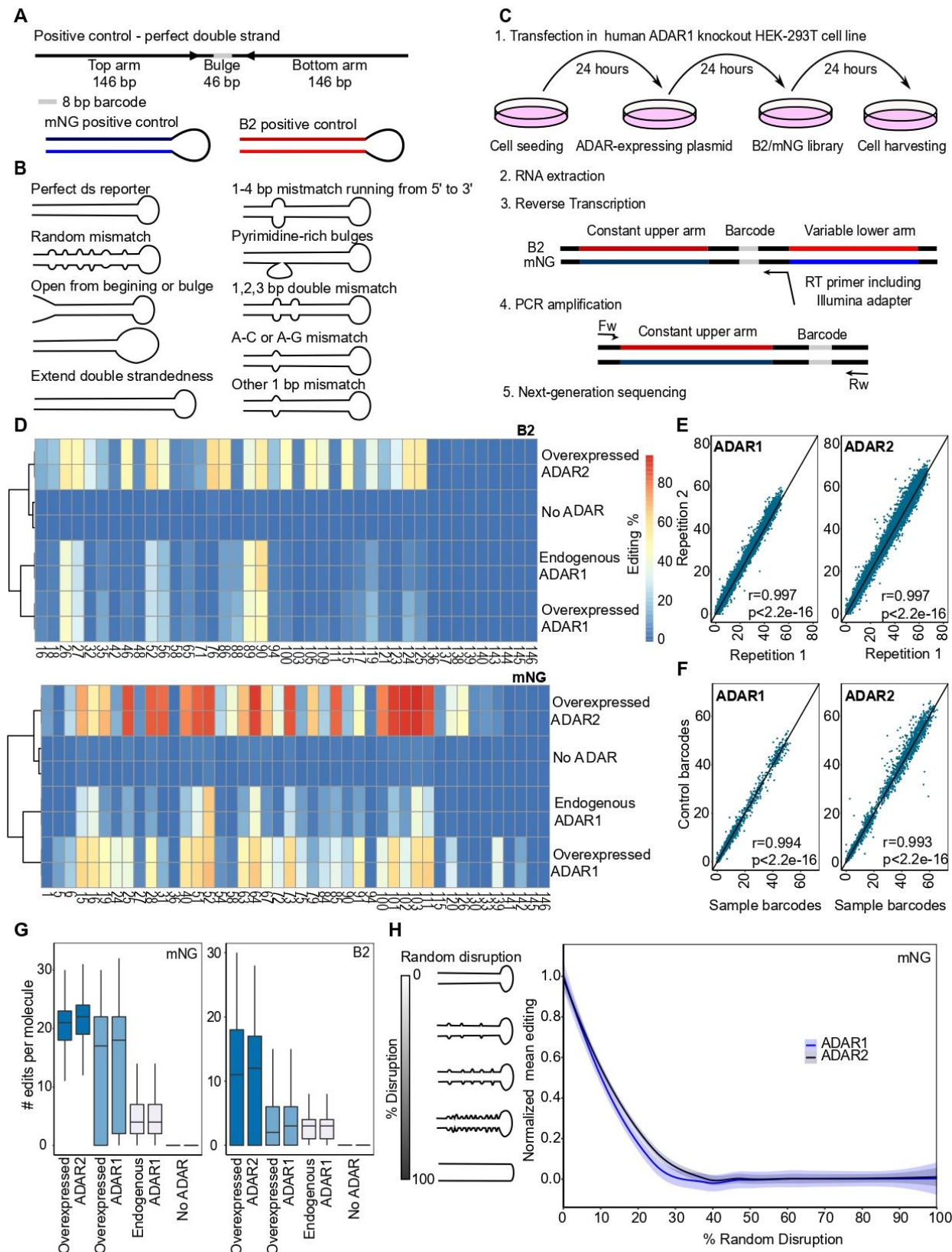
Figure 1. **Systematic screening of ADAR2 synthetic substrates in ADAR1 knockout HEK293T cell lines**. A) Design of double-stranded reporters. B2 and mNG are based on a mouse non-coding B2 element and the mNeonGreen gene, respectively. B) Repertoire of sequence series in B2 and mNG libraries. C) Experimental pipeline: Expression of the synthetic libraries in ADAR1-knockout HEK293T cells, which exogenously overexpress ADAR1 or ADAR2, and subsequent library preparation. RNA was extracted, and the constant arm and barcode of each construct were reverse transcribed. Subsequently, PCR amplification and sequencing using Novaseq 6000 platform with a 300 bp kit were performed. D) A-to-I editing levels in the B2 (upper diagram) and mNG (lower diagram) perfect double-stranded constructs in No-ADAR, ADAR1-overexpressing or ADAR2-overexpressing ADAR1-KO HEK293T cells, and wild-type HEK293T cells. E) Correlation of A-to-I levels among technical duplicates in cells overexpressing either ADAR1 or ADAR2. Each dot depicts the editing percentage of each adenosine in each construct of the B2 oligo library. The Pearson correlation coefficient and p-values are shown. F) Correlation of editing levels in B2 constructs that differ in the barcode sequences. The Pearson correlation coefficient and p-values are shown. G) Boxplots representing the distribution of numbers of editing events in the single mNG/B2 perfect double-stranded molecules in either No-ADAR expressing cells, ADAR-overexpressing cells, or wild-type HEK293T cells. H) Min-Max normalized mean editing percentage in the subset of mNG constructs containing random disruptions of double-strandedness in 5% increments.

ADAR2-mediated editing is induced 26 nt upstream of structural disruptions.

We next sought to assess whether structural disruptions within dsRNAs induce ADAR2-mediated editing at a fixed offset, given our previous discoveries of a -35 bp offset for ADAR1 [49]. To explore this, we analyzed the series of constructs into which we had systematically introduced secondary-structure disrupting sequences - either in the form of mismatches or of bulges - throughout the stem. Indeed, in both ADAR1- and ADAR2-overexpressing cells, increased editing levels were observed at a fixed offset (**Figure 2A-C & Figure S2A**). In the case of ADAR1-overexpressing cells, we recapitulated our previous observations of increased editing levels 35 bp upstream and 30 bp downstream of structural disruptions (**Figures 2B-C**) [49]. In contrast, in ADAR2-overexpressing cells, structural disruptions led to increased editing levels 26 nt upstream from the structural disruption. Though the magnitude of the increase in editing levels at position -26 following ADAR2 overexpression was lower than the increase at position -35 following ADAR1 overexpression (~1.3-1.5 mean fold at position -26 in comparison to ~3.3-6 mean fold at position -35), the phenomenon was reproducibly observed across the two different constructs as well as using different forms of structural disruption including mismatches of varying lengths (**Fig. 2B & Figure S2B-G**) and pyrimidine-rich bulges (**Figure 2C & Figure S3A-C**). The increase in editing levels at position -35 and -26 in ADAR1- and ADAR2-overexpressing cells, respectively, was dependent on the size of the mismatch, with the highest median editing increase observed in constructs carrying 3 nucleotide mismatches (**Figure 2D**). In parallel, the introduction of mismatches also led to a reproducible negative signal (indicative of adenosines resistant to editing) that was distributed in a complex - yet highly reproducible - manner with respect to the structural mismatches. The negative signal extended between positions -26 and +29. The signal was at its minimum at position 0 and +1 for ADAR1 and ADAR2, respectively, consistent with previous reports [53], with two local maximums at positions -9 and positions +6/+7 in both ADAR1 and ADAR2, and an additional ADAR2-specific local maximum at position +15 (**Figure 2B**).

In the case of ADAR1, we had previously found that structural disruptions led to a *symmetric* induction of editing, resulting in induced editing 35 bp upstream of the structural disruption on the 'top' arm of the dsRNA, and in parallel also resulting in induced editing 35 bp upstream of the structural disruption on the 'bottom' arm. Given that all results obtained thus far had only been on the basis of sequencing of the 'top' (and invariable) arm, we next amplified and sequenced also the 'bottom' variable arm of each B2 construct from ADAR1-KO HEK293T cells overexpressing ADAR2 (**Figure 2E**). A prominent peak 26 bp upstream from the structural disruption was observed on the opposite strand (**Figure 2F**), indicating that the induction of editing by human

ADAR2 is symmetric and orientation-dependent at a fixed interval as was the case for ADAR1, but in this case 26 nt upstream from structural disruptions.
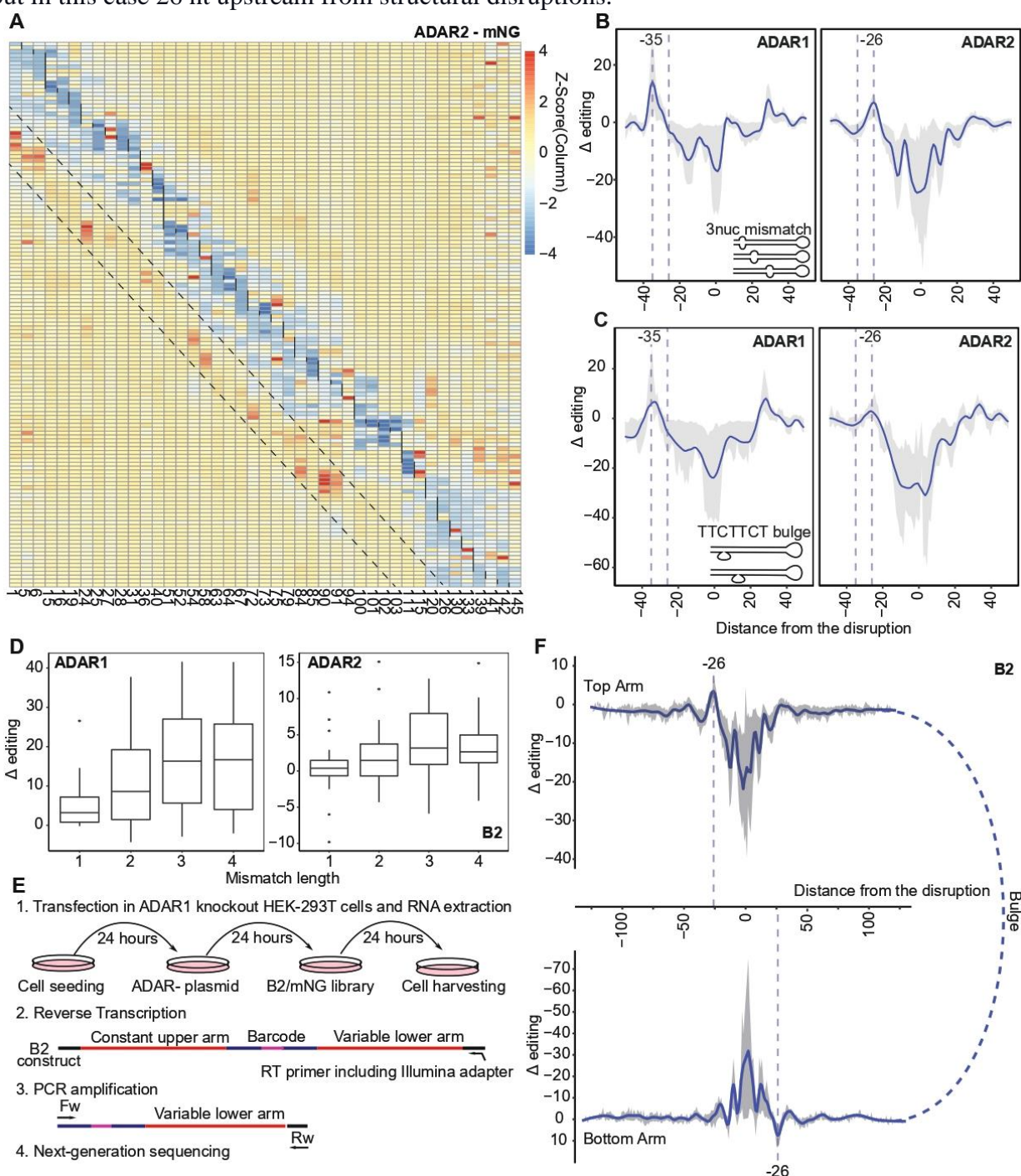
Figure 2. ADAR2-mediated editing is induced at a constant interval of 26 bp upstream from structural disruptions. A) Heatmap of a 3-nucleotide mismatch running from 5' to 3' throughout the double-stranded RNA. Each row represents a construct structurally disrupted at a specific position while each column represents an adenosine position. Delta (Δ) editing is color-coded after scaling by columns using Z-score transformation (mNG series). Black vertical lines indicate the location of the 3 nt mismatch and the parallel dashed lines highlight the ADAR2-mediated editing increase at a fixed distance upstream from the 3-nucleotide mismatch. B) ADAR1- and ADAR2-mediated editing offsets based on the subset of 3-nucleotide mismatch running throughout the mNG and B2 sequences. Mismatches differentially located in each construct get centered at 0 on the x-axis. The Δ editing level on

