HyperTRIBE: Upgrading TRIBE with enhanced editing

Weijin Xu, Reazur Rahman, Michael Rosbash

Department of Biology, Howard Hughes Medical Institute and National Center for Behavioral Genomics, Brandeis University, Waltham, Massachusetts, 02453 USA.

Corresponding author: Michael Rosbash, rosbash@brandeis.edu

We recently developed a RNA-binding protein target identification tool, TRIBE, which expresses an RBP of interest fused to the catalytic domain (cd) of the RNA editing enzyme ADAR. Here we describe HyperTRIBE, which carries a hyperactive mutation (E488Q) of the ADAR catalytic domain. HyperTRIBE identifies dramatically more targets and more faithfully recapitulates the known binding specificity of its RBP than TRIBE.

We recently developed TRIBE (Targets of RNA-binding proteins Identified By Editing) to study RBP targets in specific cells. TRIBE expresses an RBP of interest fused to the catalytic domain (cd) of the RNA editing enzyme ADAR and performs Adenosine-to-Inosine editing on RNA targets of the RBP. However, target identification is limited by the low editing efficiency of the TRIBE ADARcd. Here we describe HyperTRIBE, which carries a hyperactive mutation (E488Q) of the ADARcd, identified in a yeast screen[1]. HyperTRIBE identifies dramatically more editing sites, many of which correspond to below-threshold TRIBE editing targets. HyperTRIBE also overlaps more successfully with CLIP data and therefore more faithfully recapitulates the known binding specificity of its RBP than TRIBE.

ADAR consists of two modular parts, double-strand RNA-binding motifs (dsRBMs) and a catalytic domain[2-4]. TRIBE replaces the dsRBMs of the *Drosophila* ADAR enzyme with the RBP of interest[5]. Fusion protein binding and mRNA specificity are therefore driven predominantly by the RBP in the absence of the dsRBMs. TRIBE-dependent editing sites are identified by deep sequencing of RNA extracted from cells of interest. We were successful in identifying bona fide targets of three RBPs, but the number of targets identified by TRIBE in tissue culture was substantially reduced compared to CLIP data with the same RBP[5]. Even assuming some CLIP false positives[6,7], TRIBE may still be experiencing a high false-negative problem.

We therefore turned our attention to a “hyperactive” E488Q mutation within the cd of human ADAR2, which was identified in a yeast screen[1]. To make the companion HyperTRIBE construct, we first identified the dADARcd glutamate corresponding to human ADAR2 E488. That residue as well as surrounding sequence is highly conserved

between hADAR2 and the dADARcd (data not shown), suggesting that it should function similarly in the *Drosophila* enzyme. We then introduced the E488Q mutation into the original Hrp48 TRIBE construct via QuikChange® site-directed mutagenesis. We had no difficulty making a stable S2 cell line expressing Hrp48 TRIBE[5] but failed with Hrp48 HyperTRIBE (henceforth called HyperTRIBE), perhaps because of the greatly enhanced editing frequency (see below). Most experiments were therefore performed by transiently expressing fusion proteins in *Drosophila* S2 cells together with GFP and sorting GFP-positive cells by FACS. Editing sites were always defined as the sites in common between two experiments, i.e., >10% editing and > 20 reads/nucleotide in both biological replicates[5]. A third replicate of TRIBE minimally reduced the number of common sites, i.e., 80% were present in the third replicate (data not shown). Transient and stable expression of TRIBE in S2 cells detected comparable target gene numbers (200-300), ~40% of which were identical (Fig. S1). This rather low rate of overlap is likely due to the substantial difference between transient and stably expressing cells.

Expression of HyperTRIBE resulted in approximately 20X the number of editing events and approximately 8X the number of edited genes compared to TRIBE (Fig. 1a). Expression of the ADARcd with E488Q mutation alone does not increase the number of editing events above the endogenous level of S2 cells (Fig. 1a). This is despite the fact that the HyperADARcd is stable and expressed at comparable levels to those of the other TRIBE constructs (data not shown), so most if not all editing by HyperTRIBE -- like editing by regular TRIBE -- requires the RNA binding ability of the fused RBP[5].

