


# Precise in vivo RNA base editing with a wobble-enhanced circular CLUSTER guide RNA

Received: 15 July 2023

Accepted: 17 May 2024

Published online: 12 July 2024

 Check for updates

Philipp Reautschnig<sup>1,7</sup>, Carolin Fruhner<sup>1</sup>, Nicolai Wahn<sup>1</sup>, Charlotte P. Wiegand<sup>1</sup>, Sabrina Kragness<sup>2</sup>, John F. Yung<sup>2</sup>, Daniel T. Hofacker<sup>1</sup>, Jenna Fisk<sup>2</sup>, Michelle Eidelman<sup>3,4</sup>, Nils Waffenschmidt<sup>1</sup>, Maximilian Feige<sup>1</sup>, Laura S. Pfeiffer<sup>1</sup>, Annika E. Schulz<sup>1</sup>, Yvonne Füll<sup>1</sup>, Erez Y. Levanon<sup>3,4</sup>, Gail Mandel<sup>2</sup> & Thorsten Stafforst<sup>1,5,6,7</sup> 

Recruiting the endogenous editing enzyme adenosine deaminase acting on RNA (ADAR) with tailored guide RNAs for adenosine-to-inosine (A-to-I) RNA base editing is promising for safely manipulating genetic information at the RNA level. However, the precision and efficiency of editing are often compromised by bystander off-target editing. Here, we find that in 5'-UAN triplets, which dominate bystander editing, G•U wobble base pairs effectively mitigate off-target events while maintaining high on-target efficiency. This strategy is universally applicable to existing A-to-I RNA base-editing systems and complements other suppression methods such as G•A mismatches and uridine (U) depletion. Combining wobble base pairing with a circularized format of the CLUSTER approach achieves highly precise and efficient editing (up to 87%) of a disease-relevant mutation in the *Mecp2* transcript in cell culture. Virus-mediated delivery of the guide RNA alone realizes functional MeCP2 protein restoration in the central nervous system of a murine Rett syndrome model with editing yields of up to 19% and excellent bystander control in vivo.

Site-directed adenosine-to-inosine (A-to-I) RNA base editing is a very promising technology with a clear path for clinical application<sup>1,2</sup>. Hydrolytic deamination of A by enzymes of the ADAR family (adenosine deaminases acting on RNA) produces an inosine, which is biochemically interpreted as G in many cellular processes such as splicing or translation and, consequently, functionally substitutes A with G on the RNA level<sup>3,4</sup>. There are three catalytically active human ADAR proteins: constitutively and ubiquitously expressed ADAR1 p110, interferon-inducible ADAR1 p150 and ADAR2. In the past, ADAR

deaminase domains have been engineered into various artificial editing approaches that enable the efficient and highly programmable editing of any given target A in the transcriptome by applying customized guide RNAs and simple Watson–Crick base-pairing rules. Typical examples are the SNAP-ADAR<sup>5,6</sup>, the  $\Delta$ N-ADAR<sup>7–9</sup> and the Cas13-ADAR approaches<sup>10–12</sup>. However, even after several rounds of optimization, major limitations of such systems remain: (1) a guide RNA plus a protein component needs to be delivered; (2) nonhuman protein domains are included; and (3) global off-target editing hampers the clinical

<sup>1</sup>Interfaculty Institute of Biochemistry, Faculty of Science, University of Tübingen, Tübingen, Germany. <sup>2</sup>Vollum Institute, Oregon Health and Science University, Portland, OR, USA. <sup>3</sup>Mina and Everard Goodman Faculty of Life Sciences, Bar-Ilan University, Ramat Gan, Israel. <sup>4</sup>The Institute of Nanotechnology and Advanced Materials, Bar-Ilan University, Ramat Gan, Israel. <sup>5</sup>Gene and RNA Therapy Center (GRTC), Faculty of Medicine, University of Tübingen, Tübingen, Germany. <sup>6</sup>iFIT Cluster of Excellence (EXC2180) Image-Guided and Functionally Instructed Tumor Therapies, University of Tübingen, Tübingen, Germany. <sup>7</sup>These authors jointly supervised this work: Philipp Reautschnig, Thorsten Stafforst. ✉e-mail: [thorsten.stafforst@uni-tuebingen.de](mailto:thorsten.stafforst@uni-tuebingen.de)

development. An elegant solution to all three limitations could be to harness ubiquitously expressed endogenous ADAR enzymes for RNA base editing. It was shown recently that endogenous ADAR1 can be recruited by either chemically modified antisense oligonucleotides (ASOs)<sup>13,14</sup> or genetically encoded guide RNAs<sup>15–18</sup>. Genetically encoded guide RNAs are particularly desired for the long-lasting reversal of disease-causing G>A point mutations in vivo by viral delivery of the guide RNA component.

Genetically encoded guide RNAs currently suffer from massive bystander editing in the guide RNA–target RNA duplex. This is, on one hand, caused by the large size of the duplex (70–200 bp) and, on the other hand, by the ubiquitous presence of highly editable A bases all over the duplex. Both ADAR1 and ADAR2 prefer similar nearest neighboring bases (for example, U > A > C > G at the 5′ position relative to the target A and G > C ≈ A > U or G > C > U ≈ A at the 3′ position, respectively)<sup>19</sup>. Consequently, bystander editing is dominated by a handful of preferred triplets, particularly all four 5′-UAN triplets (N = A, U, G or C), 5′-AAG and 5′-CAG. Bystander editing can lead to unwanted recoding events in the target and might even cause ribosome stalling<sup>20</sup>. Thus, the avoidance of bystander editing represents a major engineering problem for encodable RNA base-editing systems. Today, several strategies have been suggested. In the LEAPER approach<sup>16,17</sup>, bystander editing is usually suppressed by mismatching some or even all bystander-prone A bases with G. The rationale behind this is the preference of the ADAR deaminase for a specific counter base (C > U > A or G) opposite the targeted A<sup>21</sup>. A more recent approach is U depletion of the guide RNA by keeping off-target prone A bases unpaired, thus producing single-nucleotide bulges within the guide RNA–mRNA duplex<sup>20</sup>. However, both strategies do not always work optimally as they sometimes fail to suppress bystander editing and often reduce editing efficiency, sometimes even dramatically (for example, in A-rich sequence contexts). Consequently, using these strategies to improve editing precision can cost notable editing efficiency. An alternative solution is presented in the CLUSTER approach<sup>15</sup>. Here, the presence of editable triplets in the guide RNA–mRNA duplex is minimized by subdividing the guide RNA into several functional segments. Each segment is designed to bind the target transcript in areas selected for the absence of editable A bases. Because the individual segments of the CLUSTER guide RNAs, which we refer to as recruitment sequences (RSs), can be arranged closely within a window of a few hundred nucleotides around the target site, the approach leads to a very good control of bystander editing and high editing yields. However, in target transcripts that are very rich in highly editable A bases, the RSs currently have to be placed at large distances, which sometimes compromises editing efficiency. In a somewhat related solution, the guide RNA–mRNA duplex is regularly interrupted by bulged-out nucleotides, which leads to a reduction in bystander editing<sup>18</sup>. While the latter guide RNAs are very simple to design, the CLUSTER approach allows the assembly of guide RNAs from a much larger sequence space and was shown to enable a boost of editing efficiency by computationally selecting CLUSTER guide RNAs that avoid inhibitory self-folding.

In the human transcriptome, several highly efficient and precise editing events are known, which are guided by *cis*-acting intronic editing complementary sequences (ECS) that fold back to the target site inside an exon<sup>22–28</sup>. Notably, such natural editing sites are typically not found in perfect RNA helices but rather contain bulges, mismatches and wobble base pairs, which may serve to suppress bystander editing while preserving high on-target editing yields. The exact structural layer that makes an RNA a good or poor substrate for ADAR is still underexplored, although several recent studies have started to address this issue<sup>29–31</sup>.

Here, we systematically explore the rational use of G•U wobble base pairs to affect editing precision and efficiency. The G•U wobble base pair is the most abundant type of non-Watson–Crick base pair in the transcriptome<sup>32,33</sup>. Surrounding the G•U wobble base, the RNA helix structure is perturbed, affecting the groove width, base stacking

and electrostatic profile<sup>33</sup>. Specific structural effects induced by G•U wobble bases have been shown to be important for the interaction of double-stranded RNA (dsRNA)-binding proteins with dsRNA substrates, including ADAR<sup>34</sup>. In this study, we systematically deduce rules on how to place G•U wobble base pairs to improve site-directed RNA base editing in terms of editing precision and efficiency. We apply G•U wobble base pairs in various RNA base-editing approaches including the CLUSTER approach, where it greatly improves the guide RNA design process, giving rise to highly precise and efficient circular CLUSTER guide RNAs that recode a Rett syndrome causing mutation in vivo. The disease in the model animals is caused by a loss-of-function mutation (W104>amber) in the transcription factor methyl CpG-binding protein 2 (MeCP2) that impairs its expression and binding to methylated DNA, thus deregulating >2,500 downstream genes, including many that are relevant for neuronal function<sup>35</sup>. This model system allowed us to assess key factors of successful in vivo transcript repair with endogenous Adars in the central nervous system (CNS).

## Results and discussion

### Wobble bases modulate editing depending on their orientation

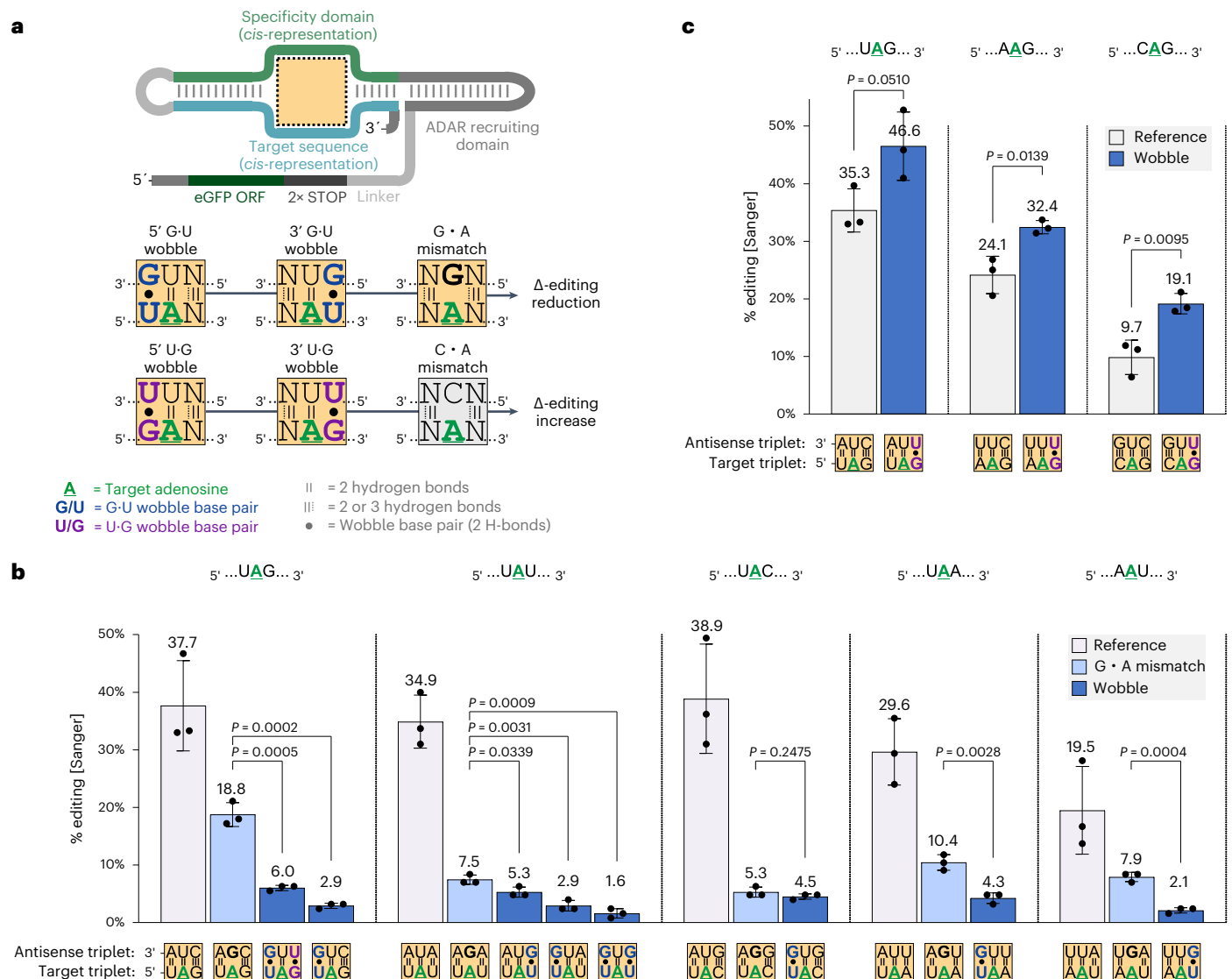
To characterize the effect of wobble base pairs in the nearest neighbor context, we first studied all 12 possible triplets that contained either a U or a G either 5′ and/or 3′ next to the target A. While the target A was always placed opposite of a U base, similar to a regular bystander site, the nearest neighboring nucleotides were either conventionally Watson–Crick base-paired or wobble base-paired (Fig. 1a). Under these circumstances, four types of wobble base pairs can occur. When the U base in a 5′-UAN or 5′-NAU triplet (N = any base) is base-paired with a G, we refer to this as a 5′-G•U or 3′-G•U wobble, respectively; when the G base in a 5′-GAN or 5′-NAG triplet is base-paired with a U, we refer to this as a 5′-U•G or 3′-U•G wobble, respectively. As a benchmark to previous studies<sup>21</sup>, we also included experiments where the target A was mismatched with a G base (G•A mismatch), to suppress editing.

The experiment used an editing reporter construct, based on the earlier R/G-guide RNA approach<sup>36,37</sup>. While the *trans*-acting guide RNA comprised only a double-stranded ADAR-recruiting domain and a single-stranded 20-nt specificity domain (SD), the *cis*-acting reporter additionally contained its own 20-nt target sequence (TS) and was located in the 3′ untranslated region (UTR) of an *eGFP* transcript to enable convenient Sanger sequencing readout (Fig. 1a). Editing was performed by transfecting the plasmid-borne editing reporter into Flp-In T-REx cells, overexpressing ADAR1 p110.

Importantly, we found that both, 5′-G•U and 3′-G•U wobble base pairs strongly suppress editing in the five triplets, 5′-UAN and 5′-AAU (Fig. 1b and Supplementary Fig. 1), which are highly editable under normal Watson–Crick base-pairing conditions and are, thus, a major source of bystander off-target events for *trans*-acting guide RNAs. In the 5′-UAN triplet, both G•U wobbles seemed to cooperate. For four of the five triplet contexts, the suppressive effect of the G•U wobble on editing significantly outcompeted the suppressive effect of the G•A mismatch (Fig. 1b). Interestingly, we found the opposite effect for the U•G wobble base pair. In particular, for the three triplets 5′-UAG, 5′-AAG and 5′-CAG, a clear editing-enhancing effect on the A directly adjacent to the 3′-U•G wobble was apparent (Fig. 1c). Because of the inability of ADARs to achieve sufficient editing at 5′-GAN triplets, the effect of 5′-U•G wobbles could not be verified (Supplementary Fig. 1).

The 5′-UAG triplet can simultaneously accommodate both a suppressive 5′-G•U and an enhancing 3′-U•G wobble base pair and, thus, allows studying their interplay. Our data suggest that the suppressive effect of the G•U wobble entirely dominates the activating effect of the U•G wobble (Fig. 1b).

It is intriguing to speculate that the enhancing effect of the 3′-U•G wobble could be combined with or replace the activating effect of the commonly used C•A mismatch at an on-target site such as 5′-UAG.



**Fig. 1 | Wobble base pairs modulate RNA editing in an orientation-dependent manner. a**, Design of the *cis*-acting editing reporter and illustration of the applied triplet base-pairing motifs. The orientation of wobble base pairs at nearest neighbor positions or the presence of G or C counter bases modulate  $\Delta$ -editing at the central A of a triplet in comparison to its fully Watson–Crick base-paired counterpart. All triplet base-pairing motifs with other background color were installed at the empty dotted outline within the *cis*-acting editing

construct to generate the results in **b,c**. **b**, Suppression of RNA editing in different target triplets using G•U wobble base pairs or G•A mismatches. **c**, Enhancing RNA editing in different target triplets using U•G wobble base pairs. The Sanger sequence analysis in **b** and **c** was performed after transfection of editing reporter plasmids into ADAR1 p110 Flp-In T-REx cells. Data in **b** and **c** are shown as the mean  $\pm$  s.d. of  $n = 3$  biological replicates. For statistical analysis, a Student's *t*-test (two-tailed, parametric) was applied.

We tested this idea using the CLUSTER guide RNA system on three endogenous targets (Supplementary Fig. 2a,b), with a mixed outcome. Only in one example (*ACTB*) was the U•G wobble a promising alternative to the C•A mismatch, although it could be useful in certain other sequence contexts for guide RNA structure optimization.

A guide RNA might find and bind to near-cognate sequences within the transcriptome to induce off-target editing. In this context, we aimed to understand whether a randomly occurring G•U wobble suppresses editing only in the context of a U-paired A or also in the rare context of a C-mismatched A, which is known to be more prone to editing. We evaluated the potential of 5'-G•U wobbles for suppressing editing at 5'-UAG sites where the A was mismatched with C. In two of three examples (*GUSB* and *NUP43*), the suppressive effect of the 5'-G•U wobble was strong enough to suppress editing even at the mismatched A (Supplementary Fig. 2c).

To see whether the enhancing and suppressing effects of wobble base pairs also apply for ADAR1 p150 and ADAR2, we selected the 5'-UAG

triplet, which is a particularly frequent site of bystander editing, and transfected the corresponding plasmid-borne editing reporters into Flp-In T-REx cells, overexpressing ADAR1 p150 or ADAR2, respectively. As expected, the underlying mechanisms were not ADAR isoform dependent and worked equally well (Extended Data Fig. 1).

### G•U wobbles improve precision and efficiency of LEAPER guides

We next applied G•U wobble base pairs in the context of unstructured 111-nt-long LEAPER guide RNAs<sup>16</sup> (Fig. 2a and Supplementary Fig. 3), which are highly prone to bystander editing. Our data (Fig. 1b) indicated that G•U wobble base pairs would be convenient to suppress bystander editing in the five highly editable triplets 5'-UAN (N = A, U, G or C) and 5'-AAU and could be combined with the G•A mismatch at all other editable triplets, such as 5'-AAG and 5'-CAG. To test this concept, *trans*-acting LEAPER guide RNAs (Fig. 2a) encoded on plasmids were cotransfected into HeLa cells with plasmids carrying the full-length



**Fig. 2 | G-U wobbles improve efficiency and precision of *trans*-acting LEAPER guide RNAs for exogenous targets.** **a**, Schematic of 111-nt-long unstructured linear LEAPER guide RNA. **b–d**, Editing heat maps of the LEAPER guide RNA-binding sites within the indicated transcripts: *AHI1* (**b**), *COL3A1* (**c**) and *BMP2* (**d**). The basic design column, LEAPER, lacks G•A mismatches and wobble base pairs. The other guide RNAs contain either G•A mismatches or G•U wobbles at G•U-amenable sites or a combination of both solutions at all bystander sites. In the

latter case, G•A mismatches are placed at sites not amenable to G•U wobbles. The triplet context for each listed editing event is given with the target A highlighted in green and all off-target A bases in blue. The position of each site is given relative to the transcript and the target A (±0 position). Editing was performed with plasmid-borne guide RNA and target in HeLa cells (endogenous ADAR). Data are shown as the mean editing percentage ± s.d. of  $n = 3$  (*AHI1* and *COL3A1*) or  $n = 5$  (*BMP2*) biological replicates.

complementary DNA (cDNA) of one of three different target genes (*AHI1* (ref. 38), *COL3A1* (ref. 39) and *BMP2* (ref. 40)) each carrying a disease-relevant W>amber STOP mutation (5'-UAG) (Fig. 2b–d). To also evaluate endogenous targets, LEAPER guide RNAs targeting a 5'-UAG within the 3' UTR of *NUP43* and *RAB7A* were transfected into HEK293FT cells expressing these genes (Extended Data Fig. 2a,b). As seen in both settings, LEAPER guide RNAs recruited endogenous ADAR to induce significant on-target editing in each of the five targets (54–80%); however, this was contaminated with massive bystander editing (>10 sites per target), as previously reported<sup>15</sup>.

First, we tested for all five targets whether G•U wobble base pairs outcompete G•A mismatches to suppress bystander editing at such

sites, where G•U wobbles are amenable (G•U at G•U-amenable sites) and compared such guide RNAs with guide RNAs that apply G•A mismatches at the same sites (G•A at G•U-amenable sites) (Fig. 2b–d and Extended Data Fig. 2a,b). Notably, we found that the G•U wobble strategy was very potent to suppress bystander editing, often but not always outcompeting the G•A mismatch approach. Moreover, the on-target editing yield was higher for four of the five targets (*AHI1*, 74% versus 51%, and *COL3A1*, 47% versus 29%, Fig. 2; endogenous *NUP43*, 36% versus 17%, and endogenous *RAB7A*, 45% versus 25%, Extended Data Fig. 2).

Second, we aimed to suppress bystander editing entirely by either fully relying on the prior-art G•A mismatch approach (G•A all off-target sites) or by combining G•A mismatches with G•U wobble base pairs



(G•A and G•U at all off-target sites). For *AHII* (Fig. 2b), only the combination of G•A mismatches with G•U wobble base pairs achieved the entire suppression of bystander editing and a good on-target editing yield (45%), while the guide RNA using only G•A mismatches suffered from residual bystander editing (position -38, 43%; position -35, 15%) and a reduced on-target efficiency (26%). These results highlight the power of G•U wobbles to improve the precision of LEAPER guide RNAs. For *COL3A1* (Fig. 2c) and *NUP43* (Extended Data Fig. 2a), the on-target yields were again clearly better when complementing G•A mismatches with wobble base pairs (41% versus 21% and 23% versus 12%, respectively). However, because of the lower yields at all edited sites, the G•A mismatch approach appeared to give slightly better precision. The *BMPR2* target transcript (Fig. 2d) gave equal on-target yields for both strategies but better precision when G•U wobble base pairs were included. In the case of *RAB7A* (Extended Data Fig. 2b), both designs performed similarly.