8

the y-axis represents the change of the editing level of an adenosine, normalized to the perfect double-stranded construct. Fitted curves depict the Loess fit of Δ editing with a span of 0.05 and the shaded region spans the 25th percentile and 75th percentile values of Δ editing per distance. Only adenosine positions, which have greater than 1% in editing on the perfect double-stranded construct, were included in the analysis. Vertical dashed lines are placed at -26 and -35. C) Subset of TTCTTCT bulges running throughout the mNG and B2 sequences. Loess fit of Δ editing with a span of 0.11. Data is exemplified as Fig. 2B. D) The mismatch size affects ADAR1- and ADAR2-mediated editing on adenosines located at -35 and -26 downstream from the mismatch, respectively. E) Library preparation: RNA was extracted, the B2 variable lower arm and barcode were reverse transcribed, and subsequently PCR amplification and sequencing using Novaseq 6000 platform with a 300 bp kit were performed. F) Depiction of the subset of 3-nucleotide mismatch running throughout the stem (B2 series) in ADAR1-knockout HEK293T cells overexpressing ADAR2. Constant and variable are illustrated under each other, and nucleotide locations are aligned. Data is shown as Figure 2B.

Differences in editing offsets among ADARs are mediated by double-stranded RNA binding domains

We next sought to understand why structural disruptions led to editing at an offset of 35 nt in the case of ADAR1, but of 26 nt in the case of ADAR2. To explore whether the offset was dictated by the catalytic domain of the two ADAR enzymes or by the RBDs, we designed two ADAR variants, by swapping the RBD domains among the ADARs: (1) An 'ADAR2-RBDs_ADAR1-deaminase' variant, harboring the catalytic domain of ADAR1 fused to the two RBDs originating from ADAR2, and (2) An 'ADAR1-RBDs_ADAR2-deaminase' variant, harboring the catalytic domain of ADAR2 fused to the three RBDs originating from ADAR1. We next used the above described human ADAR1-depleted system, into which we transfected B2 and mNG oligo libraries along with plasmids overexpressing these two ADAR variants. The two hybrids gave rise to deamination activity on both B2 and mNG positive control constructs (**Figure S4A, S4B**), albeit at substantially reduced levels in comparison to the WT counterparts (**Figure 3A**). Remarkably, we found that the offset size segregated with RBDs: 'ADAR1-RBDs_ADAR2-deaminase' showed induced editing levels at position -35, recapitulating the patterns observed in WT ADAR1 expressing cells. In parallel, 'ADAR2-RBDs_ADAR1-deaminase' exhibited induced activity at roughly -30 nt, as had similarly been observed for ADAR2 (**Figure 3B, 3C; Figure S5A-B; Figure S5E-F**). Thus, these findings suggest that the size of the offset is encoded by the differential RBD architecture.

How do the different RBDs give rise to differential offsets? We hypothesized that the RBDs might serve as molecular rulers and that the size of the offset might scale roughly linearly with the number of RBDs. Under this scenario the offset in ADAR2 with respect to ADAR1 might reflect the loss of one RBD in ADAR2, harboring 2 RBDs, in comparison to ADAR1, harboring 3 RBDs. To test this hypothesis, we designed two ADAR2 variants harboring only a single RBD by either maintaining only the first or only the second RBD, with the anticipation that these might lead to an offset potentially even smaller than -26 nt. Both mutants were active within cells, albeit at drastically different levels (**Figure 3A, S4A, S4B**), with the mutant harboring only the first RBD exhibiting very low levels of activity, in contrast to the RBD2-harboring mutant that gave rise to higher levels of editing than WT ADAR2, consistent with [54]. Nonetheless, in both cases, the size of the offset remained fixed at roughly -26, similar to WT ADAR2 (**Figure 3D; Figure S5C; S5G**). These findings thus suggest that the size of the offset is not determined by the *number* of RBDs.

The above results left open the possibility that the effect of the number of RBDs might be threshold-dependent. Under such a scenario, one or two RBDs might invariably give rise to an

9

offset of -26 whereas the addition of a third RBD might give rise to an increased increment of -35. To test this possibility, we selected two additional ADAR homologs from *Suricata suricatta* and *Octopus vulgaris,* harboring one and two RBDs, respectively, to assess whether these invariably gave rise to editing at an offset of -26 nt. The two ADAR enzymes elicited deamination activity on B2 or mNG positive control constructs **(Figure S4A, S4B)**, albeit at varying levels and with differences in substrate selectivity **(Figure 3A)**. Interestingly, the two ADAR homologs gave rise to different offsets: suricata ADAR gave rise to an offset of -35 similar to human ADAR1, whereas octopus displayed a peak at position -28, similar to human ADAR2 **(Figure 3E; Figure S5D, S5H)**. Collectively, these findings thus establish that while the size of the offsets is encoded within the RBD architecture, it is not encoded in the number of RBD domains either in a linear or a threshold-dependent manner, and instead it appears to be an inherent property that can be encoded even within a single RBD (see **Discussion**).
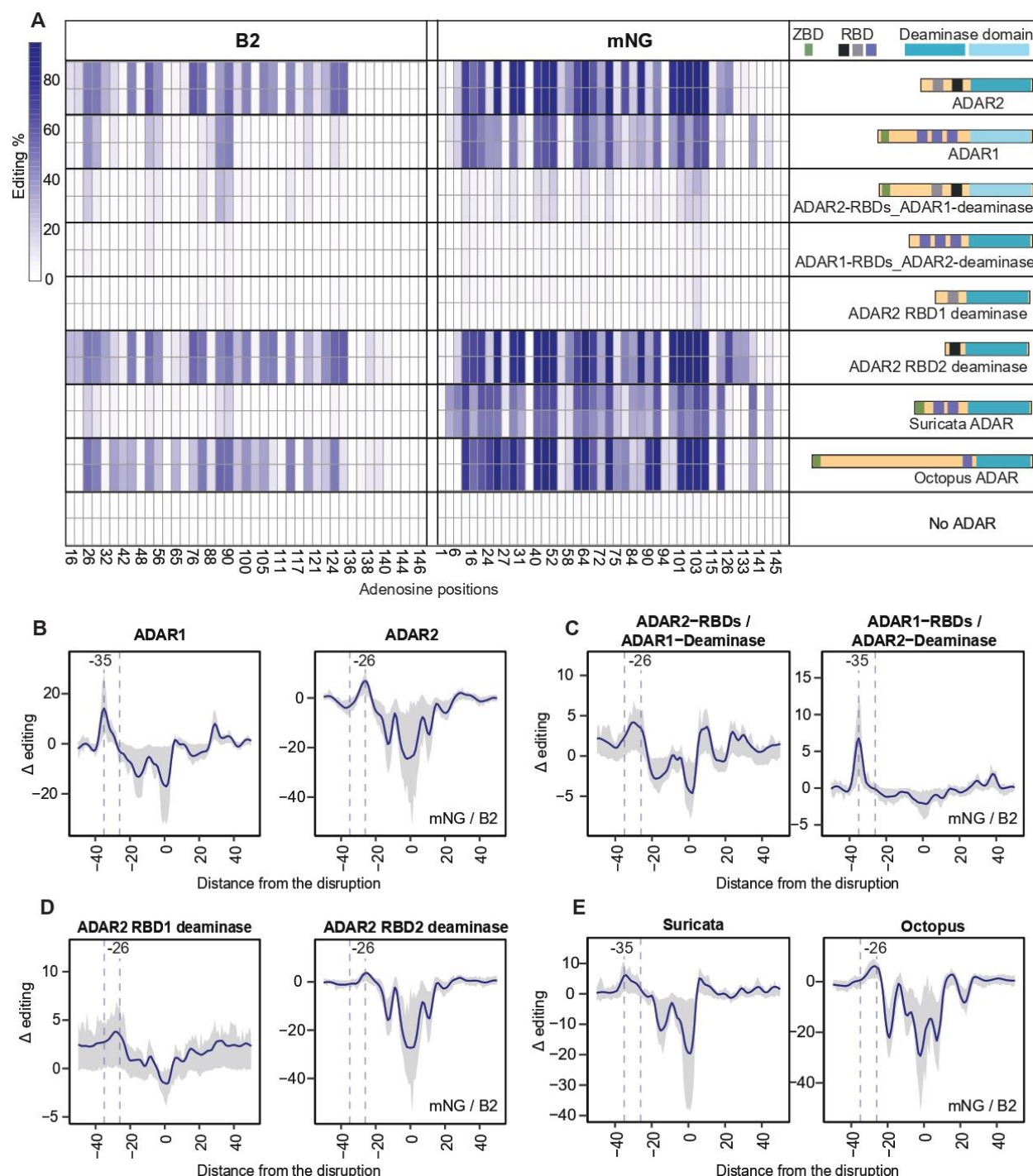
Figure 3. Editing is induced at ADAR-dependent fixed intervals upstream from structural disruptions. A) Heatmap of A-to-I editing levels in the perfect double-stranded constructs in No-ADAR and ADAR-overexpressing ADAR1-KO HEK293T cells. The adenosine positions of the B2 and mNG perfect double-stranded constructs are depicted at the bottom of the heatmap. The illustrations of each ADAR including the RBDs and deaminase domain are depicted on the right side. ZBD: Z-binding domain; RBD: RNA binding domain. B) ADAR1- and ADAR2-mediated editing offset based on subsets of 3-nucleotide mismatch running throughout the mNG and B2 sequences. Data is shown as Figure 2B. C) Editing offset based on subsets of 3-nucleotide mismatch running throughout the mNG and B2 sequences in 'ADAR2-RBDs_ADAR1-deaminase'- and 'ADAR1-RBDs_ADAR2-deaminase'-overexpressing cells. Data is shown as Figure 2B. D) Editing offset retrieved from subsets of 3-nucleotide mismatch running throughout the mNG and B2 sequences in 'ADAR2-RBD1 deaminase'- and ''ADAR2-RBD2 deaminase'-overexpressing cells. Data is shown as Figure 2B. E) Editing offset based on subsets of 3-nucleotide mismatch running throughout the mNG and B2 sequences in 'Suricata ADAR'- and 'Octopus ADAR'- overexpressing cells. Data is shown as Figure 2B.