We next compared the HyperTRIBE editing data with our previous CLIP data as well as with regular TRIBE editing data in *Drosophila* S2 cells. HyperTRIBE not only

identifies 282 (97%) of the edited sites and 220 (98%) of the edited genes identified by TRIBE (Fig. 1b, 1c), but these data also correlate well with the CLIP results: 73% of the HyperTRIBE sites are identified by CLIP. (78% for the TRIBE sites; Fig. 1c). However, there is a striking difference of 66% vs 4% in the ability of HyperTRIBE and TRIBE to recognize CLIP-identified genes, respectively (Fig. 1c and S2). This distinction is due to the many fewer sites and genes identified by TRIBE, indicating that HyperTRIBE significantly lowers the TRIBE false negative rate and thereby provides a much more complete binding signature of the RBP. We obtained similar results with FMRP HyperTRIBE (data not shown), another RBP assayed in the original TRIBE paper, indicating that the higher efficiency of HyperTRIBE is not limited to Hrp48.

The ratio of HyperTRIBE:TRIBE edited genes (8) is two to three-fold lower than the editing site ratio (20), indicating a substantial increase in the number of edited sites per gene in HyperTRIBE. Indeed, HyperTRIBE generates many more multiple-edited genes with a median of 3 edited sites per gene comparing to a median of 1 for TRIBE (Fig. 1d). Some HyperTRIBE editing sites are near the original TRIBE sites (Fig. 1e), suggesting that the higher editing rate of HyperTRIBE is due in part to its ability to edit multiple adenosines near the original ADARcd interacting region. Moreover, the data show that HyperTRIBE unique editing sites that are on the same molecule as common sites – within a single RNA-seq read – have a higher editing percentage than all unique editing sites (Fig. S3). This indicates that the HyperADARcd may edit additional, nearby adenosines without fully releasing the mRNA.

Another source of additional editing is a lower neighboring sequence preference surrounding the edited adenosines of HyperTRIBE (Fig. 1f). This sort of preference is

85 nonnegligible for CLIP[8] as well as for TRIBE[5] which leads to bias in the targets
86 identified.

87 To further address possible reasons for the additional HyperTRIBE editing, we
88 compared the editing frequencies of the sites identified by both HyperTRIBE and TRIBE.
89 We divided the HyperTRIBE editing sites into two categories, unique sites and common
90 sites; the latter were also identified by TRIBE. Although HyperTRIBE edits its unique
91 sites at similar frequencies to the much smaller number of TRIBE-edited sites, the
92 common sites experience much higher editing frequencies with HyperTRIBE than with
93 TRIBE (Fig. 2a). This is true not only on average but also when the sites are examined
94 individually (Fig. 2b). This indicates that common sites are also preferred HyperTRIBE
95 substrates, possibly due to stronger RBP binding.

96 Is it possible that the unique sites are also quantitatively rather than qualitatively
97 different between HyperTRIBE and TRIBE? This suggests that many of them might be
98 edited by TRIBE but below the required threshold. Indeed, there are 4017 adenosines that
99 meet this criterion, i.e., at least one editing event for each specific adenosine in both
100 replicates but with less than 10% editing percentage (Fig. 2c). The correspondence
101 between replicates for these adenosines and HyperTRIBE editing sites is highly
102 significant (Z-test performed, p-value≈0), and more than 30% of these 4017 adenosines
103 correspond to HyperTRIBE editing sites. This number is much greater than the number of
104 C-to-T editing/mutations, which serve as control events (data not shown). Notably, these
105 4017 below-threshold editing sites occur in 3473 different genes, which overlap well with
106 genes identified by Hrp48 HyperTRIBE and Hrp48-ADARcd CLIP (Fig. 2d). We
107 conclude that below-threshold TRIBE-edited adenosines make a significant contribution

to the extra HyperTRIBE editing sites and that much of the distinction between HyperTRIBE and TRIBE is quantitative.