According to the literature, LEAPER guide RNAs that carry numerous G•A mismatches can suffer from a loss of editing efficiency<sup>16</sup>. The reason for this might be that G•A mismatches have the lowest duplex stability among all known nucleotide mismatches<sup>41</sup>. In contrast, the thermodynamic stability of the G•U wobble is comparable to the A•U base pair<sup>32</sup>. This might partly explain why on-target efficiency often benefitted when G•A mismatches were complemented with G•U wobble base pairs. Notably, the number of suppressive G•U wobble base pairs required to improve precision can be smaller than that of G•A mismatches because one G•U wobble acts simultaneously at its 5' and 3' nearest neighbor positions. Examples can be found in Fig. 2b,c (*AHII*, positions +29 and +31; *COL3A1*, positions -5 and -3). Furthermore, we found several cases where G•A mismatching failed to fully suppress bystander editing, while the G•U wobble strategy succeeded (for example, *AHII* at positions -38, -35 and -31 and *BMPR2* at positions -32, +32 and +35; Figure 2b,d).

After targeting both exogenous and endogenous transcripts ranging from high expression levels (*AHII*, *BMPR2* and *COL3A1* all as cDNAs), over medium (*RAB7A*, normalized transcripts per million (nTPM) = 105) to low (*NUP43*, nTPM = 36) expression levels, our data suggest that the G•U wobble strategy is unaffected by target transcript abundance and, thus, widely applicable. Overall, the wobble strategy complements the prior-art G•A mismatch strategy very well and regularly improves both editing efficiency and precision.

### G•U wobbles are widely applicable to RNA base-editing tools

Bystander editing is a common problem of all RNA base-editing systems, particularly those such as the  $\lambda$ N-ADAR<sup>8</sup> and Cas13-ADAR<sup>10</sup> approaches (Fig. 3a–d and Supplementary Fig. 3) that also use genetically encoded guide RNAs. Thus, we applied the G•U wobble strategy with boxB guide RNAs (Fig. 3a) and direct repeat (DR) guide RNAs (Fig. 3c) to see whether it improves the precision of the  $\lambda$ N-ADAR and the Cas13-ADAR approaches, respectively. For this, triple-plasmid protocols were used where plasmids encoding the guide RNA, the editase ( $\lambda$ N-ADAR2Q or Cas13-ADAR2Q) and the target (*AHII* W725X) were cotransfected into HeLa cells. In contrast to the LEAPER approach, the  $\lambda$ N-ADAR and Cas13-ADAR approaches apply a hyperactive ADAR mutant and use guide RNAs with comparably short antisense part (boxB, 49 nt; DR, 59 nt). Because of the shorter duplex, the number of bystander sites is overall smaller. Nevertheless, the  $\lambda$ N-ADAR approach induced significant bystander editing (Fig. 3b). Notably, amenable sites were readily controlled by G•U wobbles but not by G•A mismatches alone. A combination of G•U wobble and G•A mismatches was able to fully suppress bystander editing at a minor cost of editing efficiency ( $65\% \pm 7\%$  to  $49\% \pm 2\%$ ; Fig. 3b). For Cas13-ADAR, the combination of G•U wobble and G•A mismatches gave the best results in terms of editing efficiency ( $25\% \pm 3\%$  versus  $20\% \pm 4\%$ ) and allowed for complete bystander suppression (Fig. 3d). However, the Cas13-ADAR system itself gave dramatically lower editing yields compared to the  $\lambda$ N-ADAR

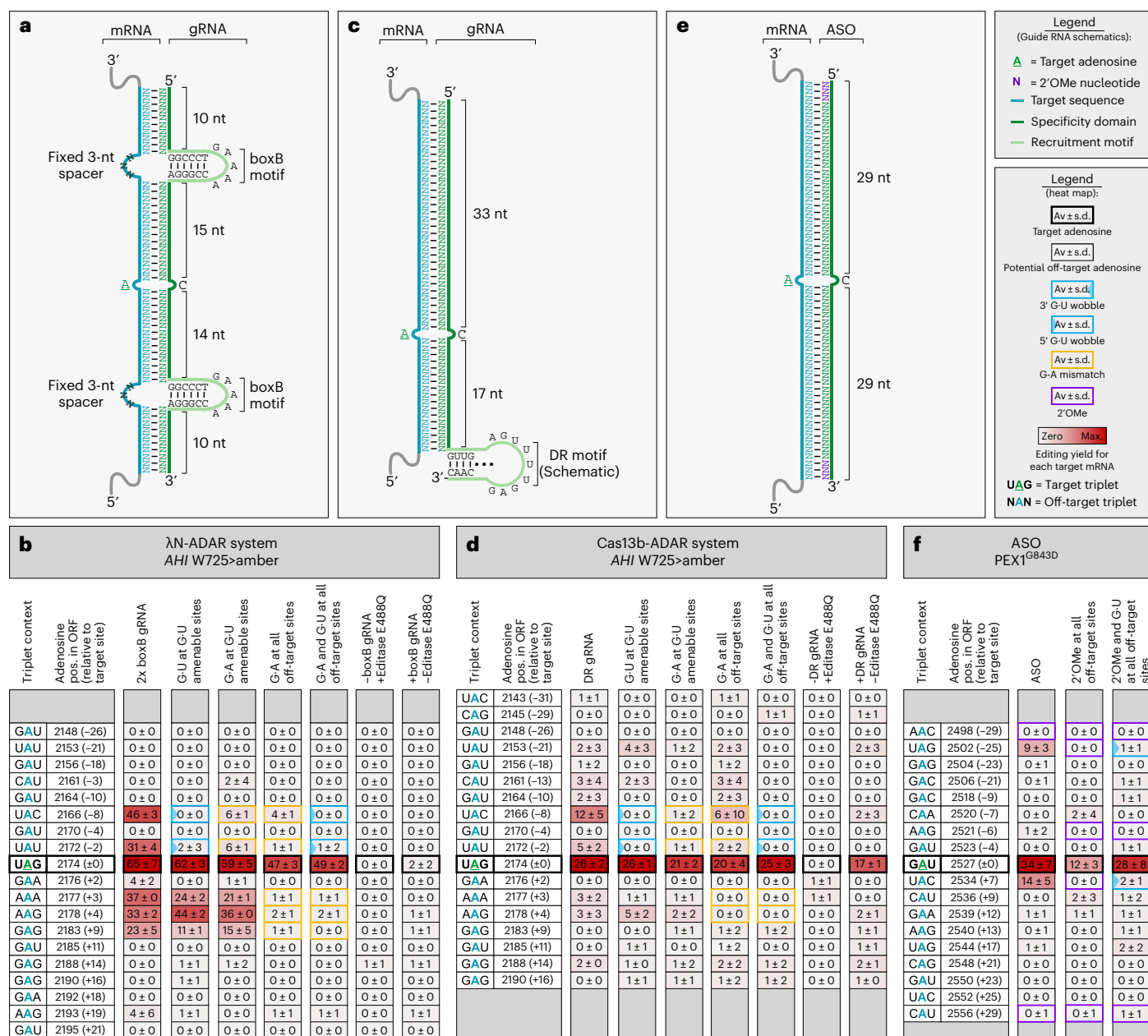
system ( $26\% \pm 2\%$  versus  $65\% \pm 7\%$ ) and showed little specificity for the Cas13b protein, as the editing yield of the DR guide RNA with and without overexpression of the editase differed only by  $\sim 9\%$  (Fig. 3d).

In ADAR-recruiting ASOs, the strategic placement of chemical modifications allows to control bystander events<sup>13,14</sup>. However, dense chemical modification, for example with 2'-O-methylated ribose (2'-OMe), can interfere strongly with editing. Thus, we evaluated G•U wobble base pairs in a case where additional chemical modifications diminished the on-target efficiency. Using a chemically modified (phosphorothioate linkage and 2'-OMe end-blocked) 59-nt-long symmetric ASO (Fig. 3e), we targeted the *PEX1* transcript, specifically the G843D substitution causative for the peroxisome biogenesis disorder Zellweger syndrome<sup>42</sup>. While placement of additional 2'-OMe modifications at the -25, -6 and +7 positions controlled bystander editing, they also reduced the on-target yield drastically ( $34\% \pm 7\%$  to  $12\% \pm 3\%$ ; Fig. 3f). By contrast, a combination of G•U wobble base pairs and 2'-OMe modifications enabled the control of bystander editing while preserving the on-target yield ( $28\% \pm 8\%$ ; Fig. 3f). The latter example shows that ASO-based approaches can also potentially benefit from wobble base pairs to maintain high on-target yields.

### Superior off-target control in A-rich target sites

The suppression of bystander editing in closest proximity to an on-target A is a common problem for all fully encoded RNA base-editing systems. Strategies such as G•A mismatching<sup>16</sup> or U depletion<sup>17</sup> often lead to a substantial loss of editing yield when they are applied too close to the on-target site. To assess the G•U wobble strategy for such a setting, we systematically tested how far the suppressive effect of a single G•U wobble base pair extends in the 5' and 3' directions. For this purpose, we again used *cis*-acting constructs that placed an editable duplex in direct extension of an ADAR-recruiting domain (R/G helix; Extended Data Fig. 3a and Supplementary Fig. 4) into the 3' UTR of an *eGFP* reporter. We then evaluated the editing yields after transfection of these constructs into ADAR1 p110-expressing Flp-In T-Rex 293 cells (Supplementary Fig. 4). To study the suppressive effect of the 3'-G•U wobble in the 5' direction, we studied a series of three 5'-UA(A)<sub>i</sub>U base-pairing motifs ( $i = 1-3$  A bases) with increasing distance between the 3'-G•U wobble and the on-target A (Supplementary Fig. 4a–c). Accordingly, we also designed a series of three 5'-U(A)<sub>i</sub>AG base-pairing motifs ( $i = 1-3$  A bases) to test the effect of a 5'-G•U wobble in the 3' direction (Supplementary Fig. 4d–f). The target triplet was either 5'-UAA or 5'-AAG, hereinafter indicated by square brackets. In both series, we also benchmarked the effect of the G•A mismatch for the same bystander off-target A site. Neither the 5'-G•U nor the 3'-G•U wobble affected the on-target editing yield negatively at any distance tested. Instead, the suppressive effect was almost entirely focused on the direct 3' and 5' neighboring base. This was in clear contrast to the G•A mismatch where not only the mismatched base but also the first and sometimes even the second neighboring base in both directions were negatively affected (Supplementary Fig. 4). Thus, the G•U wobble strategy should be particularly strong to precisely suppress bystander editing close to an on-target A. We show exemplary data on how the 5'-G•U wobble controls editing precision in a 5'-U[AA]G base-pairing motif (Extended Data Fig. 3a,c,e) and how the 3'-G•U wobble acts in the 5'-[UAA]U base-pairing motif (Extended Data Fig. 3a,d,f), always in comparison to the G•A mismatch. In both cases, the G•U wobble clearly gave a better balance of editing efficiency over editing precision.

Next, we transferred the concept to *trans*-acting CLUSTER guide RNAs (linear design), which targeted a 5'-U[AA]G site in the *BMPR2* transcript (K984, on cDNA) by harnessing endogenous ADAR in HeLa cells (Extended Data Fig. 3b,c,g). As expected, the reference guide RNA showed good on-target yields (A, 59%; Extended Data Fig. 3g) but also a strong bystander editing at the 5' neighboring A (30%). The strategically placed 5'-G•U wobble base pair was able to fully suppress this bystander editing. This was not the case with the G•A mismatch



**Fig. 3 | G•U wobble base pairs are broadly applicable to numerous side-directed RNA base-editing systems. a**, Schematic of the λN-ADAR editing system. The 2x boxB motif-containing guide RNA (84 nt) binds its target mRNA (through 49 bp) to recruit the engineered hyperactive λN-ADAR2Q editase. **b**, Editing heat map of exogenous AHI1 W725>amber mRNA targeted by the λN-ADAR system. **c**, Schematic of the Cas13b-ADAR system. The DR motif-containing guide RNA (87 nt) binds its target mRNA (through 51 bp) to recruit the engineered PspCas13b-ADAR editase carrying a hyperactive ADAR2 E488Q deaminase domain. **d**, Editing heat map of exogenous AHI1 W725>amber mRNA targeted by the Cas13b-ADAR system. **e**, Schematic of a symmetric ASO bound to its target mRNA (through 58 nt). The ASO is end-blocked by 2'-OMe, contains phosphorothioate linkages and recruits endogenous ADARs. **f**, Editing heat map of the ASO binding site within exogenous *PEX1<sup>G843D</sup>* transcript. In **b,d,f**, the

basic designs (columns 2x boxB, DR guide RNA and ASO) do not contain G•A mismatches, wobble base pairs or 2'-OMe modifications beyond the end-blocks. The other guide RNAs contain additional 2'-OMe modifications, G•A mismatches, G•U wobbles at G•U-amenable sites or a combination of these solutions at all bystander sites. In the case of a combined solution, G•A mismatches or 2'-OMe modifications are placed at sites not amenable to G•U wobbles. The triplet context for each listed editing event is given with the target A highlighted in green and all off-target A bases in blue. The position of each site is given relative to the transcript and the target A (±0 position). Editing was performed in HeLa cells using plasmid-borne guide RNAs (2x boxB and DR guide RNA) and editase (λN-ADAR2Q and Cas13b-ADAR2Q) or ASOs recruiting endogenous ADAR. Data are shown as the mean editing percentage ± s.d. of *n* = 3 biological replicates.

where only a partial suppression of bystander editing was achieved. Notably, the U depletion strategy<sup>17</sup> even increased bystander editing to 38% (Extended Data Fig. 3g). The 5'-G•U wobble also gave the best on-target efficiency of the compared bystander solutions with 47% yield, whereas a G•A mismatch reduced the yield and U depletion almost fully blocked editing, highlighting the power of the G•U wobble

strategy to achieve high efficiency and high precision in very A-rich triplet contexts where G•A mismatch and U depletion fail (Extended Data Fig. 3g and Supplementary Fig. 5a,b).

To show that this finding is highly generalizable, we performed a meta-analysis over three different target transcripts (*AHI1*, *BMP2* and *COL3A1*) representing three different A-rich target triplets (5'-[UAA]

U, 5'-U[AAG] and 5'-U[AAA]U) while using two different editing approaches, the CLUSTER guide RNA with endogenous ADAR and the boxB/ $\Delta$ N-ADAR system with engineered ADAR. Notably, not only bystander control but also on-target efficiency was significantly better with G•U wobble base pairs, demonstrating the strength of the strategy to suppress bystander editing precisely within A-rich triplets (Extended Data Fig. 3h and Supplementary Fig. 5).

**Wobble bases improve the engineering of CLUSTER guide RNAs** CLUSTER guide RNAs represent a recent strategy to harness endogenous ADAR for precise and efficient RNA base editing<sup>15</sup>. The basic concept combines an ADAR recruitment motif, a 20-nt SD that binds the target site and three or more additional RSs 15–20 nt in length that bind to the target mRNA over a larger stretch of sequence space in a multivalent fashion (Supplementary Fig. 3). In silico optimization of the guide RNA sequence is applied to choose RSs in such a way that highly editable A bases are avoided, enabling high control over bystander editing. Furthermore, guide RNAs with a high tendency to form inhibitory secondary structure are sorted out automatically, which helps to improve editing efficiency. Similar to the report for the LEAPER system<sup>17</sup>, we herein established the ribozyme-based Tornado expression system<sup>43</sup> for circularization and, thus, stabilization of guide RNAs (Extended Data Fig. 4). Starting with a simple LEAPER design, we could verify the formation of cleanly circularized guide RNAs (Supplementary Fig. 6), which gave an improved editing efficiency on the endogenous *RAB7A* transcript (Supplementary Fig. 7). Particularly notable was the positive effect of circularization on the editing yield after stable integration of the LEAPER guide RNA cassette by the PiggyBac transposase into the genome of HeLa cells, which gave moderate and stable editing yields over several weeks even with the weaker Pol2 promoter, elongation factor 1 $\alpha$  (EF1 $\alpha$ ) (Supplementary Fig. 8).

Compared to simple LEAPER guide RNAs, the flexible design of CLUSTER guide RNAs represented a considerable engineering challenge for circularization. The order of elements within the CLUSTER guide RNA, such as the SD, the RSs, the ADAR recruitment motif and the target mRNA exit points, can all be placed individually and relative to each other. On the guide RNA side, RSs can be placed 5' and/or 3' to the SD, while the ADAR recruitment motif can be flexibly placed anywhere in between. On the mRNA side, binding sites that correspond to the RS within the guide RNA can be located 5' and/or 3' of the TS, which corresponds to the SD within the guide RNA. Binding-site placement ultimately defines whether the exit points of the mRNA are close or distant relative to the ligation stem. Furthermore, the number and length of all antisense elements and linkers can be varied, which may individually affect the torsion within the guide RNA circle and, thus, its interaction with the target transcript.

To identify general design rules for circular CLUSTER guide RNAs, we tested various constructs on a luciferase reporter construct (Supplementary Figs. 9 and 10). We found that the mRNA exits should be placed approximately opposite of the SD and that the ADAR recruitment motif can be combined with the ligation stem into a larger RNA structure placed adjacent to the SD. Notably, a 5-nt bulge that separates ADAR recruitment motif and ligation stem achieved particularly high editing efficiency (Supplementary Fig. 10a–c, design L13). This ADAR recruitment motif was called the split-R/G motif and was used in subsequent designs.

After identifying general design principles for circular CLUSTER guide RNAs, we aimed to design an optimal guide RNA for a proof-of-concept in vivo study. Specifically, we aimed to target a premature STOP codon in the murine *Mecp2* transcript (W104>amber), which causes severe Rett syndrome-like symptoms in mice carrying this patient mutation<sup>44</sup>.

CLUSTER guide RNAs avoid bystander editing by choosing RSs that minimize the presence of editable A bases in the guide RNA–target RNA duplex. However, in highly A-rich target RNAs, such as *Mecp2*,

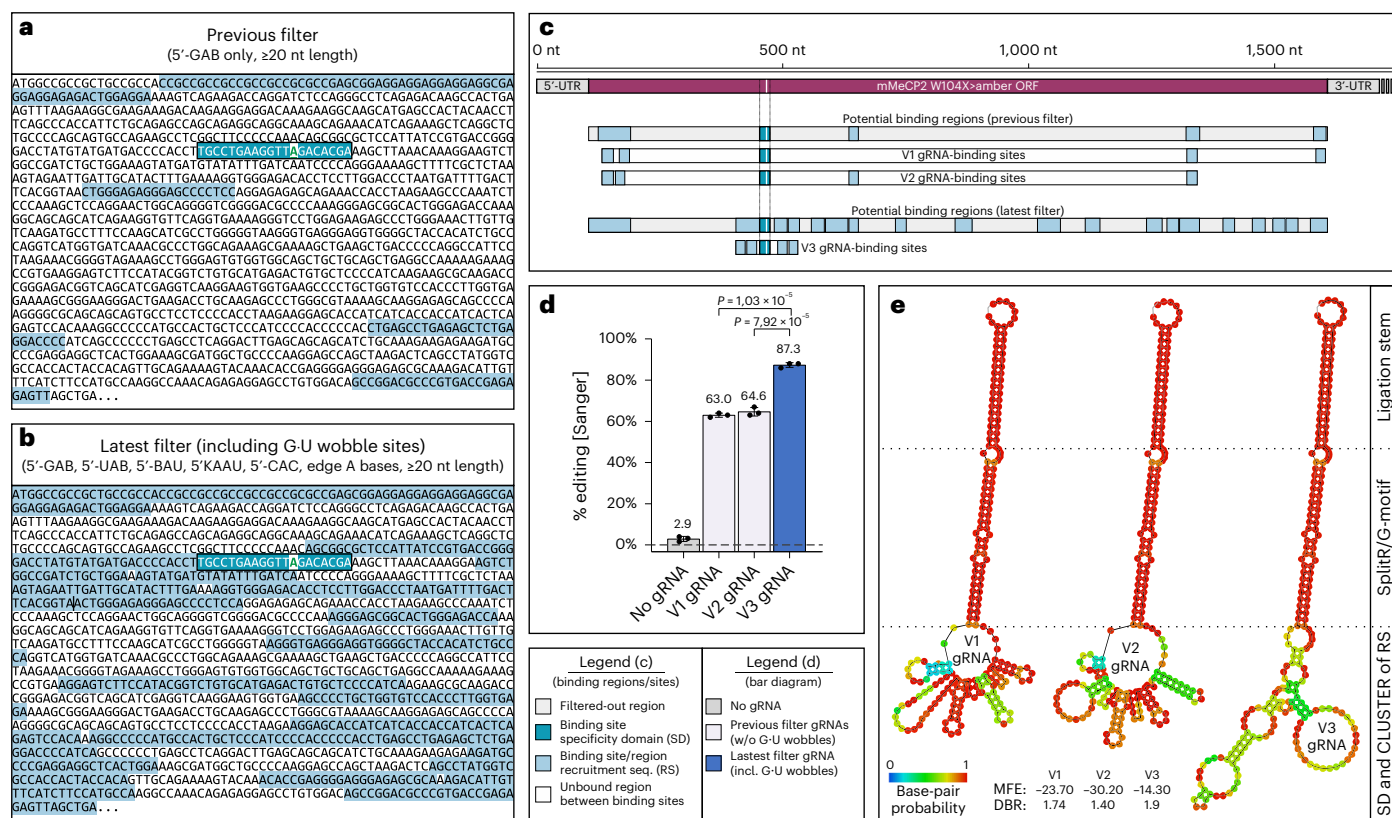
individual binding regions for RSs can be separated by long distances within the target transcript (for example, spread over several exons) and the available sequence space for secondary structure optimization can be limited. Both effects can negatively impact the editing efficiency of CLUSTER guide RNAs<sup>15</sup>. However, by applying the wobble base-pairing strategy to suppress bystander editing, the highly limiting filter set that defined eligible A bases within the RS-binding regions could now be considerably expanded; 5'-UAB, 5'-BAU triplets (B = C, G or U) and 5'-KAAU (K = G or U) sequence motifs are now included, while only 5'-GAB triplets were previously allowed<sup>15</sup>. Furthermore, 5'-CAC triplets and all A bases located at either end of a binding region (edge A bases), which we identified as being resistant to off-target editing, could be used to expand the sequence space. With the old filter settings (Fig. 4a), a circular CLUSTER guide RNA with four RSs and a split-R/G ADAR recruitment motif needs substantial space on the *Mecp2* target RNA—specifically, 1,470 nt for guide RNA V1 and 1,208 nt for guide RNA V2 (Fig. 4c). In contrast, with the new filter settings (Fig. 4b), the entire guide RNA (V3) covers only 127 nt on the murine *Mecp2* transcript (Fig. 4c). This guide RNA V3 gave significantly better on-target editing than V1 and V2 (87% versus 63% and 65%), which used the old filter settings (Fig. 4d). Partially, this might also be attributed to the larger available sequence space, which allows selecting guide RNAs with a much lower level of inhibitory secondary structure (Fig. 4e). As expected, the expansion of eligible A bases led to the appearance of bystander editing at binding sites of the RSs, which were albeit well suppressed with the G•U wobble strategy (Fig. 5).