11

26-bp offset rule can improve the efficiency of ADAR2-mediated targeted editing

To explore whether the newly identified -26 nt rule of ADAR2 might lend itself to improved design of ADAR recruiting therapeutics, we designed ADAR-recruiting RNAs (arRNA) to elicit editing on four distinct endogenous targets harboring distinct consensus motifs: PPIB-ORF:U**A**G, GAPDH-UTR:U**A**G, SMAD4-UTR:C**A**G, PPIB-UTR:U**A**G and STAT1-ORF:U**A**U via recruitment of exogenously expressed ADAR2. For each of these targets, we designed three arRNA constructs: (1) a 151-nt long arRNA containing a C opposite to the target A located between two 75-nt stretches which are perfectly complementary to the endogenous transcript. Such constructs were used in [23,55] and serve as a positive control; (2) an arRNA as in (1) but harboring 3-bp mismatch 26 or 27 nt upstream from the target adenosine, and (3) an empty vector serving as a negative control **(Figure 4A)**. Consistent with our expectations, we found that in 2 of the 5 cases (GAPDH-UTR and SMAD4-UTR), the 3-nt disruptions significantly increased editing levels with respect to the positive controls, and in a third case (PPIB-ORF) the same trend was observed albeit it did not pass statistical significance **(Figure 4B)**. The relatively low increase in these cases as well as the absence of an increase in the two remaining cases are consistent with the relatively mild effect size of induced editing at position -26 **(Figure 3B)** and may be suggestive of context-specificity remaining to be uncovered.

In some clinical contexts, it could potentially be beneficial to induce editing only in cells expressing one of the two ADAR enzymes. Given the different offsets at which ADAR1 and ADAR2 induce editing, we sought to assess whether this could be leveraged to achieve such selective editing. Indeed, we found that an arRNA with a structural disruption at an offset of 35 nt selectively induced editing by ADAR1, and
not by ADAR2, in comparison to a positive control lacking a structural disruption **(Figure 4C-D)**. Conversely, an arRNA with a structural disruption at a 26 nt offset selectively induced editing by ADAR2, and not by ADAR1 **(Figure 4D)**. These results thus suggest that engineered structural disruptions at fixed offsets can be utilized to tune the relative susceptibility of targets to editing via ADAR1 vs ADAR2.

Finally, we sought to assess whether the introduction of structural disruptions at a 26 bp offset would not only increase on-target editing levels but also decrease off-target levels. To assess this, we amplicon-sequenced the GAPDH amplicon following targeted editing via either the 'positive control' or the 'Mismatch 27' arRNA. In this analysis, we only identified a single adenosine that was edited at levels exceeding 2% across either of these two samples. Remarkably, this position was edited at levels of 6.13% in the positive control sample, which decreased to 1.05% in the 'Mismatch 27' samples **(Figure 4E-F)**. This off-target site resided 26 nt downstream of the targeted adenosine, and therefore the reduced editing levels in the 'Mismatch 27' sample are likely a direct consequence of this position no longer being base-paired in the 'Mismatch 27' arRNA. With the caveat of only relying on a single off-target site, these findings suggest that rationally designed structural disruptions within arRNAs can be designed to both increase on-target rates and decrease off-target ones.

Overall, these findings lend support to the observations that structural disruptions lead to increased ADAR2-mediated editing at a fixed offset and provide a proof of principle that this rule can allow improved recruitment of ADAR2 towards target adenosines in therapeutic settings.
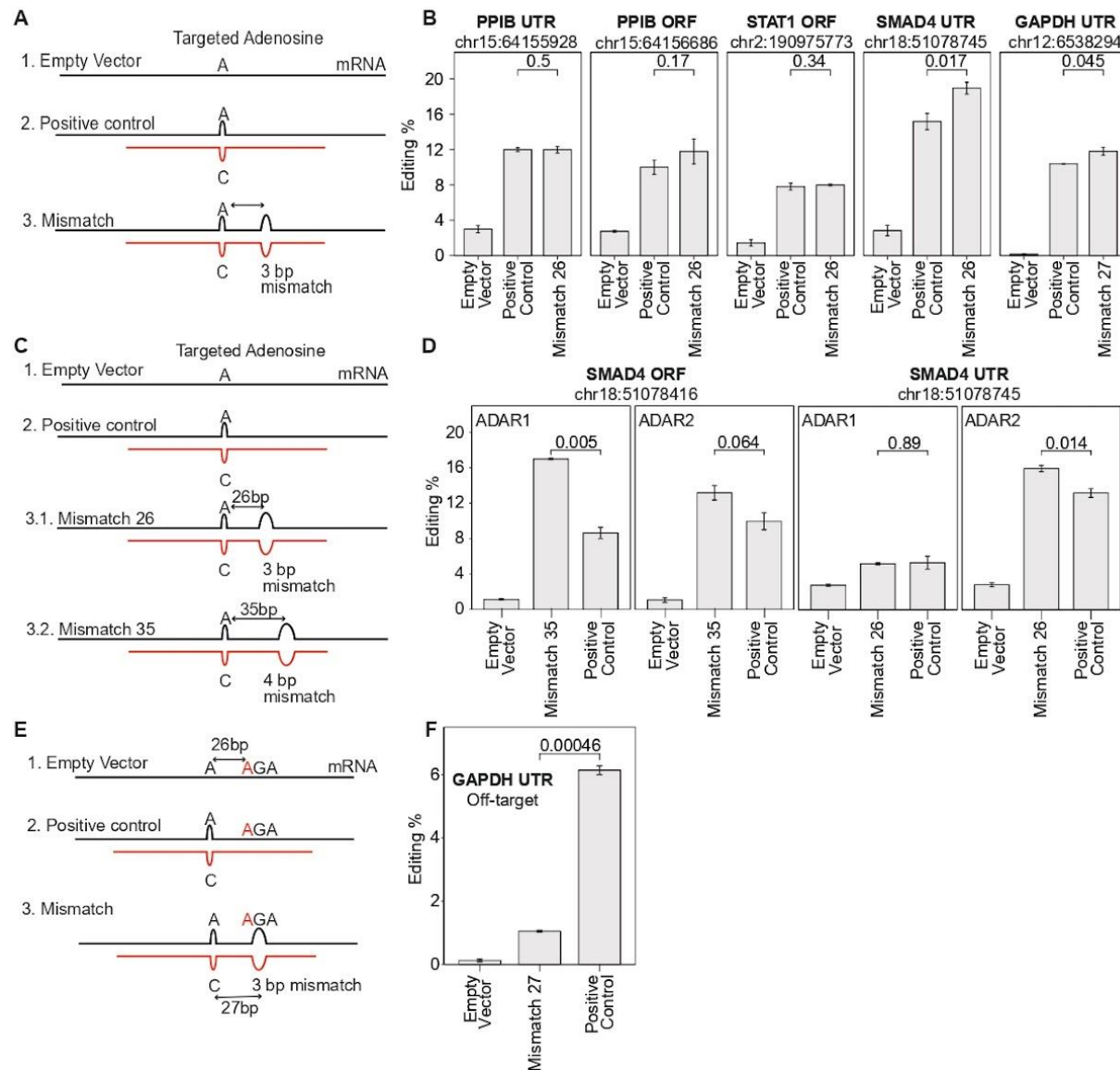
Figure 4. Targeted editing of endogenous transcripts using engineered ADAR2- and ADAR1-recruiting RNAs. A) Scheme of arRNAs targeting endogenous transcripts of *PPIB*, *SMAD4*, STAT1, and GAPDH. (1) The empty vector has no targeting oligo. (2) Positive control construct is a 151-bp-long complementary oligo, with a T to C mismatch opposite of the targeted A. (3) Mismatch 26 construct consists of an arRNA as in (2) but including a 3-bp mismatch at 26 or 27 bases away from the target A site. B) Quantification results showing the editing levels on targeted adenosine of the *PPIB*, *SMAD4*, STAT1, and GAPDH transcripts in ADAR2-expressing cells. Data is shown as the mean ± s.e.m. n = 3. The pairwise comparisons were evaluated using a t-test and the corresponding p-values are shown on the top of the barplots. C) Scheme of arRNAs targeting endogenous transcripts of *SMAD4*. (1) The empty vector has no targeting oligo. (2) Positive control construct is a 151-bp-long complementary oligo, with a T to C mismatch opposite of the targeted A. (3.1) Mismatch 26 construct consists of an arRNA as in (2) but including a 3-bp mismatch at 26 bases away from the target A site. (3.2) Mismatch 35 construct consists of an arRNA as in (2) but including a 4-bp mismatch at 35 bases away from the target A site. D) Quantification results showing the editing levels on targeted adenosine of the *SMAD4* transcript in ADAR1- and ADAR2- expressing cells. Data is shown as **Figure 4B**. E) Scheme of arRNAs targeting endogenous GAPDH transcript. F) Quantification results showing the editing levels on off-targeted adenosine of the *GAPDH* transcript in ADAR2-expressing cells. Data is shown as **Figure 4B**.