A conservative threshold of 10% editing was initially established to ensure that the TRIBE assay was not impacted by substantial numbers of false positives[5]. We began this HyperTRIBE study with the same 10% threshold, so that the new data could be directly compared with our previous TRIBE data. HyperTRIBE identifies many more sites and genes above this threshold, consistent with the original characterization of the E488Q mutation in yeast[1].

Nonetheless, the fraction of edited adenosines is low even for HyperTRIBE, presumably still reflecting the sequence and structural requirements of the ADARcd, i.e., the nearest neighbor sequence preference and the double-stranded character surrounding a bulged adenosine. Although both of these requirements are reduced for HyperTRIBE compared to TRIBE (Fig. 1f and data not shown), the major reason for the enhanced HyperTRIBE editing is probably the increased editing rate by the E488Q mutation. It either enhances base flipping, perhaps by enhanced amino acid insertion by the HyperTRIBE ADARcd into the RNA A helix, or it has a much higher rate of successful editing/flipping event[1,9]. Notably, the ADARcd loop that occupies the displaced A during flipping[9] contains amino acid 488. Better binding of the HyperTRIBE ADARcd to its substrate region might enhance the transition state lifetime and therefore editing efficiency[1,9].

We were surprised that so many of the new HyperTRIBE > 10% editing sites correspond to proper TRIBE editing sites but with editing frequencies well below the 10% threshold. The correspondence of these below threshold sites not only with

131 HyperTRIBE sites but also between replicate TRIBE experiments further indicates that
132 they are genuine editing sites. However, most of these TRIBE sites would require
133 substantial sequencing depth and expense to be identified. This can probably be avoided
134 by using HyperTRIBE, which is also a superior approach for the identification of RBP
135 targets in mammalian cells as well as neurons (data not shown).

Methods

Methods, including data availability information are at the end of the manuscript.

Acknowledgements

Special thanks are due to Aoife McMahon, a prior lab member, for her important input into this work. We thank Joshua Rosenthal, Brenda Bass, Amy Lee and Aoife McMahon for helpful comments on a very early version of this manuscript as well as current Rosbash lab members for comments and discussion. The work was supported by the Howard Hughes Medical Institute and by a NIH EUREKA grant (DA037721).

Competing interests

The authors declare that no competing interests exist.

Author Contributions

Conceptualization, W.X., M.R. and Aoife McMahon; Investigation, W.X.; Software, R.R.; Formal Analysis, W.X. and R.R.; Visualization, W.X. and R.R.; Writing – Original Draft, W.X. and M.R.; Funding Acquisition, M.R.