Very recently, circular LEAPER guide RNAs were demonstrated to recruit endogenous ADARs with moderate editing efficiency and precision in cell culture and in vivo<sup>17,18</sup>. We benchmarked our best circular CLUSTER guide RNAs (Fig. 5a) for the *Mecp2* W104>amber transcript against 111-nt-long symmetric LEAPER guide RNAs (Fig. 5b) and assessed various means of bystander editing suppression (Fig. 5 and Extended Data Fig. 5). Linear (Extended Data Fig. 5) and circularized LEAPER guide RNAs (Fig. 5b,c), using the Tornado expression system were tested by transfection into HeLa cells. As expected, the circular LEAPER-based design gave massive bystander editing at >10 sites, with yields of 20–39% at seven such sites (Fig. 5c, circular LEAPER). Again we first compared G•A mismatching<sup>16</sup>, U depletion<sup>17</sup> and G•U wobbles to suppress bystander editing at the four bystander sites, which are amenable to G•U wobble base pairing. Given the low number of amenable sites (Fig. 5c) the effects on on-target editing efficiency and bystander editing were comparably low. However, when we aimed to suppress off-target editing at the major ten bystander sites, the on-target yield of a pure G•A mismatch (17%  $\pm$  2%) or pure U depletion (16%  $\pm$  1%) solution dropped considerably, while a combination of G•U wobble and G•A mismatch (25%  $\pm$  2%) outperformed the combination of G•U wobble and U depletion (18%  $\pm$  1%) for the best-performing circular LEAPER guide RNA (Fig. 5c). In contrast, a linear CLUSTER guide RNA applying the G•U wobble strategy already achieved bystander-free 38%  $\pm$  7% on-target editing (Extended Data Fig. 5f). Furthermore, all optimized circular CLUSTER guide RNAs gave an exceptionally good editing efficiency of >84% on target with very good precision (Fig. 5d), clearly outcompeting all tested LEAPER guide RNAs (Extended Data Fig. 5). G•U wobbles entirely suppressed bystander editing at all three bystander sites where RSs bound the *Mecp2* mRNA. It might be possible that G•A mismatches or U depletion would work similarly well to suppress bystander editing at such sites but this was not tested. There was one remaining site (5'-AAG) at position –5 with a moderate editing yield of 8% that was not amenable for wobble base pairing. Editing at this site would not change the amino acid sequence of MeCP2 yet its editing was reduced by U depletion or G•A mismatching (Fig. 5d).

### In vivo proof-of-concept in a murine Rett syndrome model

Before applying the optimized circular CLUSTER guide RNA in vivo, we verified its successful circularization in cell culture with two sets of





**Fig. 4 | Wobble base pairing improves the design of CLUSTER guide RNAs. a,** Results from an in silico search for RS-binding sites within the murine *Mecp2* W104-zamber ORF when applying the previous filter settings to the GuideRNA-Forge tool. The binding site for the SD is displayed in turquoise and outlined in black. It contains the target A in green and underlined. All potential RS-binding regions  $\geq 20$  nt length, in which binding sites for RSs can be selected, are displayed in light blue. The previous filter excludes all A bases except for those within a 5'-GAB (B = U, C or G) triplet context. **b,** Results for the same in silico search with the latest filter that included A bases in a 5'-GAB, 5'-UAB, 5'-BAU, 5'-KAAU and 5'-CAC triplet context, in addition to allowing A bases independent of their sequence context at the edges of a detected binding region. **c.** Illustration of

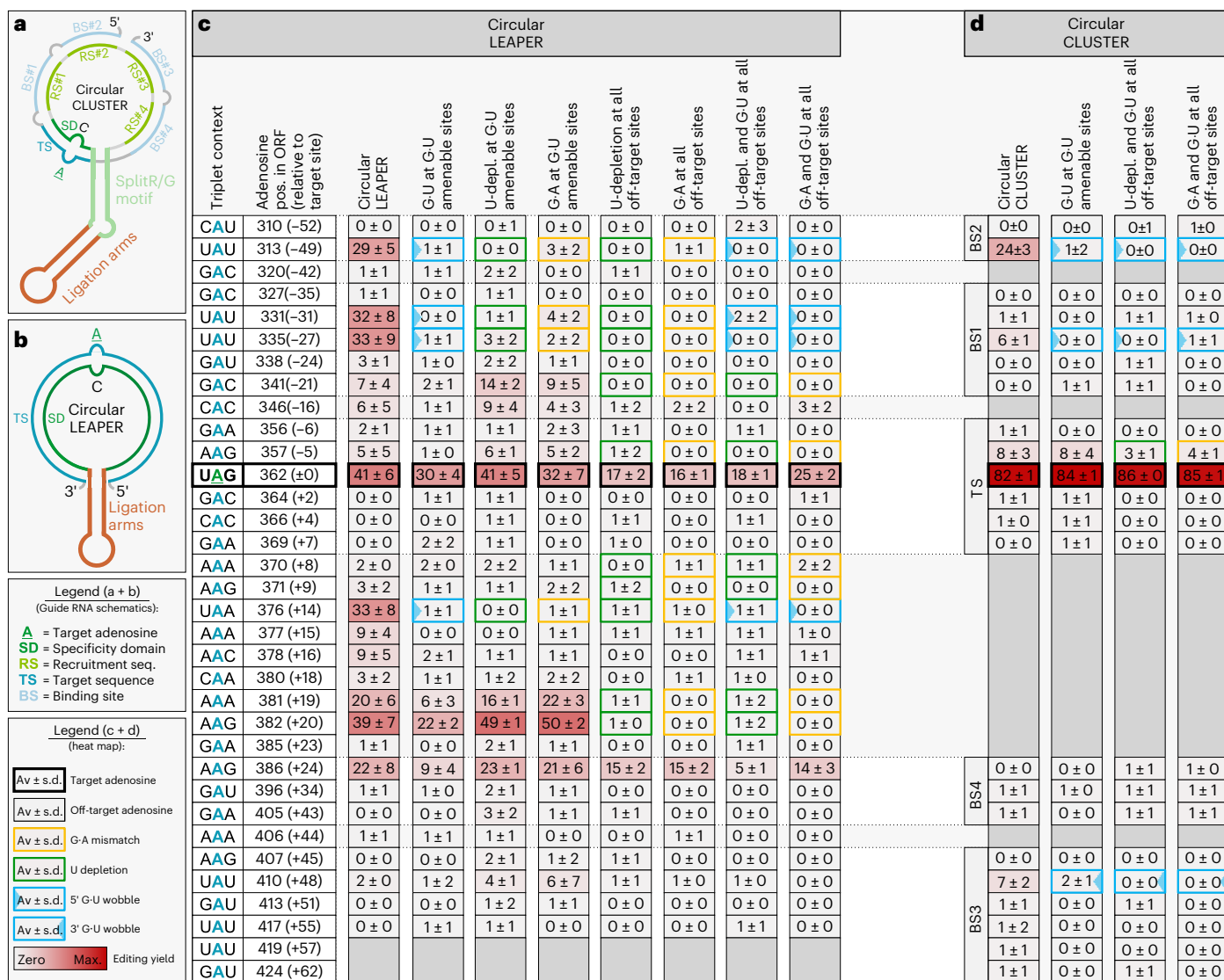
binding regions in which specific binding sites of RSs of three circular CLUSTER guide RNAs are located within the *Mecp2* ORF, generated using the previous filter (V1 and V2) or the latest filter (V3). **d**, Editing was performed with plasmid-borne guide RNA and murine *Mecp2* transcript in HeLa cells (endogenous ADAR). **e**, Secondary structure prediction of guide RNA V1–V3 generated using the ViennaRNA Package 2.0 (ref. 54). The mean free energy (MFE) of the antisense part (SD and CLUSTER of RS) is given in kcal per mol. The dot–bracket ratio (DBR) indicates the number of dots (unpaired bases) per bracket (paired bases) within the dot–bracket annotation of the antisense part. Data are shown as the mean editing percentage  $\pm$  s.d. of  $n = 3$  biological replicates. For statistical analysis, a Student's *t*-test (two-tailed, parametric) was applied.

reverse transcription (RT)–qPCR primer pairs (Supplementary Fig. 11). The use of an outward primer pair (Extended Data Fig. 6a) allowed us to verify circularization with high confidence (Extended Data Fig. 6b,c), while the use of an inward primer pair (Extended Data Fig. 6d) allowed us to quantify the strong effect (235-fold increase) of circularization on the total guide RNA abundance (Extended Data Fig. 6e), suggesting that the majority of guide RNAs are fully processed (Extended Data Fig. 6f). We chose the PHP.eB serotype for adeno-associated virus (AAV) encapsulation as it allows cargo delivery to the mouse brain after systemic administration<sup>45</sup>. Indeed, this serotype was successfully used by us before to deliver the boxB/ $\lambda$ -ADAR tool into the same Rett syndrome mouse model for mutation correction<sup>44</sup>. The targeting virus encoded the circular CLUSTER guide RNA as displayed in Fig. 5a,d (G•U at G•U-amenable sites). For the nontargeting virus control, the guide RNA's antisense parts were scrambled.

Mice were treated with  $4 \times 10^{12}$  viral genomes by retro-orbital injection and killed 4 weeks later; brain regions were analyzed separately for editing efficiency by Sanger sequencing. Editing levels differed among the seven brain regions, with clearly detectable editing in the midbrain, brainstem and thalamus, with an editing efficiency up to 19% (Fig. 6a). To better understand the key factors for successful editing, we analyzed all seven brain regions for the expression of the guide RNA (Fig. 6c), the AAV episome abundance (Fig. 6e) and the expression levels of all

catalytically active murine Adar isoforms: total Adar1 (Fig. 6g), Adar1 p150 (Fig. 6i) and Adar2 (Fig. 6k). The relative expression of guide RNA and Adars were directly compared through RT-qPCR by normalization to the geometric mean of the same three housekeeping genes *Actb*, *Rps29* and *Rnu6* (Supplementary Figs. 12 and 13), while the AAV episome abundance was determined as the number of copies per cell. Unexpectedly, editing yield correlated the least with the Adar levels (Fig. 6h,j,l;  $R^2 = 0.29-0.51$ ), even though Adar expression differed among brain regions, particularly for Adar2 (for example, thalamus versus other brain regions). This indicates that Adar abundance did not limit the editing outcome. The strongest correlation was found between guide RNA expression level and editing yield (Fig. 6d;  $R^2 = 0.87$ ), suggesting that, even under circularization, guide RNA levels still limit on-target editing. Editing also correlated well with AAV abundance (Fig. 6f;  $R^2 = 0.84$ ) and, in most brain regions, guide RNA expression seemed to also correlate well with AAV abundance. However, there were a few exceptions, such as in the cortex, where guide RNA expression was low even though AAV abundance was comparably high. This indicates, in agreement with our previous findings<sup>44</sup>, that the strength of the guide RNA promoter (for example, U6 promoter) or guide RNA stability differs among brain regions, thereby limiting editing even after successful viral delivery. In summary, these data show that the guide RNA's abundance (determined by its delivery, expression and stability) is the





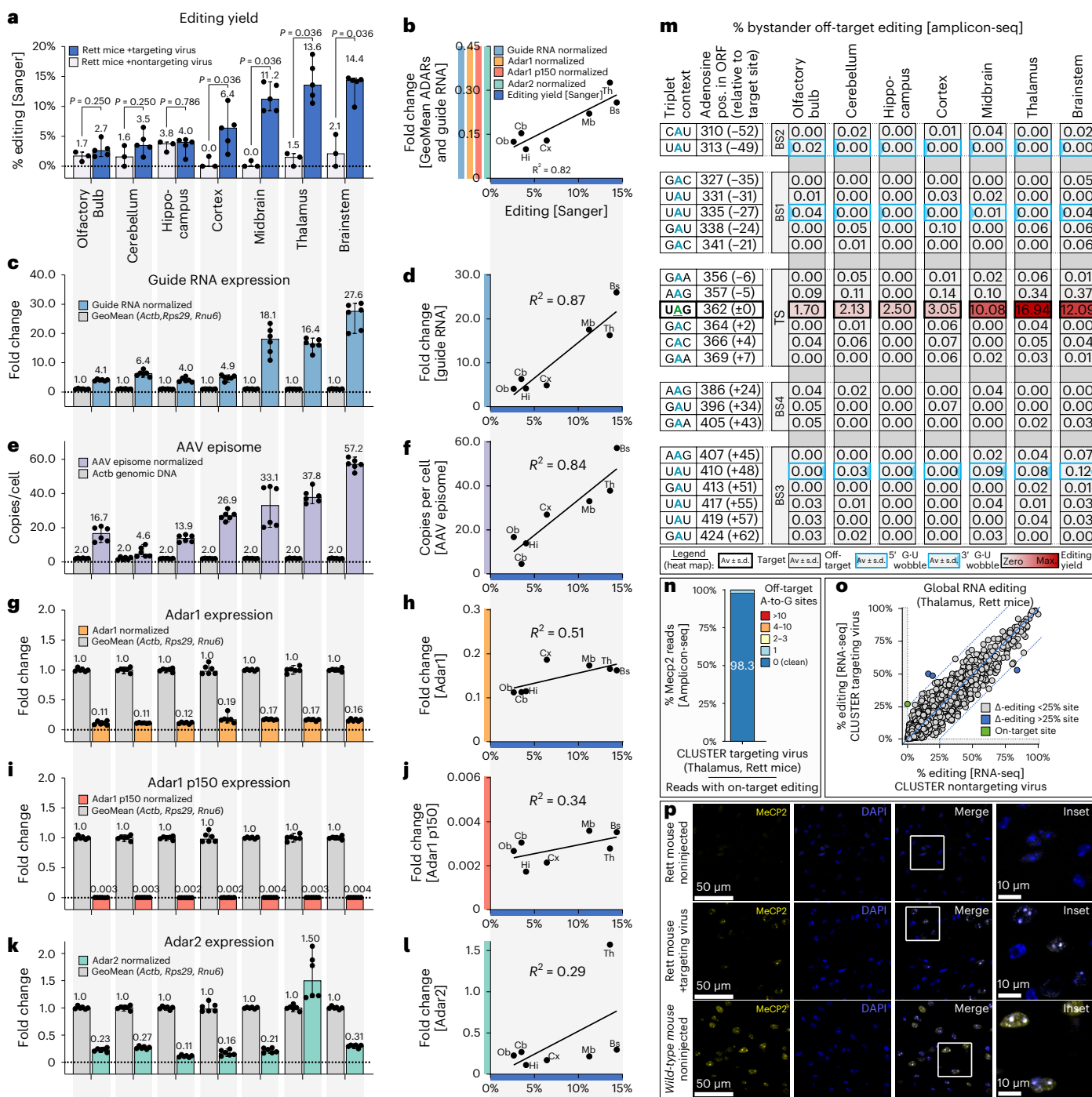
**Fig. 5 | Circular CLUSTER versus circular LEAPER *Mecp2* guide RNAs.** Editing heat map of guide RNA-binding sites within the murine *Mecp2* W104>amber transcript. The triplet context for each listed editing event is given with the target A highlighted in green and all off-target A bases in blue. The position of each site is given relative to the transcript and the target A (±0 position). Editing was performed with plasmid-borne guide RNA and murine *Mecp2* transcript in HeLa cells (endogenous ADAR). **a**, Circular CLUSTER guide RNA design. **b**, Circular LEAPER guide RNA design. **c**, Yields achieved using circular unstructured LEAPER guide RNAs containing a 111-nt-long antisense part. The basic design (column circular LEAPER) does not contain bystander solutions.

The other guide RNAs either contain G•A mismatches or apply U depletion at all bystander sites, G•U wobbles at G•U-amenable sites or a combination of G•A mismatching or U depletion with G•U wobbles at amenable sites. **d**, Yields achieved using circular CLUSTER guide RNAs containing a 100-nt-long antisense part split into a targeting sequence (20 nt) and four RSs (each 20 nt). The basic design (column circular CLUSTER) does not contain bystander solutions. The other guide RNAs contain G•U wobbles at positions -49, -27 and +48, as well as no bystander solution, G•A mismatch or depleted U at position -5. Data in **a, b** are shown as the mean editing percentage ± s.d. of *n* = 3 biological replicates.

most important factor to achieve high on-target editing in the brain, while Adar levels seem less important. This may instruct future designs of expression cassettes for CNS applications.

Particularly in a clinical setting, editing must be efficient and precise, represented by cleanly edited transcripts devoid of unintended recoding events. To evaluate bystander off-target events in the Rett mouse model, we performed deep amplicon sequencing (average read depth of 47,009 and average coverage of 45,991) of the *Mecp2* target transcript in all seven brain tissues. To ensure extensive detection of bystander events, we selected the two mice from the targeting virus group that had given the highest on-target editing yield in the thalamus. The on-target editing results matched very well with the Sanger sequencing (Fig. 6a,m and Extended Data Fig. 7). As controls,

two nontargeting virus-treated mice were used. They showed considerably lower background (~0.2%) compared to Sanger sequencing (~5%) (Fig. 6a,m and Extended Data Fig. 7). Consequently, this now enabled us to measure on-target editing yields of 1.7–3.0% with high confidence in the olfactory bulb, cerebellum, hippocampus and cortex (Fig. 6m). Notably, bystander editing was hardly detected in any brain region. At three positions (-49, -27 and +48), G•U wobble base pairs were applied to suppress bystander editing. At positions -49 and -27, this was very successful; no bystander editing was detectable. At position +48, there may have been up to 0.12% bystander editing in the brainstem but 100-fold below the on-target editing yield in that tissue. Only one bystander was detected with high confidence. This was the unresolved, silent bystander site at position -5 that was already discovered in cell



**Fig. 6 | Transcript repair in a mouse model of Rett syndrome using circular wobble-optimized CLUSTER guide RNAs and endogenous murine Adars.**

**a**, Editing yields in different brain regions, after delivery of the AAV-PHP. **e**, eB-encoded guide RNA through retro-orbital injection into Rett mice carrying the *MeCP2* W104>amber mutation and quantification by Sanger sequencing 4 weeks later. **b**, Correlation between the median editing in **a** and the geometric mean of the RT-qPCR targets in **c**, **g**, **i**, **k**. Olfactory bulb, Ob; cerebellum, Cb; hippocampus, Hi; cortex, Cx; midbrain, Mb; thalamus, Th; brainstem, Bs. **c**, Guide RNA expression quantified by RT-qPCR (normalized to the geometric mean of *Actb*, *Rps29* and *Rnu6*). **d**, As in **b** but correlating **a**, **c**, **e**, Absolute quantification of AAV episomes per cell by standard curve qPCR (normalized to *Actb*). **f**, As in **b** but correlating **a**, **e**, **g**. As in **c** but for Adar1. **h**, As in **b** but correlating **a**, **g**, **i**. As in **c** but for Adar1 p150. **j**, As in **b** but correlating **a**, **i**, **k**. As in **c** but for Adar2. **l**, As in

**b** but correlating **a**, **k**, **m**. Bystander off-target events at the guide RNA-binding sites and the TS. **n**, All amplicon reads with on-target editing binned according to their number of bystander events (based on Supplementary Fig. 14). **o**, Global RNA editing at 2,533 endogenous sites (coverage  $\geq 50$  reads, REDIPortal<sup>55</sup>). **p**, Thalamus sections stained for MeCP2 and DAPI (nuclei). For **a**, data are shown as the median editing percentage  $\pm$  95% CI determined in  $n = 3$  mice (nontargeting virus) and  $n = 5$  mice (targeting virus). For statistical analysis, a Mann-Whitney *U*-test (two-tailed, nonparametric) was applied. For **c**, **e**, **g**, **i**, **k**, **m**, **n**, data are shown as the median fold change  $\pm$  95% CI or median number of copies per cell  $\pm$  95% CI determined in  $n = 2$  mice (three technical replicates each). The NGS analysis in **o** is based on results from  $n = 1$  (nontargeting virus) or  $n = 2$  (targeting virus) mice. The results in **p** are derived from  $n = 1$  mouse per group. For **b**, **d**, **f**, **h**, **j**, **l**, the  $R^2$  values were determined by simple linear regression.

culture (Fig. 5d;  $8\% \pm 4\%$ ). Here we detected up to 0.37% bystander editing in vivo in the brainstem (Fig. 6m). If they appear at many sites, even bystander events with low editing yield may sum up to interfere with on-target editing. To address this potential issue, we studied how often an on-target edited read is damaged by an additional bystander edit. We found that 98.3% of the ~10,000 detected on-target edited reads were completely bystander free (Fig. 6n and Supplementary Fig. 14), highlighting the impressive degree of editing precision achieved in the in vivo proof of concept.

Next, we evaluated the global editing precision by interrogating transcriptome-wide RNA sequencing (RNA-seq) data collected from the thalamus of targeting and nontargeting virus-treated Rett mice (Fig. 6o and Extended Data Fig. 8a). As before, we selected the two mice from the targeting virus group that gave the highest on-target editing yield in the thalamus with Sanger sequencing. First, we applied the RNA-editing index method that monitors changes in global RNA-editing levels in a highly unbiased manner and focused the analysis on the particularly critical coding sequence space<sup>46</sup>. The A-to-G RNA-editing indices were nearly identical in both groups (Extended Data Fig. 8a), indicating that the global editing activity was overall not affected by the presence of the targeting guide RNA. Second, we tried to detect differentially edited sites between the two conditions (Fig. 6o), similar to a previous study<sup>15</sup>. However, compared to the same analysis performed with untreated Rett mice (Extended Data Fig. 8b), we detected no clear off-target sites. Only three sites fell slightly outside of the  $\pm 25\%$   $\Delta$ -editing margin but this was likely because of normal variability among mice. None of the three sites were located in the coding region and none of them were complementary to the guide RNA. Third, we tried to detect off-target events in a candidate approach and searched in silico throughout the whole murine genome for transcripts with 20-nt similarity (with up to one mismatch) to the guide RNA-binding regions. We identified 65 potential sites; however, only four of these sites were sufficiently expressed ( $\geq 20$  reads of coverage) to be evaluated and no off-target editing was detected at any of them. Overall, we were unable to identify any global off-target events, excluding mouse-to-mouse variability, which suggests a very high precision of our approach on the transcriptome-wide level.

RNA base-editing approaches that apply the overexpression of an engineered editase, such as  $\lambda$ N-ADAR2, typically suffer from substantial, editase-dependent global off-target events<sup>46–48</sup>. In our recent study—from which the noninjected mice data were derived (Extended Data Fig. 8a), we also detected numerous off-target events even though the native ADAR2 deaminase domain was fused to the  $\lambda$ N peptide. We revisited this published RNA-seq data<sup>44</sup> to compare with the CLUSTER generated data. The boxB/ $\lambda$ N-ADAR2 data were generated using the same experimental setup (mouse model, AAV serotype and application route) as in this CLUSTER study. We ran the data in parallel through the same next-generation sequencing (NGS) analysis pipelines. First, we evaluated the same 2,533 endogenous RNA-editing sites (with coverage  $\geq 50$  reads) that we evaluated for the CLUSTER approach, as shown in Fig. 6o. Again, we selected the tissue with the highest on-target editing yield, which was the brainstem in this study<sup>44</sup>. We found a notable number (25) of potential off-target sites exhibiting a  $\Delta$ -editing margin above 25% (Extended Data Fig. 8c). Importantly, the identified events included two evolutionary highly conserved, Adar2-specific editing sites in glutamate metabotropic receptor 4 (GRM4) and neuro-oncological ventral antigen 1 (NOVA1)<sup>49</sup>, for which the change in editing level could have functional impact. Second, we analyzed the editing index in the coding sequence space and found a considerable increase in the index in the presence of the  $\lambda$ N-ADAR2 deaminase (Extended Data Fig. 8c). Together, this shows, in accordance with the literature<sup>47</sup>, that the harnessing of endogenous ADAR is more precise on the transcriptome-wide level than the ectopic expression of engineered ADAR effectors, whether hyperactive or native.