13

Characterization of ADAR1 and ADAR2 sequence selectivity across diverse ADAR variants

The experimental design of the oligo-array libraries employed in this study had been primarily geared towards interrogating the impact of RNA secondary structure on editing. Nonetheless, the availability of measurements of editing levels across distinct sites and in varying sequence contexts allowed investigating the impact of sequence on editing, and the extent to which this varied across the eight ADAR variants interrogated here.

We found that across all ADAR enzymes, the position immediately upstream of the edited site were depleted of G at the upstream position, consistent with [38,39,56]. The position immediately downstream displayed less of a bias, consistent with [35,38,39,57]. **(Figure 5A)**. We next explored the extent to which the identity of nucleotides opposite of the target adenosine impacted editing across the different ADAR variants. We found that editing by all ADAR variants was induced when a C was introduced opposite of the target A **(Figure 5B)**. Introduction of A-A or A-G mismatches opposite of the edited site both substantially decreased editing at the targeted position and gave rise to increased editing at a -26 bp offset **(Figure 5B)**. Finally, we extended this analysis to mismatches occurring in the vicinity of the edited site. This analysis revealed that editing at adenosines in a 'GA' context (underlined A is edited) tends to be substantially higher when the cytidine opposite to 'G' at position -1 is mismatched with a guanosine, and even more so with an adenosine **(Figure 5C)**. We further found induced levels of editing when a 3-nt mismatch was centered around an edited site in a 'GA' context **(Figure 5D)**. These findings were consistently observed across all ADAR variants (**Figure S6A-S6B**). The facts that these sequence preferences are independent of the RBD domain structure and that they occur at sites that are in physical interaction with the deaminase domain suggest that these sequence preferences are an inherent property of the ADAR deaminase domains and that they are shared across ADAR1 and ADAR2 homologs.

Given that adenosines in GA contexts are typically edited at low efficiencies, we sought to investigate whether editing in GA contexts could be induced via the introduction of arRNAs designed to harbor a G-G or a G-A mismatch at position -1, or via guides introducing a 3-nt mismatch at the edited site. Indeed, we found that an arRNA harboring a G-A mismatch yielded the highest editing level within a SMAD4-ORF target, followed by arRNA harboring a G-G mismatch, whereas the fully complementary arRNA yielded background levels of editing **(Figure 5E-F)**. These findings are in line with reports by [58,59]. Collectively, our findings establish how editing at target sites can be induced either by introducing mismatches at a relatively distant fixed offset via a mechanism impacting recognition through the RBDs, or in close vicinity to the target site via a mechanism likely impacting recognition through the deaminase domain.
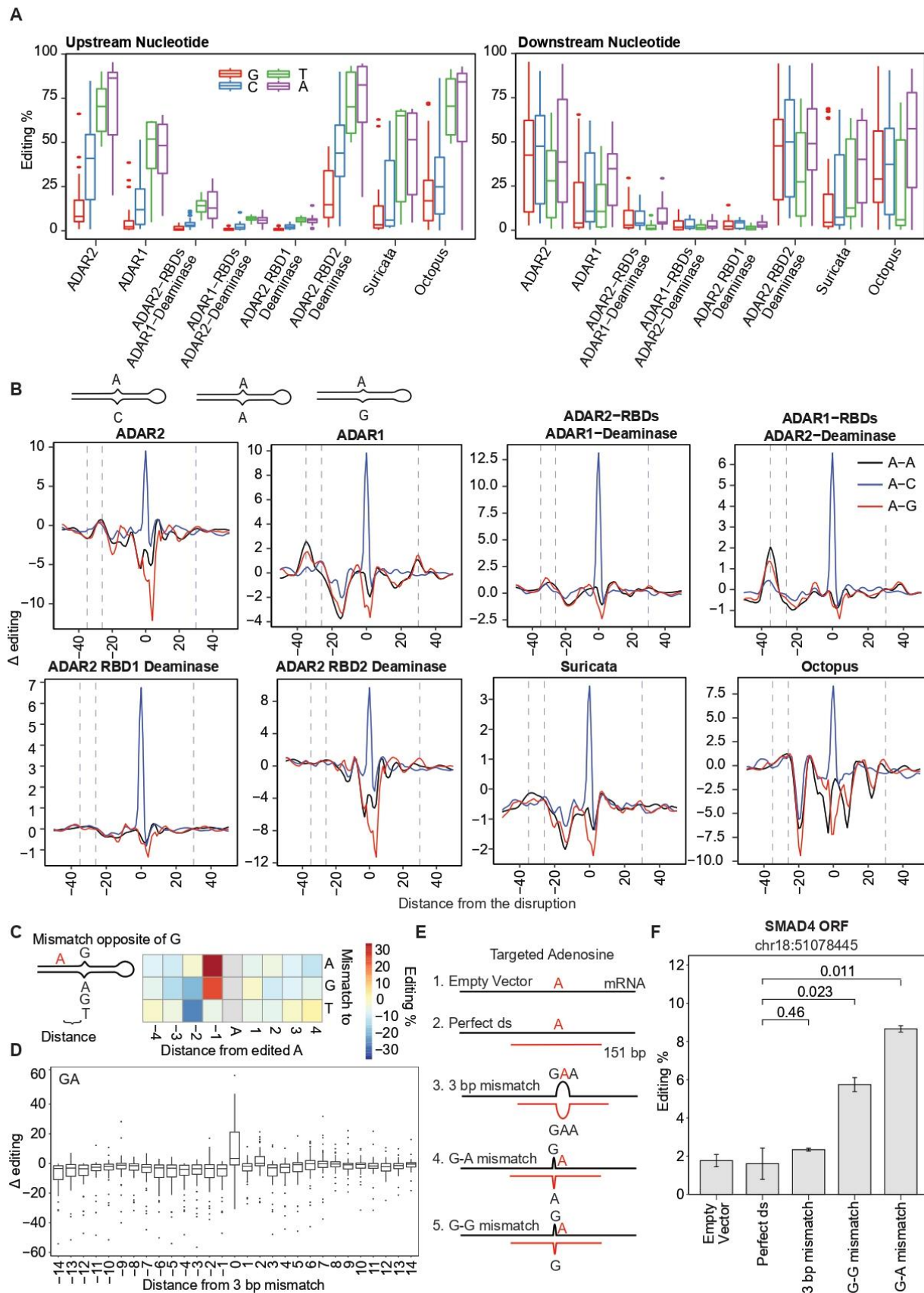
Figure 5. Sequence and structure elements involved in editing nucleation and termination among ADARs. A) Analysis of upstream (left) and downstream (right) nucleotide preference in ADAR-specific editing. Editing levels correspond to adenosines along both the B2 and mNG perfect double-stranded constructs, but As near the loop were excluded. B) Series of constructs characterized by a systematic C, A or G base opposite to A along the stem. Line charts show the effect of the different mismatches on editing. Δ of editing as a function of the distance from the disruption. Fitted curves depict the Loess fit of Δ editing with a span of 0.07. C) Subset of G-mismatching bases that neighbor the edited sites. Left - Graphical scheme. Right - On the Heatmap, the x-axis shows the distance of the disruptions to the A editing site while the y-axis represents the base to which a G is opposite. D) Effect of 3-nt mismatch running through the stem on adenosine sites within the "GA" sequence context. Mismatches differentially located in each construct get centered at 0 on the x-axis. The Δ editing on the y-axis represents the change of the editing level of an adenosine, normalized to the perfect double-stranded construct. The box plot depicts the distribution of Δ editing levels per distance. E) Scheme of arRNAs targeting endogenous *SMAD4* transcript. (1) The "Empty Vector" as negative control. (2) "Perfect ds" construct is a 151-bp-long oligo complementary to the transcript. (3) "3bp mismatch" construct consists of an arRNA as in (2) but containing a 3-nucleotide mismatch opposite to the target A site. (4-5) "G-G and G-A mismatch" constructs consist of an arRNA as in (2) but including a G-G and G-A mismatches one nucleotide upstream from the targeted A site, respectively. F) Quantification results showing the editing levels on the targeted adenosine of the *SMAD4* transcript in ADAR2-expressing cells. Data is shown as the mean ± s.e.m. n = 3. The pairwise comparisons were evaluated using t-test and the corresponding p-values are shown on the top of the barplots.

Discussion

Despite widespread interest in unraveling the determinants guiding the selectivity of ADAR1 and ADAR2, these have remained poorly understood and to a considerable extent unpredictable. It has been previously suggested that the basis for selectivity resides within mismatches [37], bulges, loops [53], and long-range tertiary pseudoknots [60,61]. Such structural elements are evolutionarily conserved [60,62] suggesting that the secondary [35] and tertiary RNA structures [63] play an important role in regulating the editing efficiency and specificity. Accordingly, mismatches and bulges have also been included in the design of prior arRNA recruiting modalities [22,55]. Yet, the rules governing such selectivity - e.g. where do structural mismatches contribute to editing? When are they prohibitive? - have remained poorly understood. Our study contributes two key insights to our understanding: First, we establish a simple rule, namely that structural disruptions of diverse types (bulges, mismatches) will give rise to induced ADAR2-mediated editing at a fixed offset of 26 bp upstream of the disruption, contrasting with ADAR1 which induces editing at a 35 bp offset. Second, we uncover that these distinct offsets by the two ADARs are encoded via the distinct RBD domains of the two enzymes.

Our work uncovers interesting commonalities and differences between the two ADAR enzymes. Activity by both enzymes is induced at a fixed offset from structural disruptions. In both cases, there is substantial evidence for symmetricity, as is evident from comparing the top and bottom strand editing levels. Moreover, in both cases, the induction of editing is orientation-specific, with editing being induced on both strands *upstream* of the structural disruption. However, the size of the offset is different (-35 vs -26 nt). In addition, the magnitude of induction is also different, with more dramatic effects being typically observed for ADAR1 than for ADAR2. Finally, for ADAR1 in addition to the major peak at -35, we had also observed a more minor peak in editing activity 30 bp downstream of the edited site. We do not observe such a downstream peak for ADAR2. This may either reflect a difference in the mechanism driving induced editing, or the lower dynamic ranges which may limit us from clearly observing such a secondary, more minor peak for ADAR2.