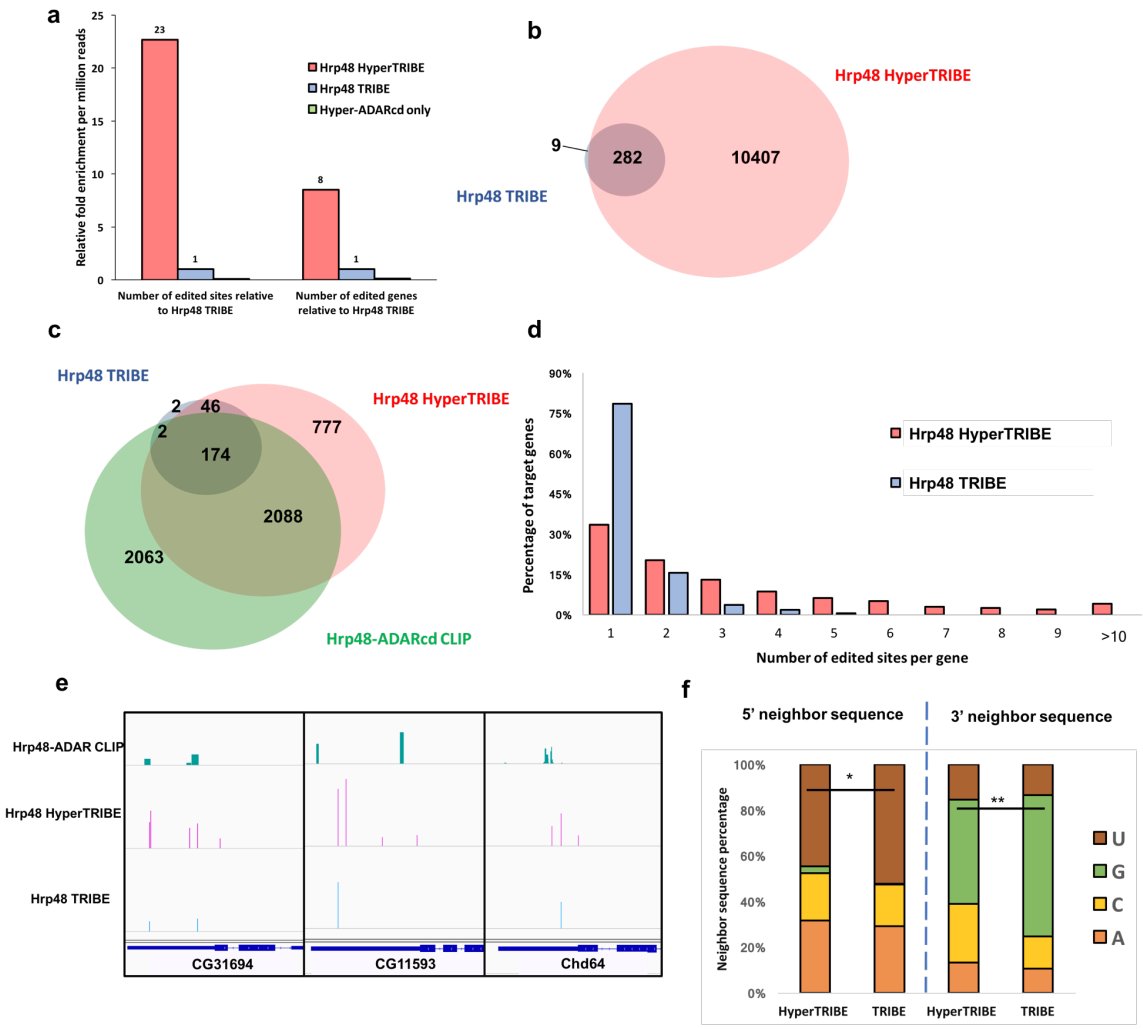

Figure 1. HyperTRIBE dramatically increases the number of targets identified compared to TRIBE

(a) Although HyperTRIBE and TRIBE increases in both the number of editing events and genes edited in S2 cells the increases are much more dramatic in HyperTRIBE-expressing cells. Numerically, HyperTRIBE identifies 10689 common edited sites in two replicates, whereas the corresponding number for Hrp48 TRIBE is 291. Both TRIBE constructs have a reproducibility of around 60% and only replicable editing events are reported. There is no comparable increase in editing sites or genes with expression of the Hyper-ADARcd is alone (11 sites identified). The number of genes and sites identified are normalized to the sequencing depth of each sample and are measured by relative fold change compared to TRIBE (see also Methods).

(b) Venn diagram of the editing sites shows that almost all of the Hrp48 TRIBE sites (blue) are also detected by HyperTRIBE (pink).

(c) Both TRIBE-and HyperTRIBE-identified genes overlap well with CLIP-identified genes. The Venn diagram shows the overlap of all genes identified by TRIBE (224 in total, blue), HyperTRIBE (3085 in total, pink) and Hrp48-ADARcd CLIP (4327 in total, green)[5]. The overlap between TRIBE and CLIP is not significantly different from the overlap between HyperTRIBE and CLIP (Z-test performed, $p=0.09$).

(d) A much larger fraction of target genes are edited at multiple sites by HyperTRIBE than by TRIBE. The histogram indicates the percentage of target genes containing 1 to more than 10 editing sites. Genes with multiple sites may be transcripts bound more stably by Hrp48.