Lastly, we studied the restoration of MeCP2 protein expression and function upon treatment using the circular CLUSTER guide RNA (targeting virus). In the noninjected Rett mice, MeCP2 expression was not detectable by single-cell immunohistochemistry<sup>44</sup> in the thalamus, as the mutation resulted in an unstable protein (Fig. 6p and Extended Data Fig. 9). In clear contrast, MeCP2 was restored in  $\sim 33.3\% \pm 4\%$  (median  $\pm 95\%$  confidence interval (CI)) of the cells in the thalamus of the treated brain. Importantly, the restored MeCP2 protein was localized in heterochromatic foci inside the nucleus, which is an accepted proxy for its in vivo binding ability to methylated DNA<sup>50,51</sup> and very similar to the positive control mouse where up to 100% of the cells showed MeCP2 protein associated with heterochromatin (Fig. 6p and Extended Data Fig. 9a,b). The abundance of foci increased with increasing editing yields over the evaluated brain tissues (Extended Data Fig. 9c). The results suggest that MeCP2 protein and its function are restored in cells where guide RNA is delivered and expressed.

## Conclusions and outlook

The strategic placement of G•U wobble base pairs is a novel approach to improve editing precision and is widely applicable to RNA base-editing tools, including the LEAPER and CLUSTER approaches,  $\lambda$ N-ADAR, Cas13-ADAR and chemically modified ASOs. The 5'-G•U wobble is particularly powerful to suppress editing at all four 5'-UAN triplets, which account for the largest burden of bystander editing. The suppressive effect of the G•U wobble is very strong and, at the same time, restricted to the direct 5' and 3' neighboring bases. In this regard, the G•U wobble differs from alternative solutions such as G•A mismatch or U depletion and makes the wobble strategy particularly powerful to suppress bystander editing close to an on-target site. As not all bystander-prone triplets are amenable to the G•U wobble strategy, we suggest to combine them with other strategies of bystander suppression, particularly with G•A mismatches at sites not amenable for the wobble strategy. Typically, but not always, a combination of wobble base pairing with G•A mismatching resulted in a better balance of editing efficiency over editing precision than the use of G•A mismatching only.

It is intriguing to speculate about the mechanism that underlies the suppressive effect of wobble base pairs. The ADAR deaminase domain strongly favors 5'-UAN triplets over 5'-CAN and 5'-GAN. A recent crystal structure analysis of the ADAR2 deaminase in complex with a dsRNA substrate shed light on the molecular basis of the nearest neighbor preference<sup>52</sup> and identified a specific steric clash between the backbone of G489 and the exocyclic amino group of the guanine of the 5' neighboring G=C base pair (Extended Data Fig. 10a–c). Thus, replacing a 5'-A=U base pair with a 5'-G•U wobble base pair would again introduce a sterically demanding exocyclic amino group in the minor groove of the dsRNA substrate. Furthermore, the guanine base is likely further shifted toward the minor groove given the noncanonical hydrogen bond pattern in a G•U wobble base pair<sup>32</sup>. This might explain the very strong and located suppressive effect of the 5' G•U wobble base pair (Extended Data Fig. 10a–c). A similar clash of the exocyclic amino group with S486 in the minor groove might explain the suppressive effect of the G•U wobble base pair at the 3' nearest neighbor position (Extended Data Fig. 10d–f).

Overall, wobble base pairs create structural perturbations within the RNA substrate that have consequences for the activity of human ADAR. Moreover, regarding natural substrates, wobble base pairs seem to be part of a structural layer of control that may not yet be fully recognized. Nevertheless, we could deduce simple rules for the rational design of *trans*-acting guide RNAs for targeted RNA base editing, which achieved very high editing efficiency with very good editing precision. While broadly applicable in many approaches, the G•U wobble strategy proved helpful to design CLUSTER guide RNAs by means of expanding the sequence space for the computation of in silico optimized circular CLUSTER guide RNAs. In a direct comparison with LEAPER guide RNAs



of similar length, we could show that CLUSTER guide RNAs can be designed with clearly advantageous efficiency and precision. It remains open whether G•A mismatching or U depletion could serve similarly well to optimize CLUSTER guide RNAs. However, given the short length of an individual RS (15–20 nt), we speculate that wobble base pairs might work better in highly A-rich binding regions where multiple G•A mismatches or bulged A bases might interfere with RS binding.

The AAV-mediated delivery of the most advanced circular CLUSTER guide RNA demonstrated effective (up to 19%) transcript repair of the *Mecp2* W104>amber mutation in a mouse model of Rett syndrome and represents a successful in vivo recruitment of endogenous Adars for site-directed RNA editing in the CNS. With a median editing yield of ~14.4%, it also represents a high in vivo editing yield for restoring a disease-causing mutation in a murine model of human disease using only endogenous Adars. Furthermore, deep amplicon sequencing showed that on-target editing was achieved with very high precision regarding bystander editing. More than 98% of the detected on-target-edited reads were free from bystander edits. Only one (silent) bystander edit was detected with high confidence, albeit with a yield of only 0.1–0.37%. Transcriptome-wide RNA-seq confirmed a stable A-to-G RNA-editing index and no off-target sites in the transcriptome could be identified. This shows that editing can be achieved with very high precision and under the control of bystander editing in vivo. Notably, it is highly promising to see that endogenous Adar can be harnessed to obtain similar editing levels to those achieved with the exogenous boxB/AN-ADAR tool in the same mouse model before<sup>44</sup>. Interestingly, the trends regarding the varying editing efficiencies in different brain regions were also comparable between endogenous and exogenous ADAR approaches. A side-by-side analysis of on-target editing, AAV episome abundance as well as guide RNA and Adar isoform expression data<sup>53</sup> suggests that not endogenous Adar expression but rather the virus-mediated guide RNA expression limited editing efficiency in the CNS of the murine Rett model. Thus, future improvements of AAV capsid engineering, application routes, guide RNA expression cassettes and/or guide RNA stability promise to enhance RNA base editing even further, breaking ground for clinical applications. Overall, harnessing endogenous ADAR with permanent AAV-driven CLUSTER guide RNAs in the CNS is an important next step toward the development of novel drug modalities that fight neurological diseases.

## Online content

Any methods, additional references, Nature Portfolio reporting summaries, source data, extended data, supplementary information, acknowledgements, peer review information; details of author contributions and competing interests; and statements of data and code availability are available at <https://doi.org/10.1038/s41587-024-02313-0>.

## References

- Diaz Quiroz, J. F., Siskel, L. D. & Rosenthal, J. J. C. Site-directed A → I RNA editing as a therapeutic tool: moving beyond genetic mutations. *RNA* **29**, 498–505 (2023).
- Rees, H. A. & Liu, D. R. Base editing: precision chemistry on the genome and transcriptome of living cells. *Nat. Rev. Genet.* **19**, 770–788 (2018).
- Nishikura, K. A-to-I editing of coding and non-coding RNAs by ADARs. *Nat. Rev. Mol. Cell Biol.* **17**, 83–96 (2016).
- Bass, B. L. RNA editing by adenosine deaminases that act on RNA. *Annu. Rev. Biochem.* **71**, 817–846 (2002).
- Stafforst, T. & Schneider, M. F. An RNA-deaminase conjugate selectively repairs point mutations. *Angew. Chem. Int. Ed. Engl.* **51**, 11166–11169 (2012).
- Vogel, P. et al. Efficient and precise editing of endogenous transcripts with SNAP-tagged ADARs. *Nat. Methods* **15**, 535–538 (2018).
- Montiel-Gonzalez, M. F. et al. Correction of mutations within the cystic fibrosis transmembrane conductance regulator by site-directed RNA editing. *Proc. Natl Acad. Sci. USA* **110**, 18285–18290 (2013).
- Montiel-Gonzalez, M. F., Vallecillo-Viejo, I. C. & Rosenthal, J. J. An efficient system for selectively altering genetic information within mRNAs. *Nucleic Acids Res.* **44**, e157 (2016).
- Vallecillo-Viejo, I. C. et al. Abundant off-target edits from site-directed RNA editing can be reduced by nuclear localization of the editing enzyme. *RNA Biol.* **15**, 104–114 (2018).
- Cox, D. B. T. et al. RNA editing with CRISPR-Cas13. *Science* **358**, 1019–1027 (2017).
- Abudayyeh, O. O. et al. A cytosine deaminase for programmable single-base RNA editing. *Science* **365**, 382–386 (2019).
- Kannan, S. et al. Compact RNA editors with small Cas13 proteins. *Nat. Biotechnol.* **40**, 194–197 (2022).
- Merkle, T. et al. Precise RNA editing by recruiting endogenous ADARs with antisense oligonucleotides. *Nat. Biotechnol.* **37**, 133–138 (2019).
- Monian, P. et al. Endogenous ADAR-mediated RNA editing in non-human primates using stereopure chemically modified oligonucleotides. *Nat. Biotechnol.* **40**, 1093–1102 (2022).
- Reautschnig, P. et al. CLUSTER guide RNAs enable precise and efficient RNA editing with endogenous ADAR enzymes in vivo. *Nat. Biotechnol.* **40**, 759–768 (2022).
- Qu, L. et al. Programmable RNA editing by recruiting endogenous ADAR using engineered RNAs. *Nat. Biotechnol.* **37**, 1059–1069 (2019).
- 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).
- Katrekar, D. et al. Efficient in vitro and in vivo RNA editing via recruitment of endogenous ADARs using circular guide RNAs. *Nat. Biotechnol.* **40**, 938–945 (2022).
- Eggington, J. M., Greene, T. & Bass, B. L. Predicting sites of ADAR editing in double-stranded RNA. *Nat. Commun.* **2**, 319 (2011).
- Licht, K. et al. Inosine induces context-dependent recoding and translational stalling. *Nucleic Acids Res.* **47**, 3–14 (2019).
- Wong, S. K., Sato, S. & Lazinski, D. W. Substrate recognition by ADAR1 and ADAR2. *RNA* **7**, 846–858 (2001).
- Higuchi, M. et al. RNA editing of AMPA receptor subunit GluR-B: a base-paired intron–exon structure determines position and efficiency. *Cell* **75**, 1361–1370 (1993).
- Huang, H. et al. Tissue-selective restriction of RNA editing of CaV1.3 by splicing factor SRSF9. *Nucleic Acids Res.* **46**, 7323–7338 (2018).
- Hanrahan, C. J. et al. RNA editing of the *Drosophila para* Na<sup>+</sup> channel transcript. Evolutionary conservation and developmental regulation. *Genetics* **155**, 1149–1160 (2000).
- Reenan, R. A., Hanrahan, C. J. & Ganetzky, B. The *mle* RNA helicase mutation in *Drosophila* results in a splicing catastrophe of the *para* Na<sup>+</sup> channel transcript in a region of RNA editing. *Neuron* **25**, 139–149 (2000).
- Ingleby, L. et al. Regulated RNA editing and functional epistasis in Shaker potassium channels. *J. Gen. Physiol.* **133**, 17–27 (2009).
- Reenan, R. A. Molecular determinants and guided evolution of species-specific RNA editing. *Nature* **434**, 409–413 (2005).
- Ramaswami, G. et al. Genetic mapping uncovers cis-regulatory landscape of RNA editing. *Nat. Commun.* **6**, 8194 (2015).
- Schneider, M. F. et al. Optimal guideRNAs for re-directing deaminase activity of hADAR1 and hADAR2 in trans. *Nucleic Acids Res.* **42**, e87 (2014).
- Uzonyi, A. et al. Deciphering the principles of the RNA editing code via large-scale systematic probing. *Mol. Cell* **81**, 2374–2387 (2021).



31. Liu, X. et al. Learning *cis*-regulatory principles of ADAR-based RNA editing from CRISPR-mediated mutagenesis. *Nat. Commun.* **12**, 2165 (2021).
32. Varani, G. & McClain, W. H. The G-U wobble base pair. A fundamental building block of RNA structure crucial to RNA function in diverse biological systems. *EMBO Rep.* **1**, 18–23 (2000).
33. Xu, D. et al. The electrostatic characteristics of G-U wobble base pairs. *Nucleic Acids Res.* **35**, 3836–3847 (2007).
34. Stefl, R. et al. Structure and specific RNA binding of ADAR2 double-stranded RNA binding motifs. *Structure* **14**, 345–355 (2006).
35. Chahrouh, M. et al. MeCP2, a key contributor to neurological disease, activates and represses transcription. *Science* **320**, 1224–1229 (2008).
36. Heep, M. et al. Applying human ADAR1p110 and ADAR1p150 for site-directed RNA editing—G/C substitution stabilizes guideRNAs against editing. *Genes (Basel)* **8**, 34 (2017).
37. Wettengel, J. et al. Harnessing human ADAR2 for RNA repair—recoding a *PINK1* mutation rescues mitophagy. *Nucleic Acids Res.* **45**, 2797–2808 (2017).
38. Dixon-Salazar, T. et al. Mutations in the *AHI1* gene, encoding Joubertin, cause Joubert syndrome with cortical polymicrogyria. *Am. J. Hum. Genet.* **75**, 979–987 (2004).
39. Smith, L. T. et al. Mutations in the *COL3A1* gene result in the Ehlers-Danlos syndrome type IV and alterations in the size and distribution of the major collagen fibrils of the dermis. *J. Invest. Dermatol.* **108**, 241–247 (1997).
40. Roberts, K. E. et al. *BMPR2* mutations in pulmonary arterial hypertension with congenital heart disease. *Eur. Respir. J.* **24**, 371–374 (2004).
41. Davis, A. R. & Znosko, B. M. Thermodynamic characterization of single mismatches found in naturally occurring RNA. *Biochemistry* **46**, 13425–13436 (2007).
42. Gartner, J. et al. Mutations in *PEX1* in peroxisome biogenesis disorders: G843D and a mild clinical phenotype. *J. Inher. Metab. Dis.* **22**, 311–313 (1999).
43. Litke, J. L. & Jaffrey, S. R. Highly efficient expression of circular RNA aptamers in cells using autocatalytic transcripts. *Nat. Biotechnol.* **37**, 667–675 (2019).
44. Sinnamon, J. R. et al. Targeted RNA editing in brainstem alleviates respiratory dysfunction in a mouse model of Rett syndrome. *Proc. Natl Acad. Sci. USA* **119**, e2206053119 (2022).
45. Chan, K. Y. et al. Engineered AAVs for efficient noninvasive gene delivery to the central and peripheral nervous systems. *Nat. Neurosci.* **20**, 1172–1179 (2017).
46. Roth, S. H., Levanon, E. Y. & Eisenberg, E. Genome-wide quantification of ADAR adenosine-to-inosine RNA editing activity. *Nat. Methods* **16**, 1131–1138 (2019).
47. Pfeiffer, L. S. & Stafforst, T. Precision RNA base editing with engineered and endogenous effectors. *Nat. Biotechnol.* **41**, 1526–1542 (2023).
48. Buchumenski, I. et al. Global quantification exposes abundant low-level off-target activity by base editors. *Genome Res.* **31**, 2354–2361 (2021).
49. Pinto, Y., Cohen, H. Y. & Levanon, E. Y. Mammalian conserved ADAR targets comprise only a small fragment of the human editosome. *Genome Biol.* **15**, R5 (2014).
50. Heckman, L. D., Chahrouh, M. H. & Zoghbi, H. Y. Rett-causing mutations reveal two domains critical for MeCP2 function and for toxicity in MECP2 duplication syndrome mice. *eLife* **3**, e02676 (2014).
51. Brown, K. et al. The molecular basis of variable phenotypic severity among common missense mutations causing Rett syndrome. *Hum. Mol. Genet.* **25**, 558–570 (2016).
52. Matthews, M. M. et al. Structures of human ADAR2 bound to dsRNA reveal base-flipping mechanism and basis for site selectivity. *Nat. Struct. Mol. Biol.* **23**, 426–433 (2016).
53. Lein, E. S. et al. Genome-wide atlas of gene expression in the adult mouse brain. *Nature* **445**, 168–176 (2007).
54. Lorenz, R. et al. ViennaRNA Package 2.0. *Algorithms Mol. Biol.* **6**, 26 (2011).
55. Picardi, E. et al. REDportal: a comprehensive database of A-to-I RNA editing events in humans. *Nucleic Acids Res.* **45**, D750–D757 (2017).

**Publisher's note** Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

**Open Access** This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

© The Author(s) 2024

## Methods

### The GuideRNA-Forge tool

The GuideRNA-Forge tool was programmed in Python (version 3.9). It is based on our previously published ‘recruitment cluster finder’ tool and, thus, follows the core principles summarized previously<sup>15</sup>. GuideRNA-Forge uses a JavaScript Object Notation file (.json) to allow for the modular input of guide RNA parts and, thus, makes the power of in silico optimization available to any conceivable multivalent design, including circular guide RNAs. An example .json file including all relevant input values or strings for the generation of a circular CLUSTER guide RNA is provided in the GitHub archive.

The filters that identify binding regions for RS placement, while excluding RSs that contain unwanted sequences (for example, immunogenic, cytotoxic or guide RNA-destabilizing motifs) are now customizable. A combination of regular expressions (regex) and modular filter building blocks allows for quick assembly of user defined filters. The filters used in this study to generate G•U wobble-containing circular CLUSTER guide RNAs are provided in the GitHub archive. Note that the filters do not apply G•U wobbles to your guide RNA. Instead, they allow binding regions that contain A bases in a 5′-GAB, 5′-UAB, 5′-BAU (B = C, G or U), 5′-KAAU (K = G or U) and 5′-CAC sequence context, as well as A bases at the edges of binding regions. As not all A bases that allow for the G•U wobble solution (5′-UAB, 5′-BAU and 5′-KAAU) are necessarily edited, we recommend determining actual bystander off-target sites experimentally. G•U wobbles can then be strategically placed adjacent to problematic A bases.

In contrast to the previous version, the GuideRNA-Forge tool can identify multiple RSs within one potential binding region, as well as search in both the 5′ and the 3′ directions starting from the SD-binding site. Furthermore, the processing speed of recombination and folding (ViennaRNA package<sup>54</sup>) was drastically increased by implementing multithreading.

The resulting guide RNAs are scored for proximity of their binding sites and minimal secondary structure. CLUSTER guide RNAs with a high median score consist of RSs that bind in close proximity to each other on the target transcript and are less likely to engage in unproductive folding of their antisense part or misfolding of the ADAR-recruiting domain.

### Vector design (guide RNA and cDNA)

The guide RNA and *cis*-acting reporter inserts including the necessary overhangs were created by hybridization and phosphorylation of oligonucleotides. The *cis*-acting editing reporters were created by cloning hybrid oligonucleotide inserts into our pcDNA3.1-based *eGFP* expression plasmid (pTS58) using Apal and AgeI as described in Supplementary Note 1. Linear CLUSTER (using HindIII and BbsI into pTS1033), LEAPER (using HindIII and BamHI into pTS1033) and boxB (using HindIII and BamHI into pTS1033) guide RNAs were created as previously described<sup>15</sup>. DR guide RNAs were created by cloning hybrid oligonucleotide inserts encoding the antisense part into the PspCas13b crRNA backbone (Addgene plasmid 103854) using BbsI. The DR hairpin is already present in the backbone and must not be a part of the insert.

As the first step to create circular CLUSTER or LEAPER guide RNAs expressed by the U6 promoter, a modified Tornado expression cassette was created by gene synthesis (Thermo Fisher, GeneArt gene synthesis) and cloned into the backbone of our in-house editing vector ‘pEdit1.2’ (ref. 37) using NheI and AgeI, thereby replacing the ADAR2 expression cassette. The resulting Tornado OHA vector U6 (pTS1541) was used as the backbone for subsequent circular guide RNA cloning. The hybrid oligonucleotide guide RNA inserts were cloned into the latter backbone using BbsI. The cloning strategy is described in Supplementary Note 2. As a first step for cloning of circular LEAPER guide RNAs with flexible poly(AC) RNA linkers (AC50), the latter were synthesized as a gene block and cloned into pTS1541 using BbsI, resulting in the Circular LEAPER AC50 cloning vector (pTS2508). Subsequent cloning of gene blocks containing circular LEAPER guide RNAs destined for the AC50

context was performed using BbsI into pTS2508. For the expression of circular guide RNAs by the EF1α promoter, the Tornado OHA vector EF1α (pTS1593) was used for subsequent circular guide RNA cloning analogous to pTS1541. For direct benchmarks between EF1α and U6 constructs, the hybrid oligonucleotide guide RNA inserts were cloned into pTS1593 (EF1α) and pTS1790 (U6) using BbsI. For stable genomic integration of linear or circular guide RNAs by the PiggyBac system, a transposase expression plasmid and the plasmid pTS1896 (transposon, U6, circular LEAPER guide RNA *RAB7A*), pTS1897 (transposon, U6, linear LEAPER guide RNA *RAB7A*), pTS1899 (transposon, EF1α, circular LEAPER guide RNA *RAB7A*) or pTS1900 (Transposon, EF1α, linear LEAPER guide RNA *RAB7A*) were used.

For AAV production, the lead circular CLUSTER guide RNA targeting the *Mecp2* W104>amber transcript was cloned into the pAAV-GFP backbone (Cell Biolabs). The cloning strategy is described in Supplementary Note 3.

The open reading frame (ORF) sequence of murine *Mecp2* W104>amber was supplied by the Mandel Lab (Vollum Institute, Oregon Health and Science University) and cloned into the pEGFP-N3 backbone (Clontech) using EcoRI and KpnI under the control of a cytomegalovirus (CMV) promoter and terminated by an SV40 poly(A) signal. The human *PEX1*<sup>G843D</sup> plasmid was supplied by the Dodt Lab (Interfaculty Institute of Biochemistry and Cell Biochemistry, University of Tübingen). The cassette is under the control of an EF1α promoter and terminated by an EF1α poly(A) signal. The dual-luciferase reporters and the disease-relevant cDNA constructs *AH1* W725>amber, *BMPR2* W298>amber and *COL3A1* W1278>amber were created as previously described<sup>15</sup>. The sequences of all cloned products were verified by Sanger sequencing. The ORFs and amino acid sequences of *Mecp2*, *PEX1* and the engineered editases are given in Supplementary Note 4. Plasmid maps of all mentioned pTS plasmids are given in Supplementary Note 5 or can be found in the literature<sup>37</sup>. Further details can be found in the Supplementary Information.