A major question left open by our study is the basis for the different offsets of ADAR1 and ADAR2. While based on the RBD swapping experiments it is clearly encoded by the RBD architecture, we rule out that this is a function of the number of RBDs, as offsets of 26 and 35 nt are achieved by variants and mutants with a distinct number of RBDs. Another possibility is that

16

the difference in offset is not due to the difference in domains, but to the difference in the size of the linker between the RBD and the deaminase domain. However, we can largely rule out this possibility as well, because in our RBD swapping experiments between ADAR1 and ADAR2 we had maintained the original linkers, and the offset sizes segregated with the RBDs. Given that single amino acids in the RBD were shown to be important in RNA recognition and binding [43], it is possible that the basis for the difference in selectivity between the two enzymes lies within such individual changes. Dissecting this systematically via genetic approaches is rendered challenging, given that mutations within RBDs oftentimes also abolish editing. Indeed, six additional RBD-disrupting ADAR mutants that we generated over the course of this study (data not shown) failed to show any substantial editing activity, consistent also with previous observations [64]. We anticipate that the structural dissection of these two enzymes bound to RNA targets will provide an answer to this question.

In attempting to understand the basis for a 26 nt offset of ADAR2, we found two potentially relevant clues in the literature. First, in a structural study of the Glu receptor target in complex with the ADAR2 RBDs, each of the two domains was found to associate with 12-14 nt. Thus, 26 nt is well within the range of the size that would be protected by two RBD [43]. While our findings suggest that a 26 nt offset can also be maintained via ADAR mutants and variants harboring a single RBD, they do leave open the possibility that an offset of 26 nt could be the combined outcome of the RBDs of two ADAR enzymes acting as a dimer, given that both ADAR1 and ADAR2 act as homodimers [65–68].

Second, our studies resonate to some extent with findings that ADAR substrates are distributed periodically at ~50 bp intervals from each other [69]. Given that we find editing induced 26 nt upstream of structural disruptions on the top strand, but also at 26 nt upstream of the disruption on the bottom strand **(Fig. 2F)**, and given our previous observations on editing symmetricity [49], it is tempting to speculate that structural disruptions could serve as a mechanism spacing edited sites at ~52 bp intervals from each other. However, in the cited study [69] the same intervals were observed for ADAR1 and ADAR2, whereas different intervals would be predicted for ADAR1 vs ADAR2 based on such a model and our findings, and thus it is unclear to us whether these findings are mechanistically related.

In our study, we also perform proof-of-principle experiments demonstrating that our improved understanding of editing specificity by the two ADAR enzymes lends itself towards the improved design of ADAR recruiting RNA sequences. We demonstrate that the offsets at a fixed distance can enhance on-target editing levels at the specified targets, potentially reduce off-target editing, and can provide some level of control over which of the two enzymes mediates it. While the effect sizes obtained in our hands are in most cases relatively modest, we anticipate that they might potentially be boosted, if combined with more potent arRNAs, such as chemically modified ones [22].

Collectively, our findings shed light on the mechanisms underlying the only partially overlapping target spectrum of ADAR1 and ADAR2, while advancing our technical toolkit to target these two enzymes towards clinically relevant targets.

Methods

ADAR plasmid generation

Full-length human ADAR2 (UniProt: P78563-2), ADAR2 RBD1 Deaminase, and ADAR2 RBD2 Deaminase coding sequences were amplified from the AAVS1-hADAR2, pYES-DEST52-hADAR2-dRBM1-Deaminase domain, and pYES-DEST52-hADAR2-dRBM2-Deaminase domain plasmids, respectively, using primers that included XbaI and EcoRI sites. Full-length human ADAR1 (UniProt: P55265-5) was amplified from the AAVS1-hADAR1 plasmid by primers containing XbaI and HindIII sites. All of those PCR products (primers in Supplementary Table 1) were subsequently digested and ligated into the corresponding restriction sites of the digested pcDNA3.1(-) vector. All the original plasmids were kind gifts of Prof. Ben-Aroya.

For designing ADAR1-ADAR2 hybrid plasmids, the ADAR1-pcDNA3.1(-) and ADAR2-pcDNA3.1(-) plasmids were used as templates for PCR reactions (primers in Supplementary Table 1) using Phusion® Hot Start II DNA polymerase (Thermo Fisher Scientific). The-gel purified DNA fragments were assembled according to Gibson Assembly® Master Mix (NEB). The assembled products were transformed using Gibson Assembly Cloning Kit (NEB), and all constructs were confirmed via Sanger sequencing on PCR-based positive clones (primers in Supplementary Table 1). Final ADAR-plasmid-containing clones were grown in ampicillin-supplemented LB liquid media, and DNA was extracted according to the QIAprep Spin Miniprep Kit (QIAGEN).

The pTwist CMV vectors containing the human-codon optimized sequences of ADAR from *Octopus vulgaris* (UniProt: A0A6P7SCW6_OCTVU), and *Suricata suricatta* (UniProt: A0A673T544_SURSU) were ordered from Twist Bioscience. Bacteria from glycerol stocks were inoculated and grown in ampicillin-supplemented LB liquid media, and plasmid DNA was extracted as previously mentioned.

Transient transfections

ADAR1-knockout HEK293T cells were grown (37°C, 5% CO_2) in Gibco Dulbecco's Modified Eagle Medium (DMEM) supplemented with 10% fetal bovine serum, 1% Penicillin and Streptomycin, and 4ug/ml Puromycin. $5x10^5$ cells were plated on a 6-well plate so that cells reached 70-90% confluency at the time of the second transfection. 24 and 48 hours after cell seeding, 1.6μg of ADAR-expressing pcDNA3.1(-) plasmid and 4μg B2 or mNG library DNA were transfected respectively according to Lipofectamine® 2000 DNA Transfection Reagent Protocol (Thermo Fisher Scientific). 24 hours later, cell harvesting was performed.

RNA processing and library preparation

Total RNA was extracted using Nucleozol (Macherey-Nagel), poly-A selected using oligo dT-beads (Dynabeads mRNA DIRECT Kit life tech), and DNase treated (Thermo Fisher Scientific). The upper constant arms or lower variable arms including 8 nucleotide barcodes of constructs were reverse transcribed, PCR amplified (primers in Supplementary Table 1), and sequenced using NovaSeq 6000 SP Reagent Kit v1.5 (300 cycles).

Data analysis of NGS data

Fastq files were assessed by a custom R script. The read-filtering process removed reads containing wrong start and end, lacking the established barcodes, and misaligning at adenosine positions. Read 1 and 2 were merged into a single sequence by custom truncation and matching. For each

barcode, the editing percentage was quantified as (G/(A+G))*100 at each adenosine position. Δ editing was calculated as the difference of editing levels at adenosine positions between each structurally altered sequence and perfect-double stranded construct, respectively.

Target RNA editing by recruiting exogenous ADAR2 using plasmid-born arRNAs.

Plasmid construction

Gene fragments containing arRNAs and KpnI sites were ordered from Twist Bioscience. All sequences were KpnI-digested and cloned into the digested EPB104 backbone (Addgene plasmid # 68369) with transcription of arRNA driven by a U6 promoter. The list of TWIST gene fragments is described in Supplemental Table 1.

Additionally, the pDECKO-mCherry plasmids expressing "Positive ctrl", "Empty Vector" and "Mismatch 35" arRNAs were retrieved from [49].

Transient transfections

5×10^5 cells were plated on a 6-well plate so that cells reached 70-90% confluency at the time of transfection. 24 hours after cell seeding, 1ug of ADAR1- or ADAR2-expressing pcDNA3.1(-) plasmid, 0.1ug of pEGFP-N1 plasmid (for assessment of transfection efficiency), and 3ug of the corresponding arRNA-expressing plasmid were transfected according to Lipofectamine® 2000 DNA Transfection Reagent Protocol. 24 hours later, the medium was changed, and 12 hours later, cells were harvested.

RNA processing and editing quantification

RNA isolation, DNase digestion, and reverse transcription were performed using NucleoZOL (Macherey-Nagel), Amplification Grade DNase I (Thermo Fisher Scientific), and MultiScribe Reverse Transcriptase cDNA synthesis kit (Thermo Fisher Scientific), respectively. The subsequent PCR with KAPA HiFi HotStart ReadyMix (Roche) was performed using transcript-specific primers (Supplementary Table 1). Finally, A-to-I editing within the target mRNA was determined via Sanger sequencing (Supplementary Table 1) and the quantitative analysis using the EditR tool [70] and MultiEditR [71].

GAPDH amplicon library preparation and analysis of sequencing data

Editing elicited by GAPDH-targeting arRNAs was quantified using Amplicon Illumina Sequencing. Total RNA was poly-A selected using oligo dT-beads (Dynabeads mRNA DIRECT Kit life tech), and DNase-treated (Thermo Fisher Scientific). The target UTR editing region was reverse transcribed, PCR amplified (primers in Supplementary Table 1), and sequenced on the Illumina Novaseq platform. Data was analyzed by a custom R script. Reads containing wrong starting and ending sequences, and GAPDH-unaligned reads were filtered out. The editing percentage was quantified as (G/(A+G)) *100 at the target adenosine position.