(e) Three examples of commonly identified genes by HyperTRIBE, TRIBE and CLIP are shown in the IGV genome browser. Hrp48-ADARcd CLIP data is from a previous publication[5]. Typically, the multiple editing sites in HyperTRIBE cluster nearby the original sites identified by TRIBE. The height of the bars indicates editing frequency for TRIBE data and CLIP signal strength for CLIP data.

(f) 5' and 3' immediate neighbor sequence preference is reduced in HyperTRIBE. ADAR shows a neighbor preference of uridine at the 5' and guanosine at the 3' side. This preference is lower in HyperTRIBE than in TRIBE (Z-test performed, $*p<0.05$, $**p<0.0001$, percentage of uridine in 5' neighbor and percentage of guanosine in 3' neighbor tested, HyperTRIBE vs TRIBE).

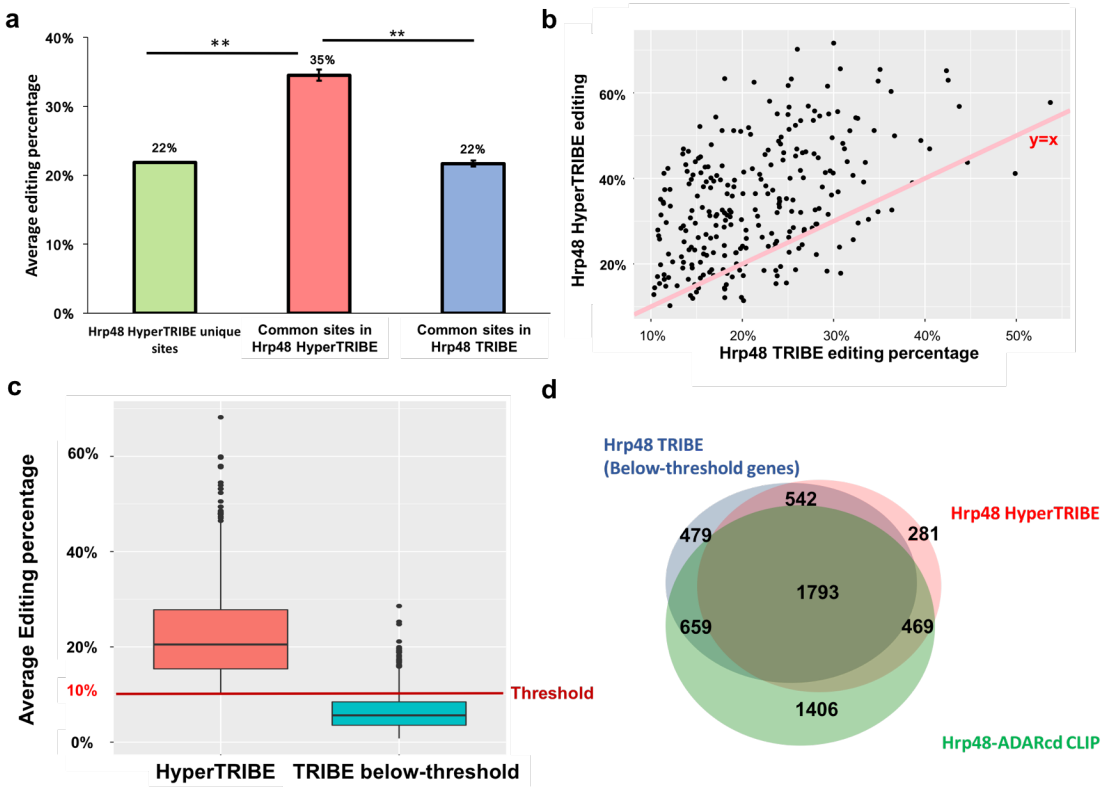

Figure 2. HyperTRIBE increases the editing frequency, which allows the detection of many below-threshold TRIBE sites

(a) Bar graph shows the editing percentage of the sites uniquely identified in HyperTRIBE as well as the editing percentage of the common editing sites in HyperTRIBE and in TRIBE, respectively. Editing percentage is shown as the weighted average. The editing percentage of the sites commonly identified by TRIBE and HyperTRIBE is increased in HyperTRIBE (Student t-test performed, **p<0.0001).