### Analysis of RNA editing

A-to-I editing yields were quantified from Sanger sequence traces using SNAP-Gene (version 4.2.11). The relative height of the signal of G was compared to the sum of G+A, as described earlier<sup>37</sup>. If a reverse primer was used for sequencing, C and T peaks were treated accordingly. For better comparability, on-target editing yields of the same target with different guide RNAs were quantified using the same sequencing primer. Off-target editing had to be evaluated with different sequencing primers in most cases because of large distances between the guide RNA-binding sites. Only the cleanest reads were used for off-target evaluations, whereby G peaks below background were counted as 0% off target. Editing events with yields below 10% were background-corrected with the negative control.

### Structural analysis

A-U base pair structures and ADAR amino acid residue structures were obtained from the substrate-bound ADAR2 dimer crystal structure (Protein Data Bank (PDB) 5HP2)<sup>56</sup>, while G•U and U•G wobble base pairs were obtained from the NMR structure of the Gria2 stem loop RNA (PDB 2L2J)<sup>57</sup>. Structures were fitted and imaged with the open-source PyMOL Molecular Graphics System (version 2.5.0)<sup>58</sup>. The results are reported in Extended Data Fig. 10.

### Cell culture

**General.** The used cell cultures were grown in DMEM (Thermo Fisher, 41965062) supplemented with 10% FBS (Thermo Fisher, 10270106) and kept in an incubator at 37 °C and 5% CO<sub>2</sub>.

**Cell line choice.** In our laboratory, HeLa cells have undergone comprehensive characterization, including assessment of ADAR1 expression, knockdown efficiency, response to interferon treatment and

optimization of transfection conditions. Consequently, HeLa cells were used for most RNA base-editing experiments, particularly those involving editing of overexpressed exogenous transcripts. However, when targeting endogenous transcripts, achieving a high percentage of positively transfected cells is crucial for robust results. Each untransfected cell harbors the unedited transcript, which can adversely impact the overall readout. HEK293FT cells offer superior transfection efficiency and were, thus, used for targeting endogenous transcripts. By integrating a single genomic copy of ADAR1 p110, p150 or ADAR2 under the control of an CMV Tet-On promoter into Flp-In T-REx cells, we generated a set of cell lines that can be used to characterize how well dsRNA substrates such as bound guide RNAs or editing reporters are accepted by specific ADAR isoforms. Notably, the ADAR expression levels in these Flp-In cell lines exceeded typical endogenous levels, which increased bystander editing and facilitated the discovery of suppressive effects of wobble base pairs on bystander off-target edits.

**ADAR Flp-In T-REx cells.** A total of  $2.5 \times 10^5$  ADAR1 p110 or p150 Flp-In T-REx cells or  $3 \times 10^5$  ADAR2 Flp-In T-REx cells were seeded on poly(DLys)-coated 24-well plates in 500  $\mu$ l of DMEM + 10% FBS + 10 ng ml<sup>-1</sup> doxycycline. After 24 h, cells were transfected with 1,300 ng of *cis*-acting editing reporter plasmid (NucleoSpin Plasmid Transfection-grade, Macherey Nagel, 740490) using a 1:3 ratio of Lipofectamine 2000 (Thermo Fisher, 11668019). Then, 72 h after transfection, the cells were harvested. As the readout method, Sanger sequencing was used. Results are reported in Fig. 1, Extended Data Figs. 1 and 3 and Supplementary Figs. 1 and 4.

**HEK293FT cells.** HEK293FT cells ( $6 \times 10^4$ ) were seeded in 24-well plates in 450  $\mu$ l of DMEM + 10% FBS. After 24 h, cells were transfected with 1,200 ng of guide RNA plasmid (NucleoSpin Plasmid Transfection-grade, Macherey Nagel, 740490) using a 1:3 ratio of FuGene 6 (Promega, E2691). Then, 48 h after transfection, cells were harvested. As the readout method, Sanger sequencing was used. Results are reported in Extended Data Fig. 2 and Supplementary Figs. 2, 6 and 7.

**HeLa cells (experiments using encodable guide RNAs).** A total of  $0.8 \times 10^5$  HeLa cells were seeded in 24-well plates in 500  $\mu$ l of DMEM + 10% FBS. Then, 24 h after seeding, cells were transfected with a total of 1,000 ng (800 ng of LEAPER or CLUSTER guide RNA plasmid and 200 ng of target-encoding plasmid) or a total of 1,200 ng (800 ng of boxB or DR guide RNA plasmid, 200 ng of target-encoding plasmid and 200 ng of  $\Delta$ N-ADAR or Cas13-ADAR editase encoding plasmid) per well using a 1:1.5 ratio of plasmid to Lipofectamine 3000 ratio. Then, 72 h after transfection, cells were harvested. As the readout method, Sanger sequencing was used. Results are reported in Figs. 2–5, Extended Data Figs. 3, 5 and 6 and Supplementary Figs. 5 and 11.

**HeLa cells (experiments using an ASO guide RNA targeting the *PEXI*<sup>G843D</sup> transcript).** A total of  $1 \times 10^5$  HeLa cells were seeded in 24-well plates in 500  $\mu$ l of DMEM + 10% FBS. Then, 24 h after seeding, cells were transfected with 300 ng of target *PEXI*<sup>G843D</sup>-encoding plasmid per well using a 1:3 ratio of plasmid to FuGene 6. Then, 48 h after seeding, the chemically modified guide RNAs were forward-transfected with 25 pmol of guide RNA and 1.5  $\mu$ l of Lipofectamine RNAiMAX reagent (Thermo Fisher, 13778150) per well. Then, 24 h after guide RNA transfection, cells were harvested. The one-step RT-PCR (*PEXI* primer set A) was followed by a nested PCR (*PEXI* primer set B) for this specific target. As the readout method, Sanger sequencing was used. Results are reported in Fig. 3f.

**HeLa cells (stable integration of guide RNAs).** A total of  $1.5 \times 10^5$  HeLa cells were seeded in 12-well plates in 1 ml of DMEM + 10% FBS. Then, 24 h after seeding, cells were transfected with a total of 2,000 ng

(1,600 ng of PiggyBac transposon vector containing the guide RNA expression cassette and 400 ng of pTS687 PiggyBac transposase vector) using a 1:3 ratio of plasmid to FuGene 6 (Promega, E2691). Then, 24 h after transfection, cells were transferred to 10-cm dishes and, after another 48 h, selection was started by addition of 5  $\mu$ g ml<sup>-1</sup> puromycin. Puromycin was refreshed on day 6 after transfection and selection was stopped 8 days after transfection. Each time when the cells were split, an aliquot was taken aside and used for RNA extraction and editing analysis. Results are reported in Supplementary Fig. 8.

**HeLa cells (experiments using the dual-luciferase reporter system).** A total of  $2.4 \times 10^4$  HeLa cells were seeded in 96-well plates. Cells were transfected 24 h after seeding with 160 ng of guide RNA plasmid and 40 ng of dual-luciferase reporter per well using a 1:1.5 ratio of plasmid to Lipofectamine 3000 and a 1:2 ratio of plasmid to P3000 reagent. For equimolar comparisons among differently sized guide RNA plasmids, their amount was adjusted accordingly and filled to a total of 160 ng with an empty pcDNA3.1 plasmid. The luciferase assay was performed 48 h after transfection. Results are reported in Supplementary Figs. 9 and 10.

**Editing readout using dual-luciferase activity.** Dual-luciferase activity was measured with the Dual-Luciferase Reporter Assay System (Promega) according to the manual in 96-well plates. Cells were washed with PBS and then lysed in 1 $\times$  passive lysis buffer (35  $\mu$ l per well) while shaking for 15 min at 700 r.p.m. at room temperature. Cell lysate (30  $\mu$ l per well) was transferred to a LumiNunc 96-well plate (VWR, 732-2696) and measured in a Spark 10M plate reader (Tecan) using the dual-luciferase reporter assay reagents (35  $\mu$ l per well) with an autoinjector. The attenuation standard settings (OD-none) of the Tecan reader were used. Each measurement was performed for 10 s, starting 5 s after injection. Per treatment, five biological replicates were analyzed, each measured in one technical replicate. For data processing, measured blank values (background) were subtracted from samples and controls and then all firefly values were divided by the corresponding *Renilla* values. The resulting normalized firefly activity of all samples was then set in ratio to the positive control to obtain the restored normalized firefly activity as a percentage.

**Editing readout using Sanger sequencing (total RNA from cell lines).** Cells were harvested in RLT buffer (Qiagen, 79216), followed by RNA isolation using either the Monarch RNA cleanup kit (NEB, T2030L) or the RNeasy Mini RNA isolation kit (Qiagen, 74104). DNase I digestion was performed according to the manual using NEB DNase I (NEB, M0303S). One-step RT-PCR was performed using the Bio-techrabbit one-step RT-PCR kit (Biotech rabbit, BR0400102) for regular substrates. For difficult substrates (for example, the murine *Mecp2* W104>amber transcript), the OneTaq one-step RT-PCR kit (NEB, E5315S) was used. All samples were mixed with nuclease-free water (10  $\mu$ l) and heated to 90 °C for 2 min immediately before RT-PCR. If necessary, a sense-oligo (1  $\mu$ l, 10  $\mu$ M) corresponding to the used guide RNA (Supplementary Table 2) was included in the volume. For the murine *Mecp2* W104>amber transcript, the heating step was performed at 95 °C for 3 min under 12.5% DMSO (5% final concentration in RT-PCR mix). After TAE buffer–agarose gel electrophoresis and PCR cleanup (NucleoSpin Gel and PCR cleanup kit, Macherey Nagel, 740609), Sanger sequencing (Microsynth AG) was performed.

**RT-qPCR experiments.** RNA isolation was performed using either the Monarch RNA cleanup kit (NEB, T2030L) or the RNeasy Mini RNA isolation kit (Qiagen, 74104) according to the manufacturer's protocol. DNase digestion was performed according to the manufacturer's protocol (rigorous two-step incubation treatment) using the Turbo DNase Kit (Thermo Fisher, AM1907). RT with 500 ng of RNA per sample



was performed using the high-capacity cDNA RT kit (Thermo Fisher, 4368814), followed by PCR cleanup using the NucleoSpin Gel and PCR cleanup kit (Macherey Nagel, 740609). RT-qPCR was executed in an Applied Biosystems QuantStudio 7 Pro real-time PCR system (96-well or 384-well qPCR plate, 20 ng or 4 ng of cDNA (10 or 2 ng  $\mu\text{l}^{-1}$ ) per well). The Fast SYBR-Green mastermix (Applied Biosystems, 4385612) was used according to the manufacturer's protocol (10  $\mu\text{l}$  or 5  $\mu\text{l}$  of SYBR-Green mix, 7.2  $\mu\text{l}$  or 2.6  $\mu\text{l}$  of nuclease-free water and 0.4  $\mu\text{l}$  or 0.2  $\mu\text{l}$  of each primer). Samples and TE buffer negative controls were measured in three technical replicates.  $\Delta\Delta C_t$  calculations were performed as previously described<sup>45</sup>. Amplification efficiency and melting curves were analyzed for each new primer pair using a cDNA dilution series (Supplementary Fig. 11). For normalization, one or more housekeeping genes from the following list were used: human U6 small nuclear RNA (snRNA; *RNU6*), murine U6 snRNA (*Rnu6*), murine  $\beta$ -actin (*Actb*) and murine ribosomal protein 29 (*Rps29*). Results are reported in Fig. 6 and Supplementary Figs. 11–13.

**qPCR experiments (AAV episome copy number).** Genomic DNA isolation from murine brain tissue was performed using the QIAamp DNA Micro Kit (Qiagen, 56304) according to the manufacturer's protocol. qPCR was executed in an Applied Biosystems QuantStudio 7 Pro real-time PCR system (384-well qPCR plate, 4 ng of genomic DNA (gDNA; 2 ng  $\mu\text{l}^{-1}$ ) per well). The Fast SYBR-Green mastermix (Applied Biosystems, 4385612) was used according to the manufacturer's protocol (5  $\mu\text{l}$  of SYBR-Green-Mix, 2.6  $\mu\text{l}$  of nuclease-free water and 0.2  $\mu\text{l}$  of each primer). Samples and AE buffer negative controls were measured in three technical replicates.  $\Delta\Delta C_t$  calculations were performed as previously described<sup>45</sup>. For normalization, the murine *Actb* housekeeping gene was used. The vector copy number per  $\mu\text{g}$  of DNA was determined by absolute quantification using a standard curve. The templates for the standard curve were prepared according to the manufacturer's protocol (Takara, 6233). For the conversion of vector copy number per  $\mu\text{g}$  of DNA into copies per cell, 6.030 pg of gDNA was assumed per male C57BL/6 nucleus according to a previous study<sup>59</sup>. Results are reported in Fig. 6.

## Animal experiments

**Animal studies.** All animal procedures were approved by the Institutional Animal Care and Use Committees of Oregon Health and Science University. A total of ten *Mecp2*<sup>G311A</sup>-carrying male mice (W104X) were killed during this study. Three mice were injected with the nontargeting virus and seven were injected with the targeting virus. From these seven animals, five were used for RT-qPCR and Sanger, amplicon and RNA-seq. The remaining two mice were used for MeCP2 immunohistochemistry.

***Mecp2* 311G>A husbandry and genotyping.** All mice were housed in conventional laboratory housing under controlled humidity (target value, 44%; minimum, 24%; maximum, 64%), temperature (target value, 21 °C; minimum, 18 °C; maximum, 24 °C) and lighting (12-h light, 12-h dark period) with free access to food (regular mouse chow) and water. The *Mecp2*<sup>G311A/+</sup> mice were maintained by crossing to pure wild-type C57BL/6J mice. Genotyping was performed using primers specific for the *Mecp2*<sup>G311A</sup> allele. Separately, sex was determined using PCR primers specific for the X and Y chromosomes (Supplementary Table 2).

**Design of AAV-encoded guide RNAs.** The targeting virus encoded the circular *Mecp2*-targeting CLUSTER guide RNA V27.2.4 (Supplementary Table 1). The latter consisted of a TS and four RSs of 20 nt length each, a split-R/G V3 ADAR-recruiting domain, a ligation stem and two ribozymes required for its circularization (Extended Data Fig. 4). For the nontargeting guide RNA virus, the sequences of all four RSs and the targeting sequence were scrambled, while A linkers,

the ADAR recruiting motif and ribozymes required for circularization were unchanged (Supplementary Table 1).

**Viral vector preparation.** The AAV-PHP.eB vectors were produced by the Penn Vector Core Facility (Perelman School of Medicine, University of Pennsylvania) and titered using droplet digital PCR. Aliquots were stored at –80 °C before use.

**Viral injections.** P30–P34 male mice were deeply anesthetized with 3% isoflurane (v/v) and placed on a prewarmed surface. For each animal, a 100- $\mu\text{l}$  volume containing 4E12 viral genomes was injected into the retro-orbital sinus. Following injections, mice were monitored for pain and distress while recovering on a heated pad before being returned to their home cage.

**RNA isolation and Sanger sequencing (in vivo samples).** Four weeks after injection, brains were dissected in 5 mM HEPES in Hanks' balanced salt solution and total RNA was isolated from individual brain regions using QIAzol reagent (Qiagen, 79306) and the Qiagen miRNeasy Kit reagent (Qiagen, 217004) according to the manufacturer's instructions. RNA was reverse-transcribed using the SuperScript III first-strand synthesis system (Invitrogen, 18080051) and primed using oligo dT. Endogenous *Mecp2* cDNA was amplified and analyzed by Sanger sequencing (Supplementary Table 2).

**RT-qPCR experiments (in vivo samples).** Total RNA was isolated from individual brain regions using QIAzol reagent (Qiagen, 79306) and the Qiagen miRNeasy Kit (Qiagen, 217004) according to the manufacturer's instructions. DNase digestion was performed on 1  $\mu\text{g}$  of RNA using Turbo DNase Kit (Thermo Fisher, AM1907) according to the manufacturer's instructions. DNase-treated RNA was then reverse-transcribed using the SuperScript III first-strand synthesis system (Invitrogen, 18080051) and was primed using oligo dT. RT-qPCR experiments were performed on an Applied Biosystems QuantStudio 6 Flex real-time PCR system. All samples were run in triplicate with a standard curve using SYBR select master mix reagent (Thermo Fisher, 4472918) for inward and outward primer sets to quantify CLUSTER guide concentrations across brain regions and circularization, respectively. CLUSTER guide expression was calculated relative to the geometric mean of three housekeeping genes (*Rnu6*, *Rps29* and *Actb*). Primers sequences are listed in Supplementary Table 2.

**Immunostaining.** Immunostaining was performed as previously described<sup>60</sup>. Briefly, mice were anesthetized using an intraperitoneal injection of 2,2,2-tribromoethanol (Sigma Aldrich, T48402) and killed by transcardial perfusion of PBS pH 7.4, followed by 4% depolymerized paraformaldehyde. Brains were cryoprotected with sucrose, embedded in freezing medium and stored at –80 °C. Sagittal whole-brain sections were cut at 25  $\mu\text{m}$  using a cryostat and stored at –20 °C until staining. Sections underwent heat-mediated antigen retrieval before blocking in PBST (0.1% Triton X-100 in PBS, pH 7.4) and 10% BSA (Sigma, A2153-500G) for 1 h at room temperature. Sections were incubated overnight at 4 °C with rabbit anti-MecP2 antibody (rabbit monoclonal antibody D4F3, RRID: AB\_2143849, Cell Signaling, 1:500) diluted in blocking buffer. Sections were washed with PBST and incubated for 1 h at room temperature with Alexa Fluor 488 donkey anti-rabbit IgG (Invitrogen, A21206 1:750) diluted in blocking buffer. Sections were washed and incubated with 300 nM DAPI (Invitrogen, D1306) in PBS for 5 min. After a final wash in PBS, sections were mounted using Fluoromount G (Invitrogen, 00-4958-02). Images were acquired using ZEN 2.3 SP1 FPS Black (version 14.0.0.0) on a Zeiss LSM 710 confocal microscope with a C-Apochromat  $\times 40$  (1.2 numerical aperture) objective and LSM T-PMT detector. Images were processed in ImageJ (version 1.54f). Murine MeCP2 protein and nuclei (DAPI) were pseudo-colored in ImageJ to highlight colocation puncta. Results are reported in Fig. 6p and Extended Data Fig. 9.



## Deep amplicon NGS experiment

Retro-orbital injections and RNA isolation were performed as explained above. Two settings were carried out, each with an independent duplicate: (1) nontargeting circular CLUSTER guide RNA and (2) *Mecp2* W104>amber transcript-targeting circular CLUSTER guide RNA. For the latter we selected the two thalamus samples that showed the highest *Mecp2* on-target editing yield during Sanger sequencing. Purified RNA was delivered to the Massively Parallel Sequencing Shared Resource Core (MPSSR) at the Oregon Health and Science University for Illumina library preparation and then transferred to the Molecular Technologies Core (MTC) at the Oregon National Primate Research Center for Illumina amplicon sequencing. The library was prepared with the TruSeq DNA nano library prep kit and sequenced with a MiSeq instrument (1 million reads, 500 cycles, 2 × 500 bp of paired-end reads; Illumina). Furthermore, 2 nM pooled libraries were spiked with 5% PhiX before the sequencing run. The sequencing reads were demultiplexed using BCL Convert (version 2.4.0). Adaptors were trimmed with FastQ (version 1.0.0). Raw FASTQ files were processed using Seqtk (version 1.3-r106). The command 'seqtk trimfq in.fastq > out.fastq' was used to trim low-quality bases. The command 'seqtk seq -q30 -n N in.fastq > out.fastq' was used to mask base calls with a quality value < 30 as N. The base-call accuracy of the remaining bases was, thus, 99.9%. FASTQ file quality was assessed with FastQC (version 0.11.9)<sup>61</sup>.

**Mapping of deep amplicon-seq reads.** Reads were aligned to the GRCm38/mm10 reference genome using BWA-MEM (version 0.7.17-r1188). The alignments were analyzed using the Integrative Genomics Viewer (version 2.16.2). The editing yields were exported from IGV (version 2.16.2) and then further processed in Excel 2016.

## Next-generation RNA-seq experiment

Retro-orbital injections and RNA isolation were performed as explained above. Two settings were carried out, each with an independent duplicate: (1) nontargeting circular CLUSTER guide RNA and (2) *Mecp2* W104>amber transcript-targeting circular CLUSTER guide RNA. For the latter, we selected the two thalamus samples that showed the highest *Mecp2* transcript on-target editing yield during Sanger sequencing. Purified RNA was delivered to CeGaT for total RNA-seq. The library was prepared from 100 ng of RNA with the KAPA RNA HyperPrep library prep kit with RiboErase (HMR) and KAPA globin depletion hybridization oligos (Roche) and sequenced with a NovaSeq 6000 (50 million reads, 2 × 100 bp of paired-end reads; Illumina). The sequencing reads were demultiplexed using Illumina bcl2fastq (version 2.20). Adaptors were trimmed with Skewer (version 0.2.2)<sup>62</sup>. No quality trimming of the reads was performed. Raw FASTQ file quality was assessed with FastQC (version 0.11.8)<sup>61</sup>. One of the nontargeting CLUSTER guide RNA samples failed quality control (RNA integrity number of 1) and, thus, had to be removed from the analysis. Results are reported in Fig. 6o and Extended Data Fig. 8. In addition, we reanalyzed brainstem samples from a similar previous study that applied the boxB/ΔN-ADAR system<sup>44</sup>. From this existing dataset, we randomly selected samples from the untreated Rett mouse negative control group and the two samples from the boxB/ΔN-ADAR targeting virus-treated group that showed the highest *Mecp2* transcript on-target editing yield. Results are reported in Extended Data Fig. 8.

**Mapping of RNA-seq reads.** Reads were uniquely aligned to the GRCm38/mm10 reference genome using STAR (version 2.7.3a) ('STAR-2.7.3a --alignSJoverhangMin 8 --alignIntronMax 1000000 --alignMatesGapMax 600000 --outFilterMismatchNoverLmax 0.3 --outFilterMismatchNoverReadLmax1 --outFilterMatchNminOverLread 0.66 --outFilterMultimapNmax 1 --outReadsUnmapped Fastx --outSAMattributes All --outSAMtype BAM Unsorted --quantMode GeneCounts --genomeLoad LoadAndKeep --limitBAMsortRAM 37580-963840 --outBAMsortingThreadN 10 --runThreadN 40 --genomeDir

mm10.STAR.7.ReadsLn100.gencodeM18 --readFilesCommand cat --readFilesIn file-name\_R1.fastq file-name\_R2.fastq --outFileNamePrefix file-name')<sup>63</sup>.