Supplementary Figures
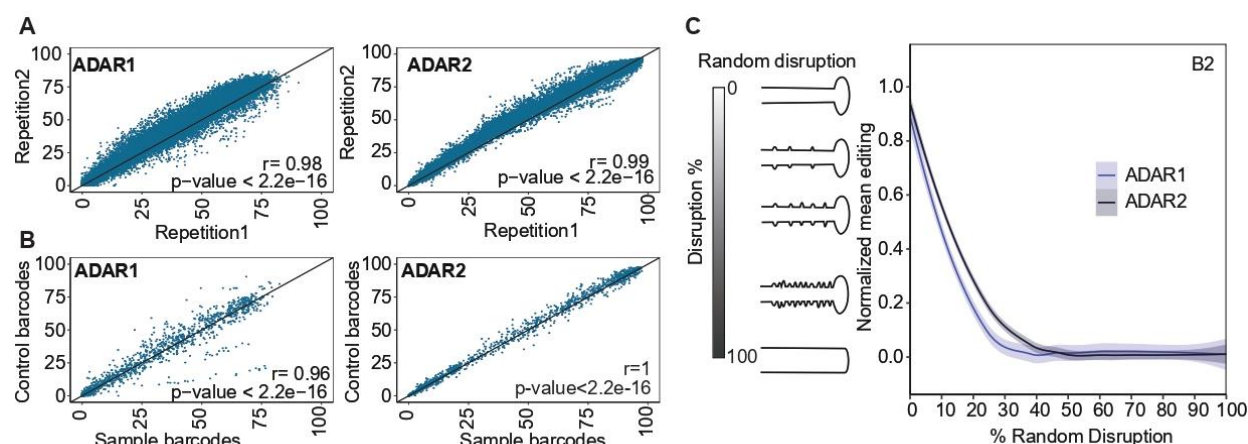
Figure S1. Systematic screening of ADAR2 synthetic substrates in ADAR1 knockout HEK-293T cell lines. A) Correlation of A-to-I levels among technical replicates in cells overexpressing either ADAR1 or ADAR2. Each dot depicts the editing percentage of an adenosine in each construct of the mNG oligo library. The Pearson correlation coefficient and p-values are shown. B) Correlation of editing levels in mNG constructs that differ in the barcode sequences. The Pearson correlation coefficient and p-values are shown. C) Min-max normalized mean editing percentage in the series of B2 constructs containing random disruptions of double-strandedness in 5% increases.
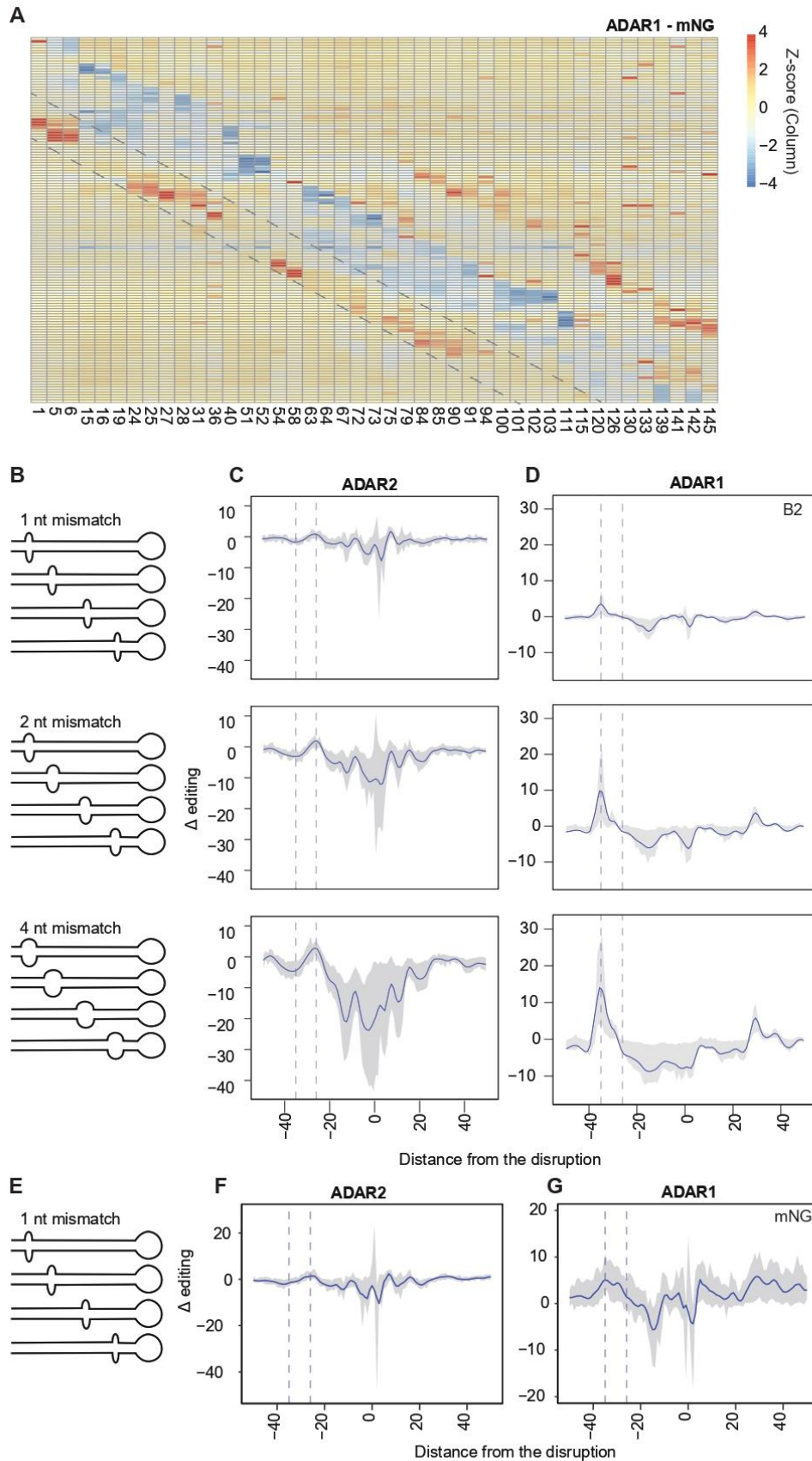
Figure S2. ADAR1- and ADAR2-mediated editing is induced at a constant interval of 26 bp upstream from structural disruptions. A) Heatmap of a 3-nucleotide mismatch running from 5' to 3' throughout the double-stranded RNA. Each row represents a construct structurally disrupted at a specific position while each column represents an adenosine position. Delta (Δ) editing is color-coded after scaling by columns using Z-score transformation (mNG series). The parallel dashed lines highlight the ADAR1-mediated editing increase at fixed distance upstream from the 3-nucleotide mismatch. **B)** Graphical scheme of subsets of B2 constructs carrying 1, 2 or 4 bp mismatches along the stem structures. **C)** ADAR2-mediated editing offset based on subsets of 1-, 2- and 4-nt mismatch running throughout the B2 sequences. Mismatches differentially located in each construct get centered at 0 on the x-axis. The Δ of the editing level on the y-axis represents the change of editing level of an adenosine, normalized to the perfect double-stranded construct. Fitted curves depict LOESS fit of Δ editing with a span of 0.05. The shaded region spans the 25th Percentile and 75th percentile values of Δ editing per distance. Only adenosine positions, which have greater than 1% in editing, on the perfect double-stranded construct were included in the analysis. Vertical dashed lines are placed at -26 and -35. **D)** ADAR1-mediated editing offset based on subsets of 1-,2 and 4- nucleotide mismatch running throughout the B2 sequences. Data is shown as in the figure S2C. **E)** Graphical scheme of subsets of mNG constructs carrying 1 bp mismatch along the stem structures. **F)** ADAR2-mediated editing offset based on the subset of 1-nucleotide mismatch running throughout the mNG sequences. Data is shown as in the figure S2B. **G)** ADAR1-mediated editing offset based on the subset of 1-nucleotide mismatch running throughout the mNG sequences. Data is shown as in the figure S2B.
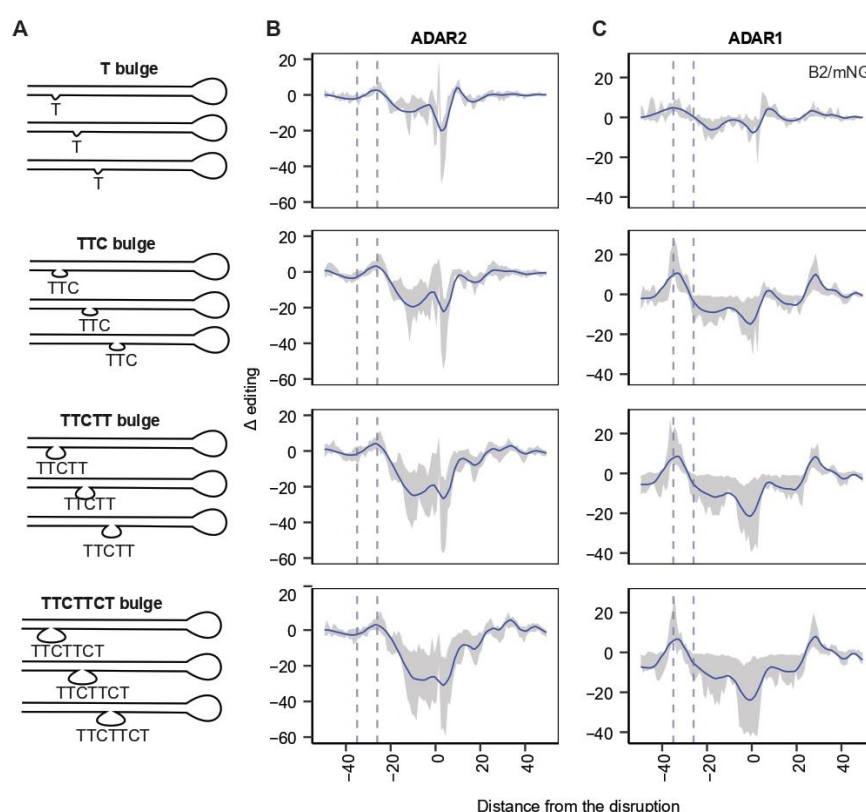
Figure S3. ADAR1- and ADAR2-mediated editing is induced at a constant interval of 35 bp and 26 bp, respectively, upstream from pyrimidine-rich bulges. A) Graphical scheme of subsets of B2 and mNG constructs carrying T, TTC, TTCTT and TTCTTCT bulges along the stem structures. **B)** ADAR2-mediated editing offsets based on subsets of T, TTC, TTCTT and TTCTTCT bulge running throughout the B2 and mNG sequences. Bulges differentially located in each construct get centered at 0 on the x-axis. The Δ editing level on the y-axis represents the change of editing level of an adenosine, normalized to the perfect double-stranded construct. Fitted curves depict LOESS fit of Δ editing with a span of 0.11. The shaded region spans the 25th-75th percentile values of Δ editing per distance. Vertical dashed lines are placed at -26 and -35. **C)** ADAR1-mediated editing offsets based on subsets of T, TTC, TTCTT and TTCTTCT bulge running throughout the B2 sequences. Data is shown as in the figure S3B.
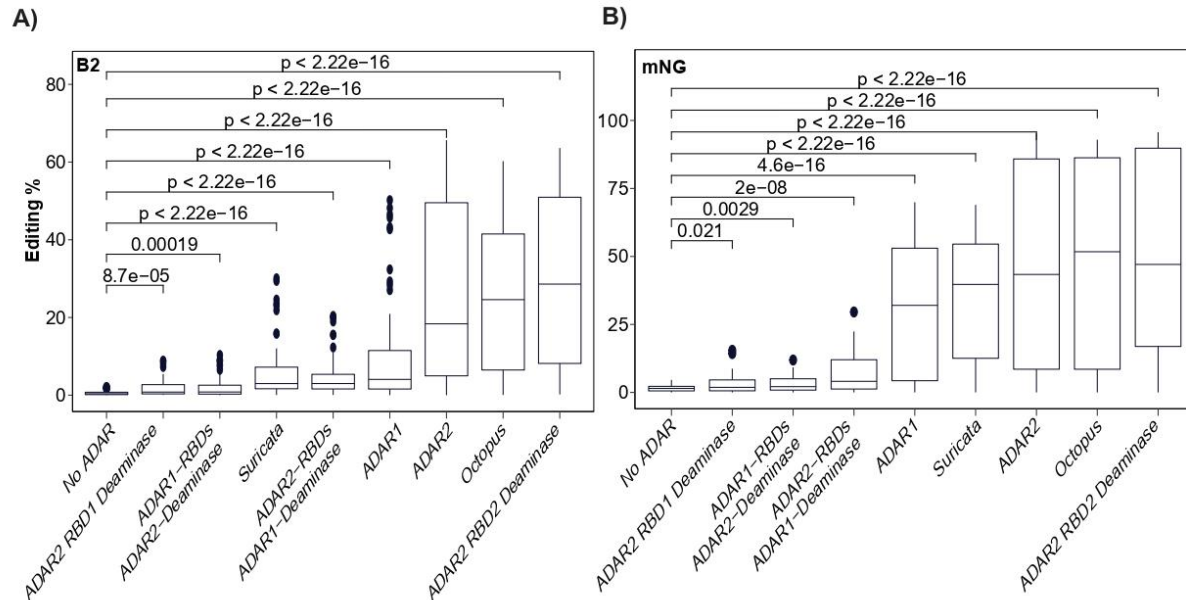
Figure S4. Editing levels of synthetic substrates among different ADARs in ADAR1-knockout HEK293T cell lines. A) Distribution of fraction of edits per position on B2 perfect double-stranded reporter. The pairwise comparisons were evaluated using Wilcoxon-test and the corresponding p-values are shown on the top of the barplots. **B)** Distribution of fraction of edits per position on mNG perfect double-stranded reporter. Data is shown as in Figure S4A.
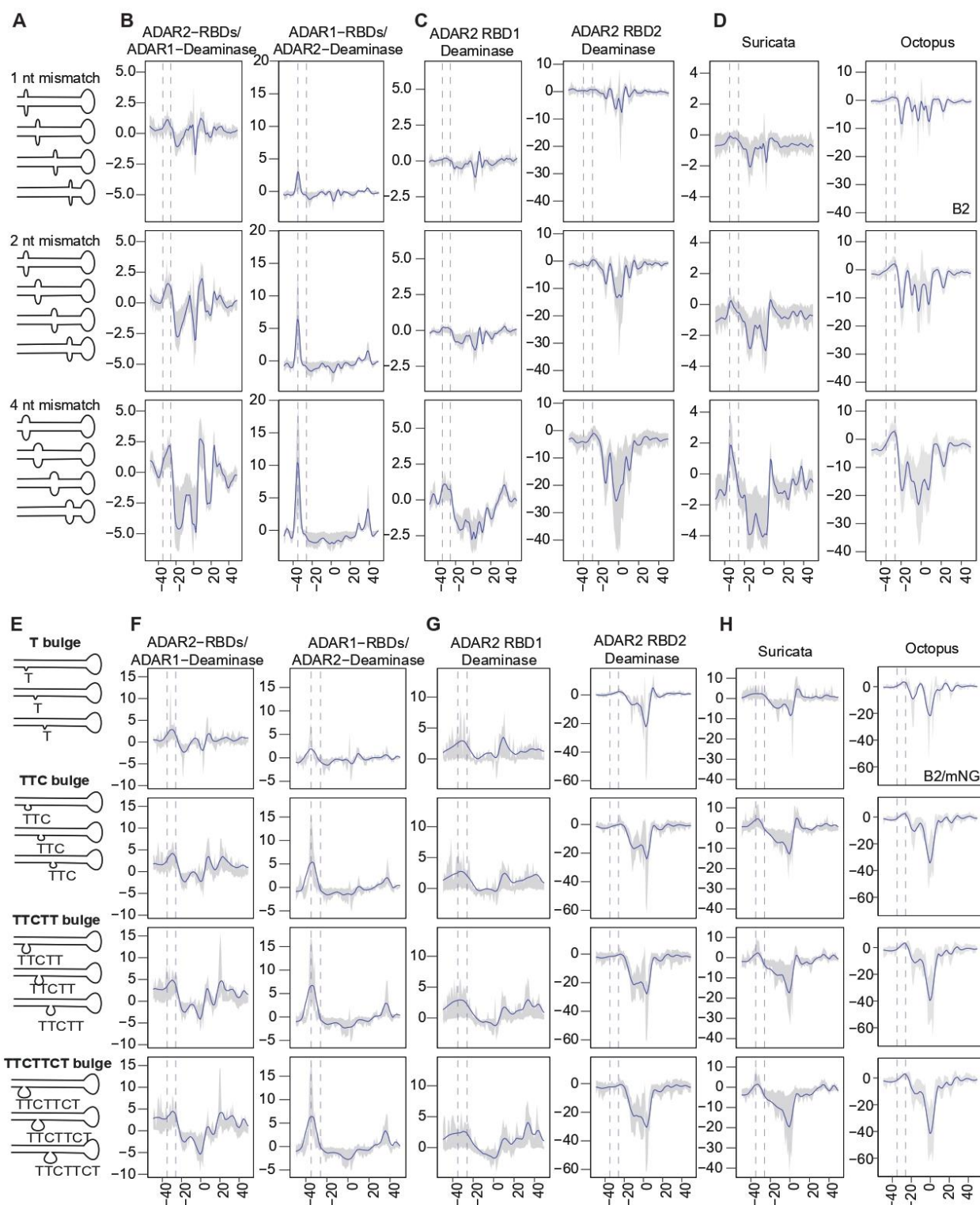
Figure S5. **Editing is induced at ADAR-specific intervals upstream from structural disruption**s. **A)** Graphical scheme of subsets of B2 constructs carrying 1, 2 or 4 bp mismatches along the stem structures. **B)** ADAR2-RBDs_ADAR1-deaminase' and 'ADAR1-RBDs_ADAR2-deaminase'-mediated editing offsets based on subsets of 1-, 2- and 4-nucleotide mismatch running throughout the mNG and B2 sequences. Mismatches differentially located in each construct get centered at 0 on the x-axis. Δ editing level on the y-axis represents the change of editing level of an adenosine, normalized to the perfect double-stranded