(b) An increase in editing percentage by HyperTRIBE is observed for most sites. A scatter plot shows editing percentage of HyperTRIBE editing sites (Y-axis) and TRIBE editing sites (X-axis). Sites identified by both TRIBE and HyperTRIBE are shown. Line represents y=x is shown for reference (pink).

(c) About 30% of the 4017 below-threshold sites in TRIBE show an elevated editing percentage in HyperTRIBE, allowing them to be identified in the analysis pipeline (Z-test performed to test enrichment, $P \approx 0$). HyperTRIBE editing sites that are present in TRIBE but below the 10% threshold are shown. Some below-threshold TRIBE editing sites have $> 10\%$ average editing

209 percentage because one replicate has >10% editing but the other replicate has
210 <10%. A red line indicating 10% threshold is shown for reference.
211 (d) 67% and 71% of TRIBE below-threshold target genes overlap with
212 HyperTRIBE and CLIP respectively. Venn diagram shows the overlap of genes
213 edited below-threshold by TRIBE (3473 in total, blue), genes identified by
214 HyperTRIBE (3085 in total, pink) and Hrp48-ADARcd CLIP genes (4327 in
215 total, green)[5]. Although small, the difference in overlap between below-threshold
216 TRIBE genes and CLIP (71%) versus overlap between HyperTRIBE genes and
217 CLIP (73%) is significant (Z-test performed, p=0.015).

1 Kuttan, A. & Bass, B. L. Mechanistic insights into editing-site specificity of ADARs. *Proceedings of the National Academy of Sciences of the United States of America* **109**, E3295-3304, doi:10.1073/pnas.1212548109 (2012).

2 Macbeth, M. R. *et al.* Inositol hexakisphosphate is bound in the ADAR2 core and required for RNA editing. *Science* **309**, 1534-1539, doi:10.1126/science.1113150 (2005).

3 Kim, U., Wang, Y., Sanford, T., Zeng, Y. & Nishikura, K. Molecular cloning of cDNA for double-stranded RNA adenosine deaminase, a candidate enzyme for nuclear RNA editing. *Proceedings of the National Academy of Sciences of the United States of America* **91**, 11457-11461 (1994).

4 O'Connell, M. A. *et al.* Cloning of cDNAs encoding mammalian double-stranded RNA-specific adenosine deaminase. *Molecular and cellular biology* **15**, 1389-1397 (1995).

5 McMahon, A. C. *et al.* TRIBE: Hijacking an RNA-Editing Enzyme to Identify Cell-Specific Targets of RNA-Binding Proteins. *Cell* **165**, 742-753, doi:10.1016/j.cell.2016.03.007 (2016).

6 Lambert, N. *et al.* RNA Bind-n-Seq: quantitative assessment of the sequence and structural binding specificity of RNA binding proteins. *Molecular cell* **54**, 887-900, doi:10.1016/j.molcel.2014.04.016 (2014).

7 Darnell, R. B. HITS-CLIP: panoramic views of protein-RNA regulation in living cells. *Wiley interdisciplinary reviews. RNA* **1**, 266-286, doi:10.1002/wrna.31 (2010).

8 Sugimoto, Y. *et al.* Analysis of CLIP and iCLIP methods for nucleotide-resolution studies of protein-RNA interactions. *Genome biology* **13**, R67, doi:10.1186/gb-2012-13-8-r67 (2012).

9 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, doi:10.1038/nsmb.3203 (2016).