**RNA-editing index.** The RNA-editing index tool<sup>46</sup> calculates the average editing level across all A bases in a set of regions. The editing index is defined as the ratio of the number of A•G mismatches to the total coverage of A bases. Genomic sites overlapping common single-nucleotide polymorphisms (dbSNP142) were excluded. The RNA-editing index tool was used to assess the overall editing for all the coding sequence regions of the mouse genome.

**Global RNA editing.** To assess global RNA editing, we used the known editing sites from the REDIPortal<sup>55</sup> and from a very small set of evolutionarily conserved A-to-I editing sites<sup>64</sup>. We calculated the editing levels in those sites using REDITools<sup>65</sup> with a coverage ≥ 50 reads for every site. For every group of each dataset (this study and the previous study<sup>44</sup>), we combined the editing levels of the replicates for each site. We then selected the 2,532 sites that were common in both datasets.

## Data analysis

Data were analyzed using Excel 2016 and GraphPad Prism 8. Figures were created with CorelDraw 2017. The manuscript was written using Word 2016. The custom GuideRNA-Forge tool was written in Python (version 3.9). Guide RNA folds were created using the ViennaRNA Package (version 2.0, <http://rna.tbi.univie.ac.at/cgi-bin/RNAWebSuite/RNA-fold.cgi>). qPCR analysis was performed using the Applied Biosystems 7500 data analysis software version 1.1 (animal experiment) and version 2.3 (remainder of the study).

## Reporting summary

Further information on research design is available in the Nature Portfolio Reporting Summary linked to this article.

## Data availability

Transcriptome-wide RNA-seq data are accessible from the National Center for Biotechnology Information (NCBI) Gene Expression Omnibus database with accession code [GSE265898](https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE265898). Deep amplicon NGS data are accessible from the NCBI Sequence Read Archive SRA database with accession code [PRJNA1100948](https://www.ncbi.nlm.nih.gov/sra/PRJNA1100948). Transcriptome-wide RNA-seq data from Sinnamon et al. are accessible from the NCBI SRA database with accession code [PRJNA849938](https://www.ncbi.nlm.nih.gov/sra/PRJNA849938). Structural data from Thuy-Boun et al. and Stefl et al. are available from PDB [5HP2](https://www.rcsb.org/structure/5HP2) and [2L2J](https://www.rcsb.org/structure/2L2J), respectively.

## Code availability

The full code of the GuideRNA-Forge tool is available upon request. The compiled tool can be obtained from GitHub (<https://github.com/recruitment-cluster-finder/GuideRNA-FORGE>)<sup>66</sup>.

## References

- Thuy-Boun, A. S. et al. Asymmetric dimerization of adenosine deaminase acting on RNA facilitates substrate recognition. *Nucleic Acids Res.* **48**, 7958–7972 (2020).
- Steffl, 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).
- Baugh, E. H. et al. Real-time PyMOL visualization for Rosetta and PyRosetta. *PLoS ONE* **6**, e21931 (2011).
- Capparelli, R. et al. DNA content differences in laboratory mouse strains determined by flow cytometry. *Cytometry* **29**, 261–266 (1997).
- Sinnamon, J. R. et al. In vivo repair of a protein underlying a neurological disorder by programmable RNA editing. *Cell Rep.* **32**, 107878 (2020).

61. Andrews, S. FastQC: a quality control tool for high-throughput sequence data. Available online at <http://www.bioinformatics.babraham.ac.uk/projects/fastqc/>. (2010).
62. Jiang, H. et al. Skewer: a fast and accurate adapter trimmer for next-generation sequencing paired-end reads. *BMC Bioinf.* **15**, 182 (2014).
63. Dobin, A. et al. STAR: ultrafast universal RNA-seq aligner. *Bioinformatics* **29**, 15–21 (2013).
64. Gabay, O. et al. Landscape of adenosine-to-inosine RNA recoding across human tissues. *Nat. Commun.* **13**, 1184 (2022).
65. Lo Giudice, C. et al. Investigating RNA editing in deep transcriptome datasets with REDIttools and REDlportal. *Nat. Protoc.* **15**, 1098–1131 (2020).
66. recruitment-cluster-finder / GuideRNA-FORGE. *GitHub* <https://github.com/recruitment-cluster-finder/GuideRNA-FORGE> (2024).

## Acknowledgements

We gratefully acknowledge generous support from the Deutsche Forschungsgemeinschaft (DFG, German Research Foundation; STA 1053/10-1 and STA 1053/11-1 to T.S.) and the Baden-Württemberg Ministry of Science as part of the Excellence Strategy of the German Federal and State Governments with an intramural innovation grant (IG-Reauts-2019-05 to P.R.). Further generous support was provided by the International Rett Syndrome Foundation (grant 3806 to T.S.), by the VolkswagenStiftung (Momentum Grant 96 876 to T.S.) and by the National Institutes of Health (grant NS110868 to G.M.). E.Y.L. was supported by the Israel Science Foundation grants 2039/20 and 2637/23. We also want to thank J. J. C. Rosenthal and J. D. Quiroz (Eugene Bell Center for Regenerative Biology and Tissue Engineering, The Marine Biological Laboratory) and G. Dodt (Interfaculty Institute of Biochemistry, University of Tübingen) for supplying us with the plasmids encoding the  $\lambda$ N-ADAR editases and PEX1<sup>G843D</sup> that were applied in Fig. 3. In addition, we want to thank J. B. Li and I. Jarmoskaite (Department of Genetics, Stanford University) for insights and discussion on engineering improved ADAR recruitment motifs. Lastly, we want to acknowledge technical support from the CeGaT GmbH (Illumina RNA-seq), from the MPSSR at the Oregon Health and Science University (Illumina library

preparation) and from the MTC at the Oregon National Primate Research Center (Illumina amplicon sequencing).

## Author contributions

P.R. and T.S. contributed equally to this work as the senior authors. P.R. and T.S. conceptualized and analyzed the overall study. P.R. and N. Wahn conceptualized and realized the software implementation. T.S., G.M. and P.R. conceptualized and analyzed the animal experiment. S.K., J.F.Y. and J.F. conceptualized, performed and analyzed the animal experiment. E.Y.L., M.E., S.K., P.R. and T.S. conceptualized and analyzed the NGS experiments. P.R., C.F., C.P.W., D.T.H., N. Waffenschmidt, M.F., L.S.P., A.E.S. and Y.F. performed all other wet lab experiments. P.R. and T.S. wrote the manuscript. All authors proofread the manuscript.

## Competing interests

T.S., G.M. N.W., C.P.W., L.S.P., A.E.S. and P.R. hold patents on site-directed RNA editing. T.S., C.P.W., A.E.S. and P.R. are inventors of a filed patent based on the work published here. T.S. is a cofounder and shareholder of AIRNA Bio. G.M. is a cofounder and shareholder of Vico Therapeutics. The other authors declare no conflicts of interest.

## Additional information

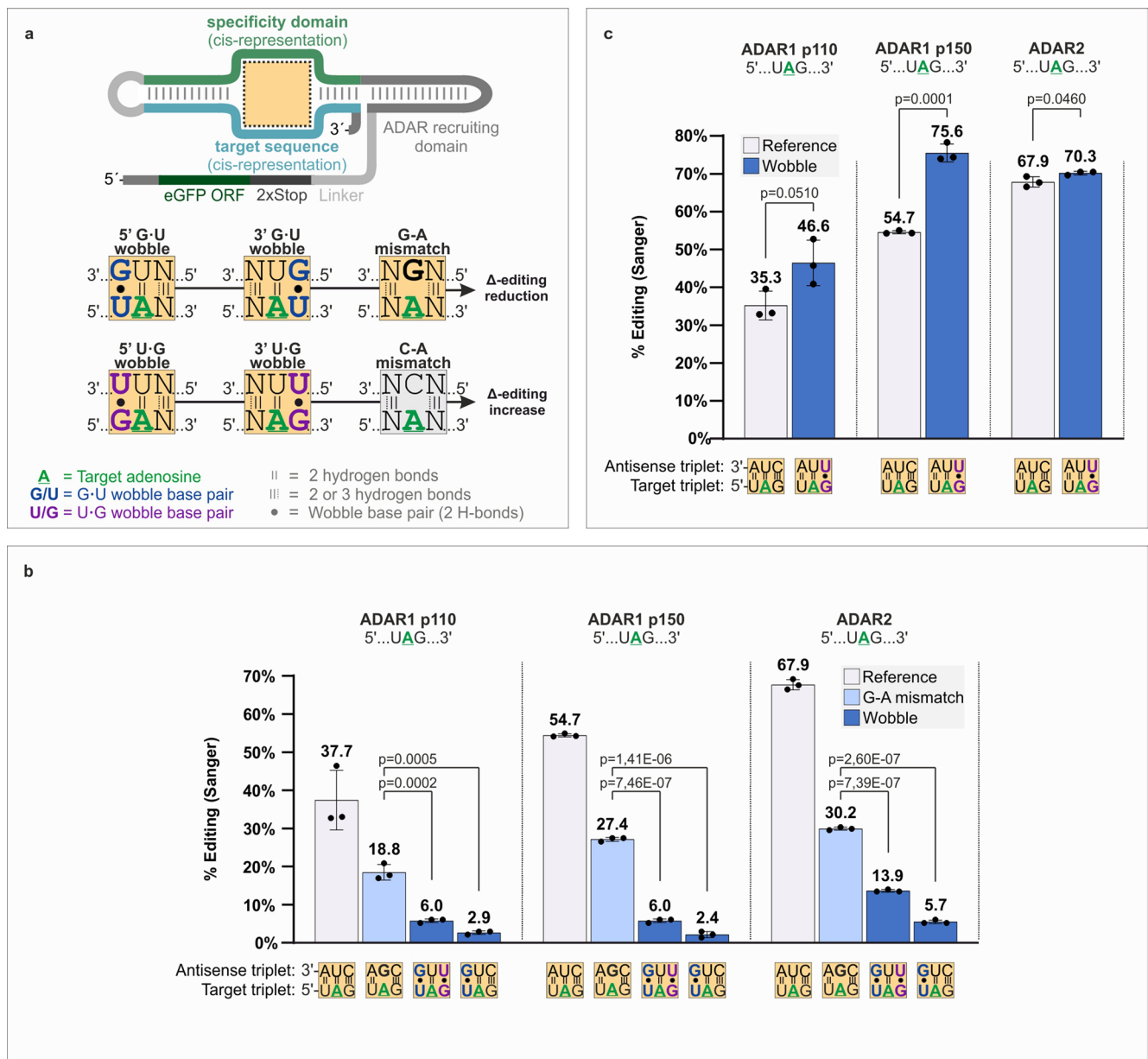
**Extended data** is available for this paper at <https://doi.org/10.1038/s41587-024-02313-0>.

**Supplementary information** The online version contains supplementary material available at <https://doi.org/10.1038/s41587-024-02313-0>.

**Correspondence and requests for materials** should be addressed to Thorsten Stafforst.

**Peer review information** *Nature Biotechnology* thanks Michael Jantsch and the other, anonymous, reviewer(s) for their contribution to the peer review of this work.

**Reprints and permissions information** is available at [www.nature.com/reprints](http://www.nature.com/reprints).



**Extended Data Fig. 1 | The enhancing and suppressing effects of G-U and U-G wobble base pairs on RNA editing apply to all catalytically active ADAR isoforms. (a)** Schematic of the utilized cis-acting editing reporter construct. **(b)** Editing yields when applying G-A mismatches or G-U wobbles for editing suppression at the 5'-UAG triplet in either ADAR1 p110-, ADAR1 p150- or ADAR2-Flp-In T-REx cells. The exact position of each mismatch or wobble is indicated below each bar. The target adenosine is highlighted in green, bold and

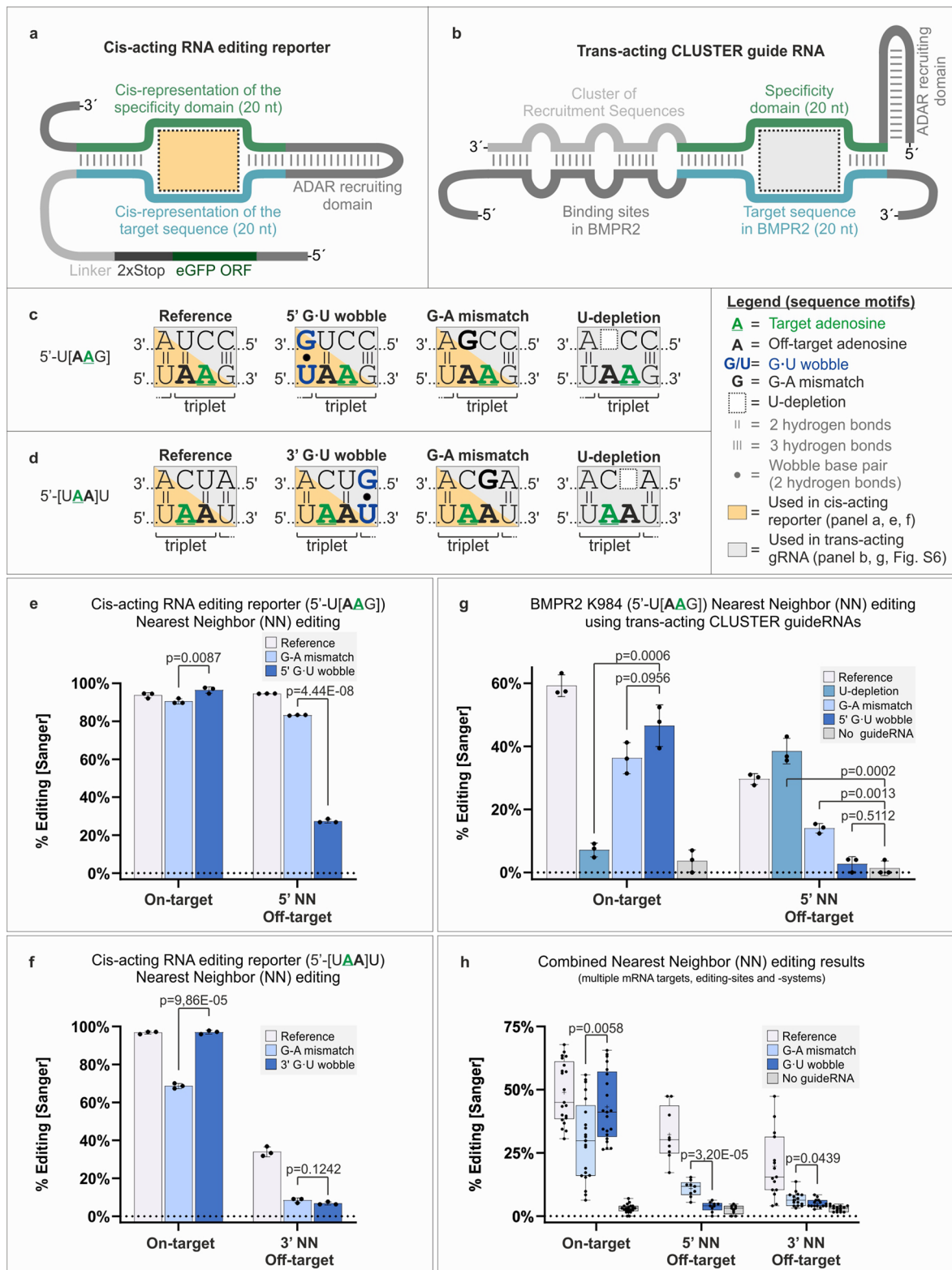
underlined. Bases that induce wobbles or mismatches are highlighted in dark blue and bold. The 5' and 3' orientation of the triplet and the used ADAR isoform are shown above the grouped bars. **(c)** As b) but regarding editing yields when applying enhancing U-G wobbles. Data for b) and c) are shown as the mean  $\pm$  s.d. of N = 3 biological replicates per bar. For statistical analysis, a student's *t*-test (two-tailed, parametric) was applied.

a							b						
Endogenous NUP43 3'UTR							Endogenous RAB7A 3'UTR						
Triplet context	Adenosine pos. from 5'-end (relative to target site)	LEAPER	GU at GU amenable sites	GA at GU amenable sites	GA at all off-target sites	GA&GU at all off-target sites	Triplet context	Adenosine pos. from 5'-end (relative to target site)	LEAPER	GU at GU amenable sites	GA at GU amenable sites	GA at all off-target sites	GA&GU at all off-target sites
GAC	2424 (-75)	1±1	1±0	2±2	1±1	1±0	GAG	1595 (-55)	0±0	0±0	0±0	0±0	0±0
GAG	2429 (-70)	2±3	1±1	1±1	0±0	1±0	GAA	1602 (-48)	0±0	0±0	0±0	0±0	0±0
GAC	2431 (-68)	1±1	1±1	1±0	0±0	1±0	AAC	1603 (-47)	12±2	1±0	0±0	0±0	0±0
UAA	2436 (-63)	34±3	0±1	0±0	0±0	1±1	GAU	1614 (-36)	0±0	0±0	0±0	0±0	0±0
AAU	2437 (-62)	17±6	2±2	1±1	1±2	0±0	CAG	1619 (-31)	7±3	2±1	0±0	0±0	0±0
GAC	2440 (-59)	1±1	1±1	1±1	0±0	0±0	GAG	1621 (-29)	0±0	0±0	0±0	0±0	0±0
CAA	2442 (-57)	5±2	5±1	2±1	0±0	0±1	UAC	1627 (-23)	33±6	3±0	1±1	2±1	0±0
AAU	2443 (-56)	12±0	10±3	3±0	0±0	3±1	CAG	1631 (-19)	15±4	6±1	1±0	0±0	3±2
AAU	2444 (-55)	2±0	2±0	2±0	1±1	0±0	GAA	1633 (-17)	2±2	1±1	0±0	0±0	0±0
CAC	2447 (-52)	5±3	0±1	1±1	0±0	0±0	AAU	1634 (-16)	17±4	1±0	2±1	1±1	1±1
UAU	2453 (-46)	21±3	0±0	1±1	0±0	1±1	GAA	1640 (-10)	1±1	0±0	0±0	0±1	0±0
UAG	2456 (-43)	48±3	1±0	2±1	1±0	1±0	AAA	1641 (-9)	20±6	3±1	0±0	0±0	0±0
GAC	2458 (-41)	1±2	0±0	0±0	0±0	0±0	AAU	1642 (-8)	15±3	0±0	0±0	0±0	0±0
CAA	2460 (-39)	5±6	1±1	1±1	0±1	0±1	CAG	1646 (-4)	0±0	0±0	0±0	0±0	0±0
AAU	2461 (-38)	14±2	1±1	0±0	0±0	0±0	UAG	1650 (±0)	54±6	45±3	25±3	29±2	32±4
GAA	2465 (-34)	4±6	1±1	0±0	1±1	0±0	CAG	1656 (+6)	0±0	0±0	0±0	0±0	0±0
AAG	2466 (-33)	23±3	17±2	2±1	0±0	0±0	UAU	1659 (+9)	34±7	0±0	0±0	0±0	0±0
CAU	2473 (-26)	1±1	1±1	3±1	2±1	2±2	UAC	1666 (+16)	12±5	0±0	0±0	2±2	1±1
UAC	2477 (-22)	27±4	1±0	3±1	2±0	1±1	CAG	1668 (+18)	2±5	0±0	0±0	0±1	1±1
UAA	2484 (-15)	25±7	0±0	4±2	1±1	0±0	UAG	1671 (+21)	10±4	0±0	0±0	0±0	0±0
AAU	2485 (-14)	6±1	2±2	1±1	2±4	1±1	GAC	1673 (+23)	1±1	1±1	0±0	0±0	0±0
CAA	2490 (-9)	2±2	1±1	1±2	1±2	0±0	CAC	1675 (+25)	0±0	0±0	0±0	1±2	2±2
AAU	2491 (-8)	3±3	0±0	0±0	0±0	0±0	CAA	1677 (+27)	10±5	0±0	0±0	0±1	0±0
UAA	2493 (-6)	29±3	0±0	1±1	0±1	0±0	AAG	1678 (+28)	7±6	1±0	0±0	0±0	0±0
AAU	2494 (-5)	4±2	0±0	0±0	0±0	0±0	GAA	1680 (+30)	0±0	0±0	0±0	0±0	0±1
CAU	2497 (-2)	3±2	0±0	1±2	0±0	0±0	AAU	1681 (+31)	2±2	0±0	0±0	1±1	0±1
UAG	2499 (±0)	57±6	36±3	17±6	12±3	23±3	UAU	1684 (+34)	19±5	0±0	0±0	0±0	0±0
UAU	2503 (+4)	29±4	7±4	3±2	0±0	2±1	UAC	1688 (+38)	0±0	0±0	0±0	0±0	0±0
GAA	2507 (+8)	0±0	0±0	0±0	1±1	0±0	UAU	1697 (+47)	0±0	0±0	0±0	0±0	0±0
AAU	2508 (+9)	11±4	0±0	1±0	0±0	0±0	CAA	1700 (+50)	0±1	1±2	0±0	0±0	0±0
UAU	2511 (+12)	44±6	0±0	1±1	0±0	0±0	AAA	1701 (+51)	4±7	1±2	2±1	1±1	0±0
UAA	2516 (+17)	48±6	0±0	9±5	0±0	1±1	AAG	1702 (+52)	1±1	0±0	1±1	1±1	0±0
AAG	2517 (+18)	26±9	1±1	1±1	0±0	0±0	GAC	1704 (+54)	0±0	0±0	0±0	0±0	0±0
UAA	2537 (+38)	3±0	0±0	0±0	0±0	1±1							
AAU	2538 (+39)	2±2	1±1	1±0	1±1	0±0							
UAA	2540 (+41)	31±5	3±1	5±6	2±2	1±1							
AAA	2541 (+42)	25±6	15±2	10±9	13±1	4±7							
AAU	2542 (+43)	2±1	0±0	0±0	0±0	0±0							
UAU	2544 (+45)	27±2	2±0	2±2	1±1	0±1							
UAA	2546 (+47)	92±3	90±1	91±3	92±2	88±2							
UAU	2548 (+49)	50±8	38±3	40±6	35±2	35±3							
UAU	2550 (+51)	17±3	2±1	2±0	0±0	0±1							
GAG	2558 (+59)	0±0	0±0	0±0	0±0	0±0							
GAU	2560 (+61)	2±1	1±0	1±0	1±0	1±1							
GAG	2564 (+65)	0±0	0±1	1±1	0±0	0±0							

**Extended Data Fig. 2 | G-U wobbles improve efficiency and precision of trans-acting LEAPER guide RNAs for endogenous targets. (a-b)** Editing heat-map of the LEAPER guide RNA binding site within the indicated transcripts. The basic design column, LEAPER, lacks G-A mismatches and wobble base pairs. The other guide RNAs contain either G-A mismatches or G-U wobbles at G-U amenable sites or a combination of both solutions at all bystander sites. In the latter case, G-A mismatches are placed at sites not amenable to G-U wobbles. The triplet context for each listed editing event is given with the target adenosine highlighted in

green and all off-target adenines in blue. The position of each site is given relative to the transcript and the target adenosine (±0 position). The editing events detected at position 2546 (+47) and 2548 (+49) of the NUP43 transcript were natural events and were thus skipped in regard to bystander suppression. Editing was performed with plasmid-borne guide RNA and endogenous target transcripts in HEK293FT cells (endogenous ADAR). Data are shown as the mean editing percentage ± s.d. of N = 3 biological replicates.