construct. Fitted curves depict LOESS fit of Δ editing with a span of 0.05. The shaded region spans the 25th Percentile and 75th percentile values of Δ editing per distance. Only adenosine positions, which have greater than 1% in editing, on the perfect double-stranded construct were included in the analysis. Vertical dashed lines are placed at -35 and -26. **C)** 'ADAR2 RBD1 deaminase' and 'ADAR2 RBD2 deaminase'-mediated editing offsets based on subsets of 1-, 2- and 4-nucleotide mismatch running throughout the mNG and B2 sequences. Data is shown as Figure S4B. **D)** 'Suricata'- and 'Octopus' ADAR-mediated editing offsets based on subsets of 1-, 2- and 4-nucleotide mismatch running throughout the mNG and B2 sequences. Data is shown as Figure S4B. **E)** Graphical scheme of subsets of B2 constructs carrying T, TTC, TTCTT, and TTCTTCT bulges along the stem structures. **F)** 'ADAR2-RBDs_ADAR1-deaminase' and 'ADAR1-RBDs_ADAR2-deaminase'-mediated editing offsets based on subsets of T, TTC, TTCTT and TTCTTCT bulge running throughout the B2 sequences. Data is shown as in the figure S3B. **G)** 'ADAR2 RBD1 deaminase' and 'ADAR2 RBD2 deaminase'-mediated editing offsets based on subsets of T, TTC, TTCTT and TTCTTCT bulge running throughout the B2 sequences. Data is shown as in the figure S3B. **H)** 'Suricata' and 'Octopus' ADAR-mediated editing offsets based on subsets of T, TTC, TTCTT and TTCTTCT bulge running throughout the B2 sequences. Data is shown as in the figure S3B.

Figure S6. Effect of mismatching nearby bases on the opposite strand among ADARs. A) Graphical scheme of constructs harboring A, G, or T opposite to G. **B)** On the heatmaps, the x-axis shows the distance from the mismatch while the Y-axis shows to which base a "G" is mismatched.

References

1. Ramaswami, G. & Li, J. B. RADAR: a rigorously annotated database of A-to-I RNA editing. *Nucleic Acids Res.* **42**, D109–D113 (2014).

2. Porath, H. T., Carmi, S. & Levanon, E. Y. A genome-wide map of hyper-edited RNA reveals numerous new sites. *Nat. Commun.* **5**, 4726 (2014).

3. Bazak, L. *et al.* A-to-I RNA editing occurs at over a hundred million genomic sites, located in a majority of human genes. *Genome Res.* **24**, 365–376 (2014).

4. Gabay, O. *et al.* Landscape of adenosine-to-inosine RNA recoding across human tissues. *Nat. Commun.* **13**, 1184 (2022).

5. Sommer, B., Köhler, M., Sprengel, R. & Seeburg, P. H. RNA editing in brain controls a determinant of ion flow in glutamate-gated channels. *Cell* **67**, 11–19 (1991).

6. Burns, C. M. *et al.* Regulation of serotonin-2C receptor G-protein coupling by RNA editing. *Nature* **387**, 303–308 (1997).

7. Hoopengardner, B., Bhalla, T., Staber, C. & Reenan, R. Nervous system targets of RNA editing identified by comparative genomics. *Science* **301**, 832–836 (2003).

8. Lev-Maor, G. *et al.* RNA-editing-mediated exon evolution. *Genome Biol.* **8**, R29 (2007).

9. Solomon, O. *et al.* Global regulation of alternative splicing by adenosine deaminase acting on RNA (ADAR). *RNA* **19**, 591–604 (2013).

10. Tang, S. J. *et al.* Cis- and trans-regulations of pre-mRNA splicing by RNA editing enzymes influence cancer development. *Nat. Commun.* **11**, 799 (2020).

11. Kawahara, Y. *et al.* Redirection of silencing targets by adenosine-to-inosine editing of miRNAs. *Science* **315**, 1137–1140 (2007).

12. Tomaselli, S. *et al.* ADAR enzyme and miRNA story: a nucleotide that can make the difference. *Int. J. Mol. Sci.* **14**, 22796–22816 (2013).

13. Liddicoat, B. J. *et al.* RNA editing by ADAR1 prevents MDA5 sensing of endogenous dsRNA as nonself. *Science* **349**, 1115–1120 (2015).

14. Yang, S. *et al.* Adenosine deaminase acting on RNA 1 limits RIG-I RNA detection and suppresses IFN production responding to viral and endogenous RNAs. *J. Immunol.* **193**, 3436–3445 (2014).

15. Samuel, C. E. Adenosine deaminase acting on RNA (ADAR1), a suppressor of double-stranded RNA–triggered innate immune responses. *J. Biol. Chem.* **294**, 1710–1720 (2019).