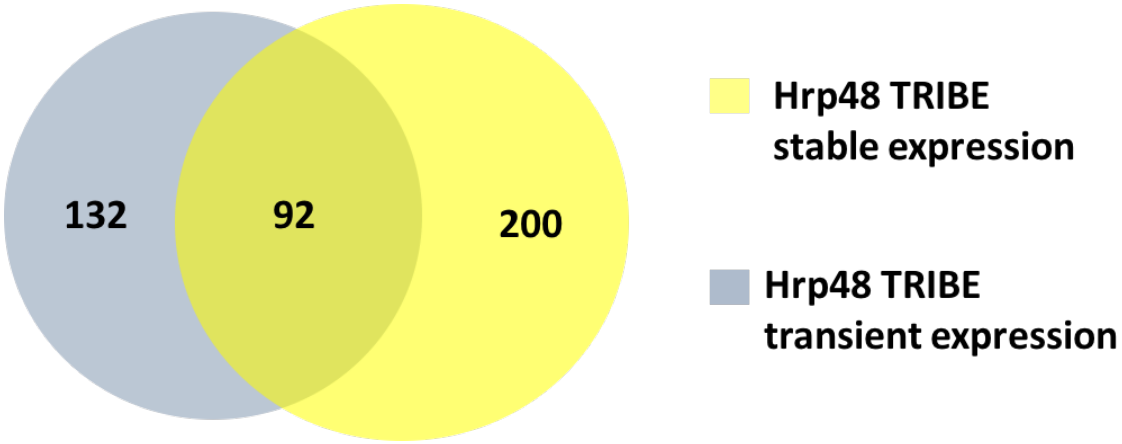

Figure S1. Comparison of Hrp48 TRIBE target genes between stable and transient expression

Venn diagram shows the overlap of genes identified by expressing TRIBE in S2 cells transiently (224 in total, blue) and stably (292 in total yellow).

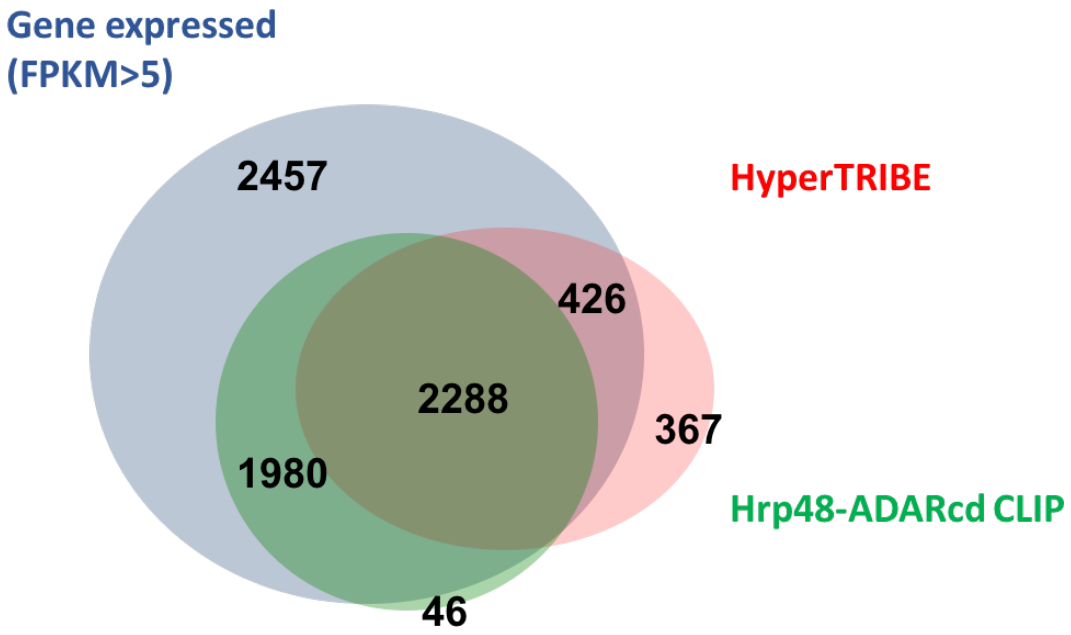

Figure S2. Genes identified by HyperTRIBE overlap well with CLIP-identified genes

Venn diagram shows the overlap of genes identified by HyperTRIBE (3085 in total, pink), Hrp48-ADARcd CLIP (4327 in total, green) and all expressed genes in the cells (7151 in total, blue). HyperTRIBE-identified genes are significantly enriched for CLIP-positive genes (Z-test performed, $p<0.0001$, ratio of overlap with HyperTRIBE genes for CLIP-positive and CLIP-negative genes compared), indicating specificity.