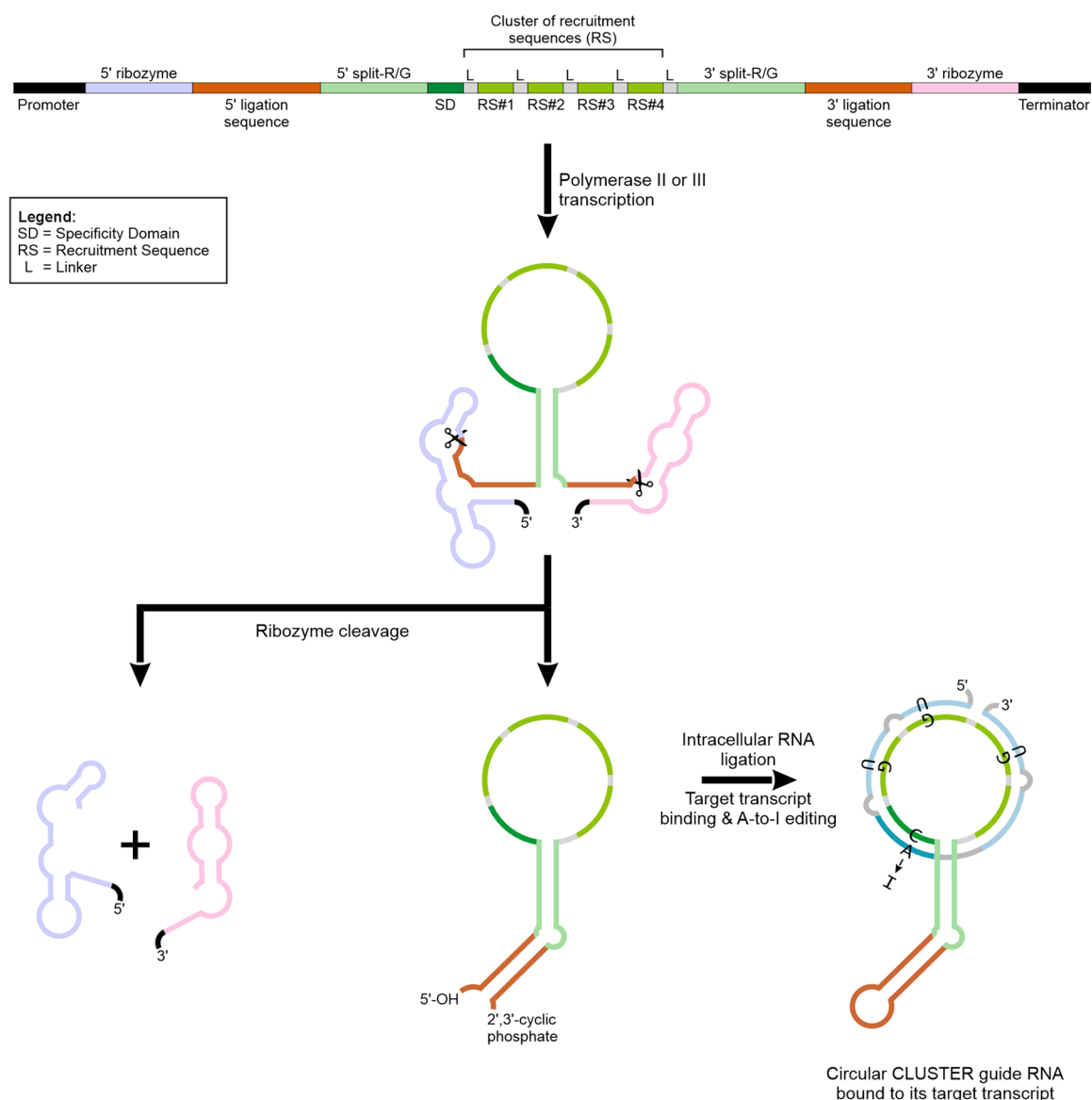


Extended Data Fig. 3 | See next page for caption.

**Extended Data Fig. 3 | G·U wobbles significantly outperform previous solutions in bystander suppression at nearest neighbor (NN) sites.**

(a) Design of the cis-acting RNA editing reporter constructs used in e) and f). (b) Design of the linear trans-acting CLUSTER guide RNAs used in g). (c) Illustration of the base-pairing motifs applied to the constructs in a) (ocher) and b) (grey) that were used to suppress bystander editing at the 5'-U[**AAG**] sequence motif in e) and g). (d) As c) but for the 5'-[**UAA**]U sequence motif and the results in f). (e) Suppression of bystander editing at the 5'-NN **A** and concurrent effect on the target **A** editing yield in a 5'-U[**AAG**] sequence motif. (f) Similar to e) but targeting the **A** in a 5'-[**UAA**]U sequence motif and suppressing bystander editing at a 3'-NN **A**. Editing in e) and f) was performed in ADAR1 p110 Flp-In T-REx cells using plasmid-borne reporters as displayed in a). (g) Suppression of bystander editing at a 5'-NN **A** when editing a 5'-U[**AAG**] site (K984) of the human BMPR2 transcript using trans-acting CLUSTER

guide RNAs. (h) Meta-analysis of bystander suppression at 5'- and 3'-NN sites over three different target transcripts (AH11, BMPR2 and COL3A1), three different A-rich target triplets flanked either 5' and/or 3' by uridines (5'-[**UAA**]U, 5'-U[**AAG**], 5'-U[**AAA**]U), and applying two trans-acting site-directed RNA editing systems (CLUSTER/endogenous ADAR, boxB/ $\Delta$ N-ADAR2Q). The full data set of the meta-analysis is shown in Supplementary Fig. S5. Editing in g) and h) was performed in HeLa cells using plasmid-borne editase ( $\Delta$ N-ADAR2Q) and/or guide RNA (2x boxB, CLUSTER). Data in e) – g) are shown as the mean editing percentage  $\pm$  s.d. of N = 3 biological replicates. Data in h) are shown as the mean editing percentage  $\pm$  s.d. of three 5'-NN off-target sites, seven on-target sites, and five 3'-NN off-target sites containing N = 3 biological replicates per site. The center line represents the median, the plus-sign the mean. The whiskers extend to the minimum and maximum values. The box limits the 25<sup>th</sup> and 75<sup>th</sup> percentile. For statistical analysis, a student's *t*-test (two-tailed, parametric) was applied.



**Extended Data Fig. 4 | Circularization of a split-R/G CLUSTER guide RNA using the Tornado expression system.** The guide RNA is expressed from either a polymerase III (for example U6) or a polymerase II (for example CAG) promoter with a corresponding terminator. 5' and 3' ribozymes auto-cleave the primary transcript resulting in a 5'-OH and a 2',3'-cyclic phosphate overhang, respectively, which are subsequently ligated by the endogenous RTCB ligase. In the final guide

RNA, the ligation stem (orange) forms a nearly continuous duplex with the split-R/G motif (light green), only interrupted by a five-nucleotide bulge. The circular guide RNA binds to its target transcript. Strategic placement of wobble base pairs within the recruitment sequences (RS) and/or the specificity domain (dark green) suppresses off-target editing, while a C-A mismatch within the specificity domain (SD) promotes on-target A-to-I RNA editing.

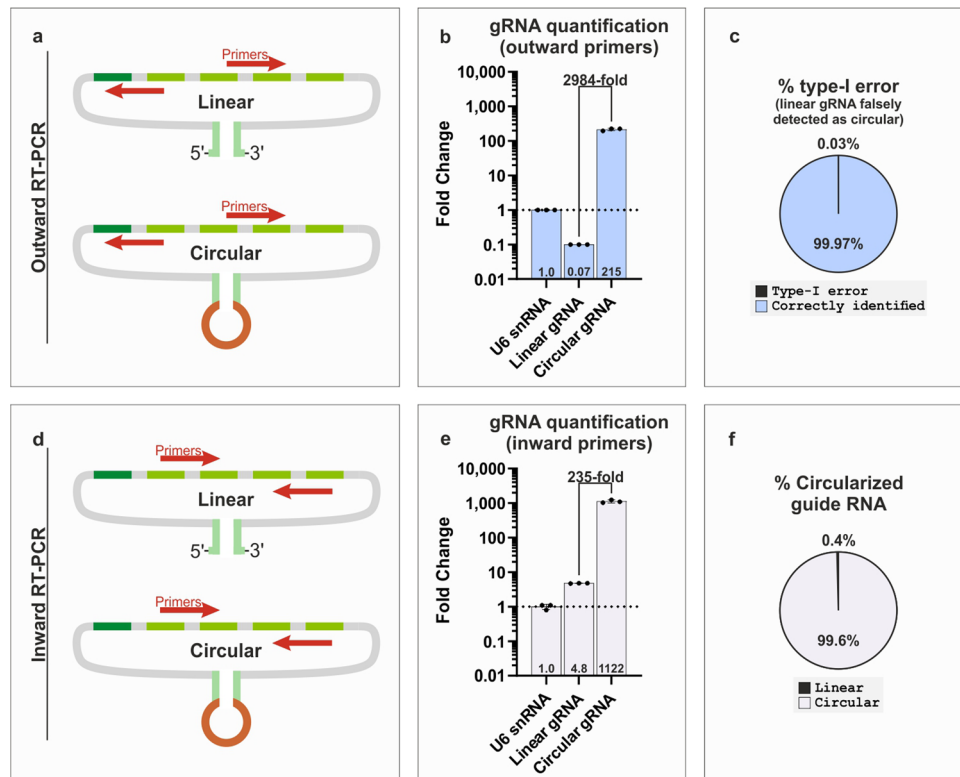




**Extended Data Fig. 5 | See next page for caption.**

**Extended Data Fig. 5 | Extended data set for editing Mecp2 W104amber with LEAPER and CLUSTER guide RNAs.** Panels (a) to (d) give schematic representations of the used linear and circular LEAPER and CLUSTER guide RNAs and their binding to the target mRNA. LEAPER guide RNAs were 111 nt long, centered with a C-A mismatch on the target adenosine. CLUSTER guide RNAs combined an ADAR-recruiting domain (R/G motif), a 20 nt long specificity domain containing the C-A mismatch with the target adenosine at position 8, and a cluster of three recruitment sequences of 20 nt length each. The circular CLUSTER guide RNA combines the ligation stem for circularization with the ADAR-recruiting motif into a split-R/G motif. (e) Editing heat-map using linear and circular LEAPER guide RNAs as seen in a) and b). (f-h) Editing heat map using linear and circular CLUSTER guide RNAs as seen in c) and d). Notably, the linear R/G version 21 guide RNA achieved already highly precise editing with an impressive  $38 \pm 7\%$  on-target editing yield, clearly better than the best circular

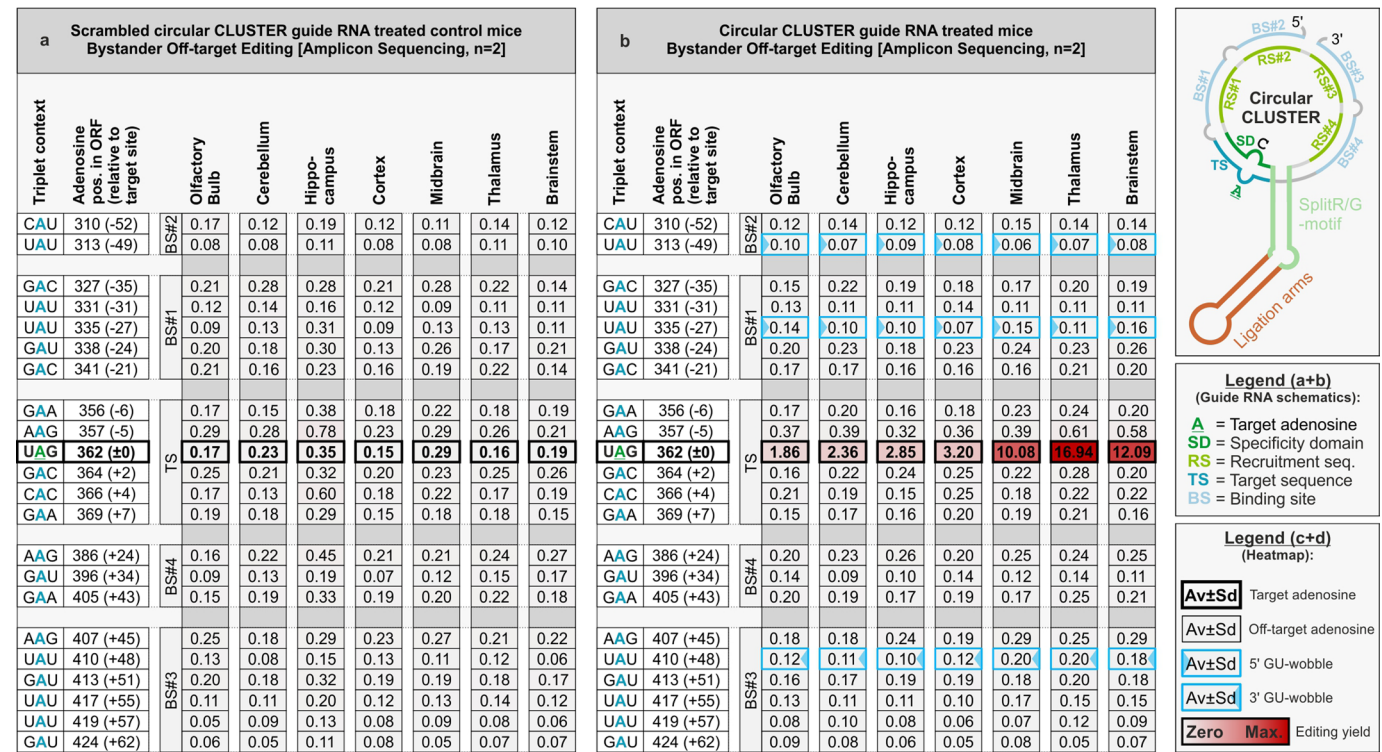
LEAPER guide RNA design ( $25 \pm 2\%$  on-target yield) with similar precision. A circular CLUSTER guide RNA with six recruitment sequences, but lacking the ADAR recruitment motif, achieved  $54 \pm 7\%$  on-target editing with very high precision after applying the GU wobble strategy. However, adding a split-R/G recruiting motif achieved even better on-target editing yields of  $75 \pm 1\%$  to  $82 \pm 1\%$ , even though only 4 recruitment sequences were applied. GU wobbling was required to eliminate bystander editing at the binding sites of recruitment sequences. The triplet context for each listed editing event is given with the edited adenosine highlighted in bold green. The position of each bystander site (bold blue) is given relative to the target adenosine ( $\pm 0$  position). Editing was performed with plasmid-borne guide RNA and murine Mecp2 in HeLa cells (endogenous ADAR). Data are shown as the mean editing percentage  $\pm$  s.d. of  $N = 3$  biological replicates.

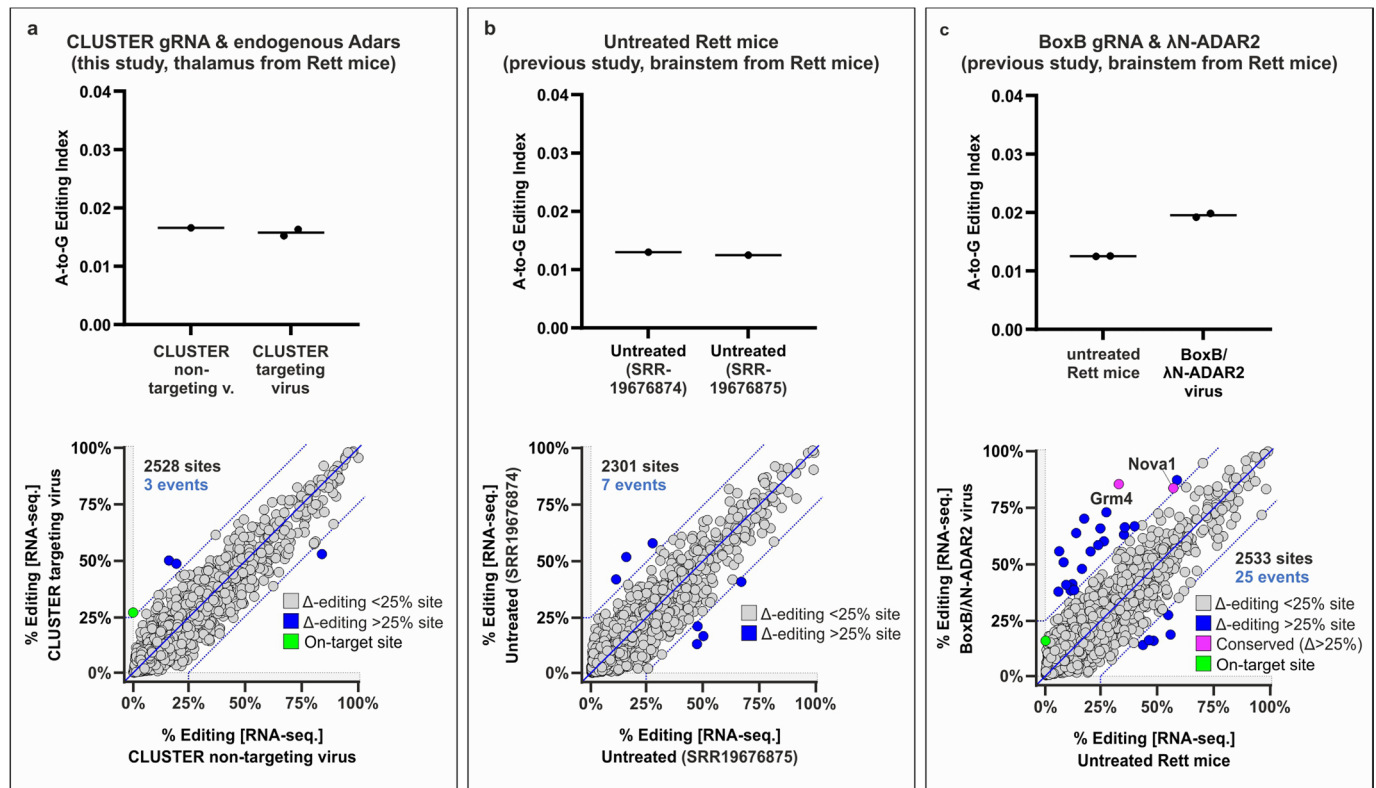


**Extended Data Fig. 6 | Verifying guide RNA circularization via quantitative RT-qPCR.** The circular state of the split-R/G CLUSTER guide RNA (V27.2.4, pTS2108) targeting mMeCP2 W104Amber was verified by using quantitative RT-PCR on in vitro samples from transfected HeLa cells. As reference a linear version of the same guide RNA was used. The dataset was normalized to human *RNU6* snRNA. The amplicon of the outward primer pair in (a) is 187 bp long and requires reverse transcription and PCR through a strong hairpin structure. It can thus not be directly compared to the inward primer pair in (d) which is smaller (75 bp) and does not contain inhibitory secondary structures. (a) Binding sites of the outward primer pair, which can only amplify circular guide RNAs. (b) Fold change of linear and circular split-R/G CLUSTER guide RNAs (V27.2.4, pTS2108, pTS2228) relative to endogenous human *RNU6* as detected by using the outward primer pair. The 2984-fold difference between linear and circular shows that the endogenous circularization of the guide RNAs through the RTCB ligase (Extended Data Fig. 4) was successful and that the outward primer pair can identify circular

guide RNAs with high confidence (low type-I error rate). (c) Type-I error rate calculated from (b) amounts to only 0.03%, thus 99.97% of circular guide RNAs are correctly identified. (d) Binding sites of the inward primer pair, which can amplify both linear and circular guide RNAs. (e) Fold change of linear and circular split-R/G CLUSTER guide RNAs (V27.2.4, pTS2108) relative to endogenous human *RNU6* snRNA as detected by using the inward primer pair. The 235-fold difference between linear and circular shows the high level of total guide RNA enrichment achieved after circularization. (f) Percentage of successfully circularized guide RNAs calculated from (e). Under the assumption that the 235-fold increase detected for circular guide RNAs compared to linear ones solely depends on circularization, 99.6% of all guide RNAs were successfully processed by the ribozymes of The tornado expression system and then ligated by RTCB. Data in (b) and (e) are shown as the mean editing percentage  $\pm$  s.d. of  $N = 3$  biological replicates.