16. Mannion, N. M. *et al.* The RNA-editing enzyme ADAR1 controls innate immune responses to RNA. *Cell Rep.* **9**, 1482–1494 (2014).

17. Liddicoat, B. J., Chalk, A. M. & Walkley, C. R. ADAR1, inosine and the immune sensing system: distinguishing self from non-self. *Wiley Interdiscip. Rev. RNA* **7**, 157–172 (2016).

18. Slotkin, W. & Nishikura, K. Adenosine-to-inosine RNA editing and human disease. *Genome Med.* **5**, 105 (2013).

19. Ishizuka, J. J. *et al.* Loss of ADAR1 in tumours overcomes resistance to immune checkpoint blockade. *Nature* **565**, 43–48 (2019).

20. Aquino-Jarquin, G. Novel Engineered Programmable Systems for ADAR-Mediated RNA Editing. *Mol. Ther. Nucleic Acids* **19**, 1065–1072 (2020).

21. Fukuda, M. *et al.* Construction of a guide-RNA for site-directed RNA mutagenesis utilising intracellular A-to-I RNA editing. *Sci. Rep.* **7**, 41478 (2017).

22. Merkle, T. *et al.* Precise RNA editing by recruiting endogenous ADARs with antisense oligonucleotides. *Nat. Biotechnol.* **37**, 133–138 (2019).

23. Qu, L. *et al.* Programmable RNA editing by recruiting endogenous ADAR using engineered

RNAs. *Nat. Biotechnol.* **37**, 1059–1069 (2019).

24. Katrekar, D. *et al.* In vivo RNA editing of point mutations via RNA-guided adenosine deaminases. *Nat. Methods* **16**, 239–242 (2019).

25. 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. *Proc. Natl. Acad. Sci. U. S. A.* **110**, 18285–18290 (2013).

26. Buchumenski, I. *et al.* Global quantification exposes abundant low-level off-target activity by base editors. *Genome Res.* **31**, 2354–2361 (2021).

27. Tan, M. H. *et al.* Dynamic landscape and regulation of RNA editing in mammals. *Nature* **550**, 249–254 (2017).

28. Chalk, A. M., Taylor, S., Heraud-Farlow, J. E. & Walkley, C. R. The majority of A-to-I RNA editing is not required for mammalian homeostasis. *Genome Biol.* **20**, 1–14 (2019).

29. Morse, D. P., Aruscavage, P. J. & Bass, B. L. RNA hairpins in noncoding regions of human brain and Caenorhabditis elegans mRNA are edited by adenosine deaminases that act on RNA. *Proc. Natl. Acad. Sci. U. S. A.* **99**, 7906–7911 (2002).

30. Levanon, E. Y. *et al.* Systematic identification of abundant A-to-I editing sites in the human transcriptome. *Nat. Biotechnol.* **22**, 1001–1005 (2004).

31. Athanasiadis, A., Rich, A. & Maas, S. Widespread A-to-I RNA editing of Alu-containing mRNAs in the human transcriptome. *PLoS Biol.* **2**, e391 (2004).

32. Hundley, H. A. & Bass, B. L. ADAR editing in double-stranded UTRs and other noncoding RNA sequences. *Trends Biochem. Sci.* **35**, 377–383 (2010).

33. Herb, A., Higuchi, M., Sprengel, R. & Seeburg, P. H. Q/R site editing in kainate receptor GluR5 and GluR6 pre-mRNAs requires distant intronic sequences. *Proc. Natl. Acad. Sci. U. S. A.* **93**, 1875–1880 (1996).

34. Källman, A. M., Sahlin, M. & Ohman, M. ADAR2 A-->I editing: site selectivity and editing efficiency are separate events. *Nucleic Acids Res.* **31**, 4874–4881 (2003).

35. Dawson, T. R., Sansam, C. L. & Emeson, R. B. Structure and sequence determinants required for the RNA editing of ADAR2 substrates. *J. Biol. Chem.* **279**, 4941–4951 (2004).

36. Ohman, Kallman & Bass. In vitro analysis of the binding of ADAR2 to the pre-mRNA encoding the GluR-B R/G site. *RNA* (2000).

37. Wong, S. K., Sato, S. & Lazinski, D. W. Substrate recognition by ADAR1 and ADAR2. *RNA* **7**, 846–858 (2001).

38. Lehmann, K. A. & Bass, B. L. Double-stranded RNA adenosine deaminases ADAR1 and ADAR2 have overlapping specificities. *Biochemistry* **39**, 12875–12884 (2000).

39. Eggington, J. M., Greene, T. & Bass, B. L. Predicting sites of ADAR editing in double-stranded RNA. *Nat. Commun.* **2**, 319 (2011).

40. Grice, L. F. & Degnan, B. M. The origin of the ADAR gene family and animal RNA editing. *BMC Evol. Biol.* **15**, 4 (2015).

41. Wang, Y., Park, S. & Beal, P. A. Selective Recognition of RNA Substrates by ADAR Deaminase Domains. *Biochemistry* **57**, 1640–1651 (2018).

42. Goodman, R. A., Macbeth, M. R. & Beal, P. A. ADAR proteins: structure and catalytic mechanism. *Curr. Top. Microbiol. Immunol.* **353**, 1–33 (2012).

43. Stefl, R. *et al.* The solution structure of the ADAR2 dsRBM-RNA complex reveals a sequence-specific readout of the minor groove. *Cell* **143**, 225–237 (2010).

44. Herbert, A. *et al.* A Z-DNA binding domain present in the human editing enzyme, double-stranded RNA adenosine deaminase. *Proc. Natl. Acad. Sci. U. S. A.* **94**, 8421–8426 (1997).

45. Barraud, P. & Allain, F. H.-T. ADAR proteins: double-stranded RNA and Z-DNA binding domains. *Curr. Top. Microbiol. Immunol.* **353**, 35–60 (2012).

46. Hubbard, N. W. *et al.* ADAR1 mutation causes ZBP1-dependent immunopathology. *Nature* **607**, 769–775 (2022).

47. Jiao, H. *et al.* ADAR1 averts fatal type I interferon induction by ZBP1. *Nature* **607**, 776–783 (2022).

48. de Reuver, R. *et al.* ADAR1 prevents autoinflammation by suppressing spontaneous ZBP1 activation. *Nature* **607**, 784–789 (2022).

49. Uzonyi, A. *et al.* Deciphering the principles of the RNA editing code via large-scale systematic probing. *Mol. Cell* **81**, 2374–2387.e3 (2021).

50. Schaffer, A. A. *et al.* The cell line A-to-I RNA editing catalogue. *Nucleic Acids Res.* **48**, 5849–5858 (2020).

51. Hajji, K. *et al.* ADAR2 enzymes: efficient site-specific RNA editors with gene therapy aspirations. *RNA* **28**, 1281–1297 (2022).

52. Bhalla, T., Rosenthal, J. J. C., Holmgren, M. & Reenan, R. Control of human potassium channel inactivation by editing of a small mRNA hairpin. *Nat. Struct. Mol. Biol.* **11**, 950–956 (2004).

53. Lehmann, K. A. & Bass, B. L. The importance of internal loops within RNA substrates of ADAR1. *J. Mol. Biol.* **291**, 1–13 (1999).

54. Macbeth, M. R., Lingam, A. T. & Bass, B. L. Evidence for auto-inhibition by the N terminus of hADAR2 and activation by dsRNA binding. *RNA* **10**, 1563–1571 (2004).

55. Yi, Z. *et al.* Engineered circular ADAR-recruiting RNAs increase the efficiency and fidelity of RNA editing in vitro and in vivo. *Nat. Biotechnol.* **40**, 946–955 (2022).

56. Morse, D. P. & Bass, B. L. Long RNA hairpins that contain inosine are present in Caenorhabditis elegans poly(A)+ RNA. *Proc. Natl. Acad. Sci. U. S. A.* **96**, 6048–6053 (1999).

57. Polson, A. G. & Bass, B. L. Preferential selection of adenosines for modification by double-stranded RNA adenosine deaminase. *EMBO J.* **13**, 5701–5711 (1994).

58. 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.* **42**, e87 (2014).

59. Doherty, E. E. *et al.* ADAR activation by inducing a syn conformation at guanosine adjacent to an editing site. *Nucleic Acids Res.* **50**, 10857–10868 (2022).

60. Reenan, R. A. Molecular determinants and guided evolution of species-specific RNA editing. *Nature* **434**, 409–413 (2005).

61. Rieder, L. E., Staber, C. J., Hoopengardner, B. & Reenan, R. A. Tertiary structural elements determine the extent and specificity of messenger RNA editing. *Nat. Commun.* **4**, 2232 (2013).

62. Aruscavage, P. J. & Bass, B. L. A phylogenetic analysis reveals an unusual sequence conservation within introns involved in RNA editing. *RNA* **6**, 257–269 (2000).

63. Tian, N. *et al.* A structural determinant required for RNA editing. *Nucleic Acids Res.* **39**, 5669–5681 (2011).

64. Valente, L. & Nishikura, K. RNA binding-independent dimerization of adenosine deaminases acting on RNA and dominant negative effects of nonfunctional subunits on dimer functions. *J. Biol. Chem.* **282**, 16054–16061 (2007).

65. Jaikaran, D. C. J., Collins, C. H. & MacMillan, A. M. Adenosine to inosine editing by

ADAR2 requires formation of a ternary complex on the GluR-B R/G site. *J. Biol. Chem.* **277**, 37624–37629 (2002).

66. Cho, D.-S. C. *et al.* Requirement of dimerization for RNA editing activity of adenosine deaminases acting on RNA. *J. Biol. Chem.* **278**, 17093–17102 (2003).

67. Poulsen, H. *et al.* Dimerization of ADAR2 is mediated by the double-stranded RNA binding domain. *RNA* **12**, 1350–1360 (2006).

68. Thuy-Boun, A. S. *et al.* Asymmetric dimerization of adenosine deaminase acting on RNA facilitates substrate recognition. *Nucleic Acids Res.* **48**, 7958–7972 (2020).

69. Song, Y. *et al.* irCLASH reveals RNA substrates recognized by human ADARs. *Nat. Struct. Mol. Biol.* **27**, 351–362 (2020).

70. Kluesner, M. G. *et al.* EditR: A Method to Quantify Base Editing from Sanger Sequencing. *CRISPR J* **1**, 239–250 (2018).

71. Kluesner, M. G. *et al.* MultiEditR: The first tool for the detection and quantification of RNA editing from Sanger sequencing demonstrates comparable fidelity to RNA-seq. *Mol. Ther. Nucleic Acids* **25**, 515–523 (2021).