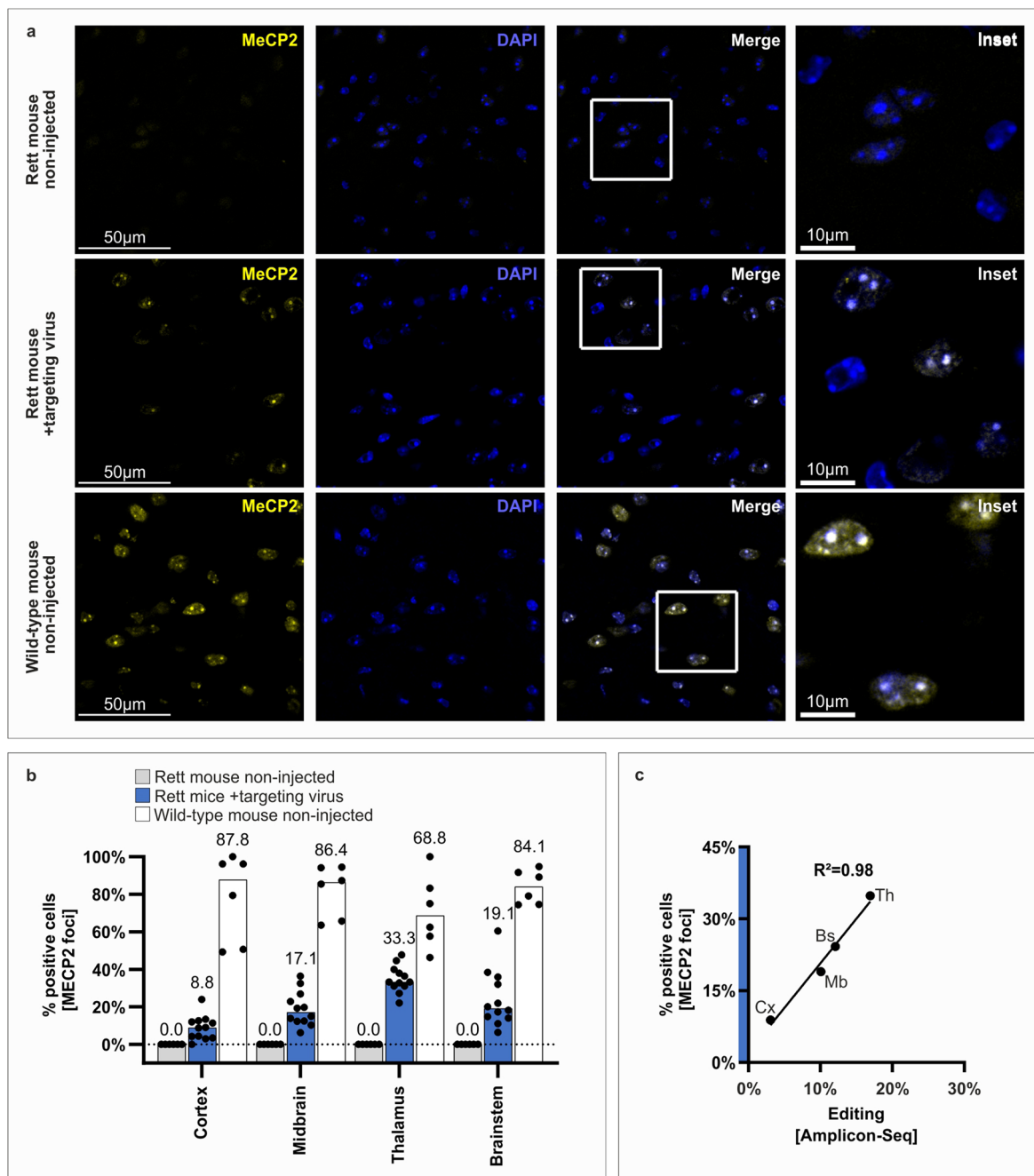






**Extended Data Fig. 8 | Analysis of global off-target events by RNA-seq. (a)** Rett mice injected with wobble-optimized circular CLUSTER guide RNA (targeting virus) versus scrambled guide RNA (nontargeting virus). **(a) (top)** A-to-G RNA editing index in coding sequences. An index of 1 means 1% of RNA nucleotides mapped to genomic adenosines are guanosines. The index was stable (no change in global editing activity due to the targeting guide RNA). **(a) (bottom)** Global RNA editing at 2,528 endogenous sites (coverage  $\geq 50$ , REDportal<sup>55</sup>). With only 3 sites showing > 25% Δ-editing (beyond blue dotted lines), no clear off-target editing was detectable in comparison to panel b. **(b)** Reanalyzed RNA-seq data (previous study<sup>44</sup>). With respect to the tissue selected in panel c), data from the brainstems of two untreated Rett mice are compared. **(b) (top)** Both untreated mice show a stable editing index. **(b) (bottom)** Global RNA editing at 2,301 endogenous sites (coverage  $\geq 50$ ). These sites are also displayed in panel a) (bottom) and c) (bottom). The 7 sites showing > 25% Δ-editing are considered normal mouse to mouse and technical variability. **(c)** Reanalyzed RNA-seq data

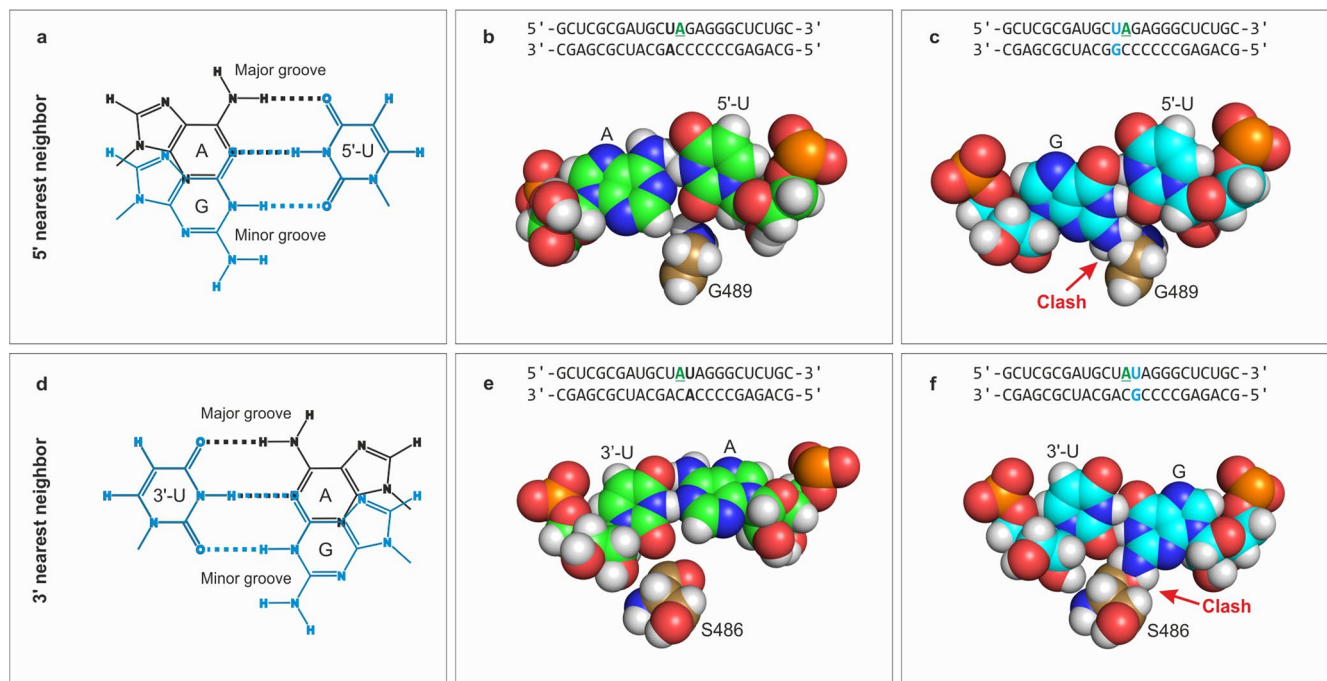
(previous study<sup>44</sup>). Rett mice injected with boxB/λN-ADAR2 virus were compared to untreated ones. The ADAR2 catalytic domain in the virus is from the native ADAR2 protein. **(c) (top)** The A-to-G index increase by 0.007 in the coding space reflects a substantial and somewhat concerning increase in global editing, probably due to the overexpressed ADAR2 catalytic subunit. **(c) (bottom)** Global RNA editing at 2,533 endogenous sites that include the 2,528 sites shown in panel a) (coverage  $\geq 50$  reads). 25 sites showing > 25% Δ-editing after injection of the λN-ADAR2 targeting virus. This is in clear contrast to the very high precision seen for CLUSTER guide RNAs recruiting endogenous ADARs in a) (bottom) or the untreated mice in b) (bottom). Importantly, the potential off-target sites included two ADAR2-specific, highly evolutionary conserved sites in GRM4 (glutamate transmission) and NOVA1 (splicing)<sup>49</sup>. Data in a), b) and c) are based on results from N = 1–2 mice per group. Thalamus and brainstem were selected for showing the highest editing yields in their respective study.



**Extended Data Fig. 9 | Immunohistochemical evidence supporting repair of MeCP2 protein and function after treatment of Rett mice with the circular CLUSTERguide RNA. (a)** Confocal images of sections of thalamus stained for MeCP2 protein and DAPI to indicate nuclei. MeCP2 puncta staining, co-localizing with DAPI, only seen in middle and lower panels indicating restoration of endogenous protein expression. Nuclei (DAPI) were pseudo-colored to highlight colocalization puncta. **(b)** Each bar represents the total percentage of MeCP2 protein foci positive cells relative to DAPI positive cells in each mouse group. Each datapoint represents one microscopic field of view. Non-injected Rett mice show background levels of immunofluorescence<sup>44</sup>. The distribution of positive nuclei in non-injected wild-type mice reflects the cellular heterogeneity within the brain. Analysis was performed by confocal imaging in three sections from

each brain region indicated in the figure. We selected the four brain regions with the highest editing yields. Data is shown as the median percentage of cells showing MeCP2 protein foci  $\pm$  c.i. (confidence interval 95%) from  $N = 2$  mice per bar, as counted in 3 (non-injected Rett mice) or 6 (injected Rett mice) fields of view per animal. **(c)** High correlation between the median editing yields (amplicon sequencing) and the percentage of cells with MeCP2 protein foci in the four brain regions in a). The abbreviations stand for Cortex (Cx), midbrain (Mb), thalamus (Th), and brainstem (Bs). Data is shown as one datapoint per brain region. Each datapoint represents the median result of the respective tissue from  $N = 2$  mice injected with targeting virus per readout. For technical reasons the correlated editing yields and protein foci percentages are from different animals.





**Extended Data Fig. 10 | Proposed mechanisms for the suppressive effect of G-U wobble base pairs on RNA editing when placed at nearest neighbor sites.**

**(a)** Superposition of a 5' G-U wobble base pair (blue) on top of a Watson-Crick AU base pair (black) arranged according to panel b) and c). The shifted Guanine (blue) results in the projection of an exocyclic amino group into the minor groove, which is a distinctive feature of the GU wobble base pair<sup>32</sup>. **(b)** hADAR2d/Bdf2 crystal structure indicating the close proximity of the 5' nearest neighbor AU base pair and Glycine 489. The position of the observed base pair (black, bold) and the target adenosine (green, bold, underlined) is highlighted in the Bdf2 duplex sequence. **(c)** A 5' G-U wobble base pair fitted into the hADAR2d/Bdf2 crystal structure at the 5' nearest neighbor position predicts a steric clash

between the exocyclic amino group of the Guanine and the Glycine 489. The clash would require extensive accommodation of the complex and thus cause structural perturbations, that might negatively affect the deamination rate. The position of the observed wobble base pair (blue, bold) and the target adenosine (green, bold, underlined) is highlighted in the Bdf2 duplex sequence. **(d)** As a) but arranged according to panel e) and f). **(e)** As b) but showing the 3' nearest neighbor AU base pairs proximity to Serine 486. The AU base pair was fitted into the position of the original CG base pair of the hADAR2d/Bdf2 crystal structure. **(f)** As c) but showing the Guanosine of the 3' G-U wobble base pairs clashing with Serine 486. The 3' G-U wobble base pair was fitted into the position of the original CG base pair of the hADAR2d/Bdf2 crystal structure.

Reporting Summary

Nature Portfolio wishes to improve the reproducibility of the work that we publish. This form provides structure for consistency and transparency in reporting. For further information on Nature Portfolio policies, see our [Editorial Policies](#) and the [Editorial Policy Checklist](#).

Statistics

For all statistical analyses, confirm that the following items are present in the figure legend, table legend, main text, or Methods section.

- |                                     |  |
|-------------------------------------|--|
| n/a                                 | Confirmed  |
| <input type="checkbox"/>            | <input checked="" type="checkbox"/> The exact sample size ( <i>n</i> ) for each experimental group/condition, given as a discrete number and unit of measurement   |
| <input type="checkbox"/>            | <input checked="" type="checkbox"/> A statement on whether measurements were taken from distinct samples or whether the same sample was measured repeatedly  |
| <input type="checkbox"/>            | <input checked="" type="checkbox"/> The statistical test(s) used AND whether they are one- or two-sided<br><i>Only common tests should be described solely by name; describe more complex techniques in the Methods section.</i>   |
| <input type="checkbox"/>            | <input checked="" type="checkbox"/> A description of all covariates tested   |
| <input checked="" type="checkbox"/> | <input type="checkbox"/> A description of any assumptions or corrections, such as tests of normality and adjustment for multiple comparisons   |
| <input type="checkbox"/>            | <input checked="" type="checkbox"/> A full description of the statistical parameters including central tendency (e.g. means) or other basic estimates (e.g. regression coefficient) AND variation (e.g. standard deviation) or associated estimates of uncertainty (e.g. confidence intervals) |
| <input type="checkbox"/>            | <input checked="" type="checkbox"/> For null hypothesis testing, the test statistic (e.g. <i>F</i> , <i>t</i> , <i>r</i> ) with confidence intervals, effect sizes, degrees of freedom and <i>P</i> value noted<br><i>Give P values as exact values whenever suitable.</i>                     |
| <input checked="" type="checkbox"/> | <input type="checkbox"/> For Bayesian analysis, information on the choice of priors and Markov chain Monte Carlo settings  |
| <input checked="" type="checkbox"/> | <input type="checkbox"/> For hierarchical and complex designs, identification of the appropriate level for tests and full reporting of outcomes  |
| <input checked="" type="checkbox"/> | <input type="checkbox"/> Estimates of effect sizes (e.g. Cohen's <i>d</i> , Pearson's <i>r</i> ), indicating how they were calculated  |

Our web collection on [statistics for biologists](#) contains articles on many of the points above.

Software and code

Policy information about [availability of computer code](#)

Data collection	The full code of the GuideRNA-Forge tool is available upon request. The compiled tool will be available online prior to publication under: <a href="https://github.com/recruitment-cluster-finder/GuideRNA-FORGE">https://github.com/recruitment-cluster-finder/GuideRNA-FORGE</a>
Data analysis	Data were analyzed using Excel 2016 and GraphPad Prism 8, figures were created with CorelDraw 2017, the manuscript was written with Word 2016, the custom "GuideRNA Forge" tool was written in python (version 3.9) and used the Vienna RNA package (version 2.0) ( <a href="https://www.tbi.univie.ac.at/RNA/">https://www.tbi.univie.ac.at/RNA/</a> ), gRNA folds were created using the RNAfold web server (University of Vienna, <a href="http://rna.tbi.univie.ac.at/cgi-bin/RNAWebSuite/RNAfold.cgi">http://rna.tbi.univie.ac.at/cgi-bin/RNAWebSuite/RNAfold.cgi</a> ), qPCR analysis was performed using the 7500 data analysis software version 1.1 (animal experiment) and version 2.3 (rest of the study), structural analysis was performed using the open-source PyMOL™ Molecular Graphics System (version 2.5.0), Sanger sequence traces were analyzed using SNAP-Gene (version 4.2.11), microscopy images were acquired using ZEN 2.3 SPI FPS Black (version 14.0.0.0) and pseudo colored using ImageJ (version 1.54f), RNA-seq. reads were demultiplexed using Illumina bcl2fastq (version 2.20), RNA-seq. adapters were trimmed with Skewer (version 0.2.2), RNA-seq. FASTQ file quality was assessed with FastQC (version 0.11.8), RNA-seq. reads were aligned to the GRCm38/mm10 reference genome using STAR (version 2.7.3a), Deep amplicon-seq. reads were demultiplexed using BCL Convert (version 2.4.0), Deep amplicon-seq. adapters were trimmed with FastQ (version 1.0.0), Deep amplicon-seq. FASTQ file quality was assessed with FastQC (version 0.11.9), Deep amplicon-seq. reads were quality trimmed and their low quality bases masked using Seqtk (version 1.3-r106),

Deep amplicon-seq. reads were aligned to the GRCm38/mm10 reference genome using BWA-MEM (version 0.7.17-r1188),  
Deep amplicon-seq. alignments were analyzed using the Integrative Genomics Viewer (version 2.16.2).

For manuscripts utilizing custom algorithms or software that are central to the research but not yet described in published literature, software must be made available to editors and reviewers. We strongly encourage code deposition in a community repository (e.g. GitHub). See the Nature Portfolio [guidelines for submitting code & software](#) for further information.

## Data

Policy information about [availability of data](#)

All manuscripts must include a [data availability statement](#). This statement should provide the following information, where applicable:

- Accession codes, unique identifiers, or web links for publicly available datasets
- A description of any restrictions on data availability
- For clinical datasets or third party data, please ensure that the statement adheres to our [policy](#)

Transcriptome-wide RNA-seq data are accessible via the NCBI GEO database with accession code GSE265898. Deep amplicon sequencing data are accessible via the NCBI Sequence Read Archive (SRA) database with accession code PRJNA1100948. Transcriptome-wide RNA-seq data from Sinnamon et al. are accessible via the NCBI SRA database with accession code PRJNA849938. Structure data from Thuy-Boun et al. and Stefl et al. are available with Protein Data Bank ID 5HP2 and 2L2J, respectively.

## Research involving human participants, their data, or biological material

Policy information about studies with [human participants or human data](#). See also policy information about [sex, gender \(identity/presentation\), and sexual orientation](#) and [race, ethnicity and racism](#).

Reporting on sex and gender

Reporting on race, ethnicity, or other socially relevant groupings

Population characteristics

Recruitment

Ethics oversight

Note that full information on the approval of the study protocol must also be provided in the manuscript.

## Field-specific reporting

Please select the one below that is the best fit for your research. If you are not sure, read the appropriate sections before making your selection.

☒ Life sciences ☐ Behavioural & social sciences ☐ Ecological, evolutionary & environmental sciences

For a reference copy of the document with all sections, see [nature.com/documents/nr-reporting-summary-flat.pdf](https://nature.com/documents/nr-reporting-summary-flat.pdf)

## Life sciences study design

All studies must disclose on these points even when the disclosure is negative.

Sample size	Experiments for evaluating editing yields via Sanger sequencing were mostly done in n=3 biological replicates in rare cases in n=4 or n=5. When possible (in bar diagrams or box plots) data points are displayed individually. In case of heat maps (in-vitro experiments) the mean $\pm$ s.d. was displayed. The evaluation of the editing yields via Sanger sequencing for the animal experiment was done with n=3-5 mice per group, with each indicated brain region being measured individually as one technical replicate. Amplicon-seq. and RNA-seq. for the animal experiment were done with n=2 mice per group. In case of heat maps (in-vivo experiment) the median $\pm$ 95% confidence interval was displayed. qPCR quantifications for the animal experiment were done with n=2 mice and n=3 technical replicates per tissue per mouse. The evaluation of the fold change via the dual-luciferase assay for cell culture experiments was done with n=5 biological replicates with one technical measurement each. No sample size calculation was performed. The sample sizes for cell culture experiments were selected based on the prior knowledge of variation. The group sizes for animal experiments were selected based on "Sinnamon, J. R., et al. (2022). "Targeted RNA editing in brainstem alleviates respiratory dysfunction in a mouse model of Rett syndrome." Proc Natl Acad Sci U S A 119(33): e2206053119."
Data exclusions	One of the non-targeting CLUSTER gRNA samples from the next-generation RNA-sequencing experiment failed quality control (RIN value 1) and thus had to be removed from the analysis.
Replication	All experiments could be reproduced, as shown in the manuscript, the number of replications and nature of replicates is always given in the figure caption.



## Randomization

All samples were treated according to the same protocols side-by-side with the respective controls and thus, there was no requirement for randomization.

## Blinding

No blinding was performed due to the involvement of several experimentators and several laboratories.

## Reporting for specific materials, systems and methods

We require information from authors about some types of materials, experimental systems and methods used in many studies. Here, indicate whether each material, system or method listed is relevant to your study. If you are not sure if a list item applies to your research, read the appropriate section before selecting a response.

### Materials & experimental systems

n/a	Involved in the study
<input type="checkbox"/>	<input checked="" type="checkbox"/> Antibodies
<input type="checkbox"/>	<input checked="" type="checkbox"/> Eukaryotic cell lines
<input checked="" type="checkbox"/>	<input type="checkbox"/> Palaeontology and archaeology
<input type="checkbox"/>	<input checked="" type="checkbox"/> Animals and other organisms
<input checked="" type="checkbox"/>	<input type="checkbox"/> Clinical data
<input checked="" type="checkbox"/>	<input type="checkbox"/> Dual use research of concern
<input checked="" type="checkbox"/>	<input type="checkbox"/> Plants

### Methods

n/a	Involved in the study
<input checked="" type="checkbox"/>	<input type="checkbox"/> ChIP-seq
<input checked="" type="checkbox"/>	<input type="checkbox"/> Flow cytometry
<input checked="" type="checkbox"/>	<input type="checkbox"/> MRI-based neuroimaging

## Antibodies

## Antibodies used

MeCP2 (D4F3) XP Rabbit mAb (Monoclonal IgG, Cell Signaling Technology, #3456S), Donkey  $\alpha$  Rabbit AlexaFluor 488 (Polyclonal IgG, Invitrogen, A21206)

## Validation

MeCP2 (D4F3) XP Rabbit mAb (Commonly used commercial antibody validated in KO mouse tissue sections previously in our lab. RRID: AB\_2143849, PMID: 35939700)

Donkey  $\alpha$  Rabbit AlexaFluor 488 (Commonly used commercial antibody with 135 IHC-IF citations as of Jan. 2024. RRID: AB\_2535792 Recent publications: PMID: 36865524, PMID: 36321664, PMID: 38017073)

## Eukaryotic cell lines

Policy information about [cell lines and Sex and Gender in Research](#)

## Cell line source(s)

HeLa (cat. no. CCL-2, ATCC)  
HEK-Flp-In T-Rex-A1p110, A1p150, A2 (cat. no. R78007, Thermo Fisher scientific, stably transfected with ADAR1 p110 vector in our lab)  
293FT (cat. no. R70007, Thermo Fisher scientific)

## Authentication

Authentication via STR profiling was performed by the commercial suppliers before purchase of the material. Cell lines were not additionally authenticated by us.

## Mycoplasma contamination

The HeLa cell line has been tested as mycoplasma-free by the commercial suppliers and in house. 293FT and HEK-Flp-In T-Rex (A1p110, A1p150, A2) cell lines were purchased mycoplasma-free, but not tested for mycoplasma in-house.

Commonly misidentified lines  
(See [ICLAC](#) register)

None were used.

## Animals and other research organisms

Policy information about [studies involving animals](#); [ARRIVE guidelines](#) recommended for reporting animal research, and [Sex and Gender in Research](#)

## Laboratory animals

Species: mouse, strain: C57BL/6, disease: Rett syndrome caused by Mecp2 311G>A mutation.  
All groups consisted of hemizygous male mice.  
The animals were between P30 and P34 at the time of AAV treatment.

## Wild animals

No wild animals were used in this study.

## Reporting on sex

The findings in this study apply to male mice. Their sex was identified via PCR (see methods section). Female homozygous Rett mice (Mecp2<sup>-/-</sup>) are hard to obtain due to the inability to use hemizygous Rett males (Mecp2<sup>-/y</sup>) as breeders as they have a decreased life span and Rett-like symptoms that begin to emerge early in life. Homozygous females are also not a model for the human disease because the human Rett females are heterozygous. The hemizygous males are the optimal choice for our experimentation because

we are able to accurately study the Mecp2 RNA editing efficiency and protein recovery without worrying about the mosaicism that Mecp2+/- females demonstrate.

Field-collected samples No field-collected samples were used in this study.

Ethics oversight All animal procedures were approved by the Institutional Animal Care and Use Committees of Oregon Health and Science University (IACUC protocol number: IP00000284).

Note that full information on the approval of the study protocol must also be provided in the manuscript.

## Plants

Seed stocks No plants were used in this study.

Novel plant genotypes No plants were used in this study.

Authentication No plants were used in this study.