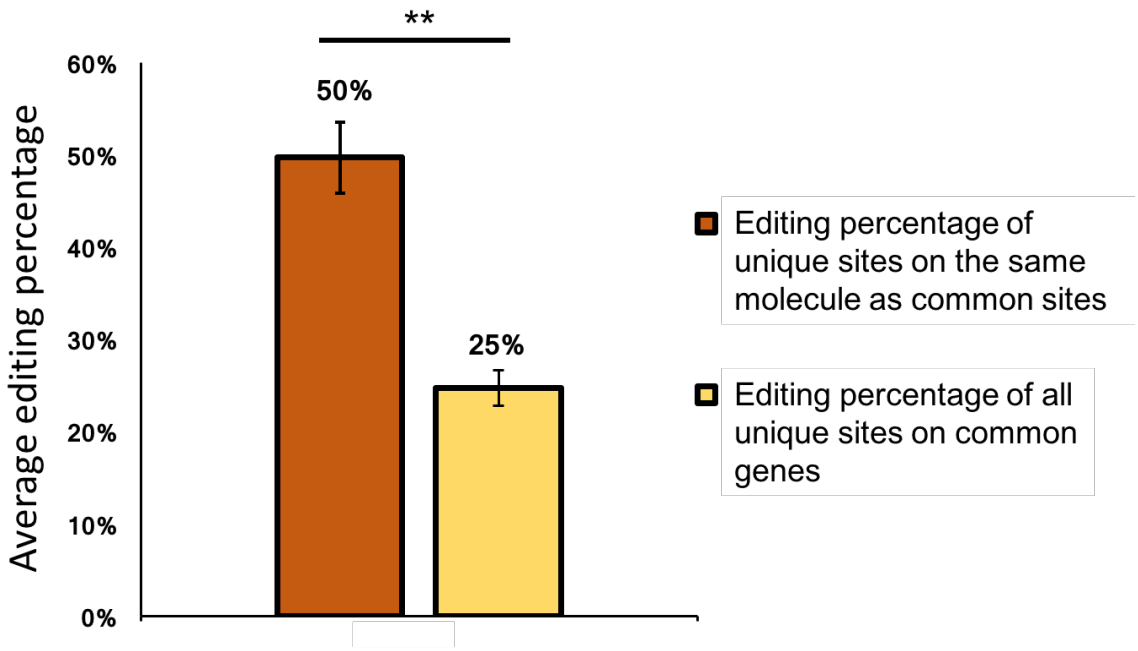

Figure S3. HyperTRIBE editing retains the 3' UTR binding preference of Hrp48 protein

Both TRIBE methods show 3'UTR editing preference, although the preference is lower for HyperTRIBE. Editing preference is calculated as the distribution of editing sites in each mRNA region normalized to the distribution of sequence reads in each mRNA region. (n.s.=$p>0.05$, **$p<0.001$, Z-test performed HyperTRIBE compared against TRIBE)

Methods

Molecular Biology

RBP-ADARcd with E488Q mutation was created by performing Quikchange® Site-directed Mutagenesis on pMT-RBP-ADARcd-V5 plasmid[5]. Primers 5'-TCGAGTCCGGTCAGGGGACGATTCC and 5'-GGAATCGTCCCCTGACCGGACTCGA were used to induce point mutation to the underlined nucleotide. Transient expression of TRIBE constructs was performed by co-transfecting pMT TRIBE plasmids with pActin-EGFP to *Drosophila* S2 cells using Cellfectin® II from Thermo Fisher Scientific. Cells were allowed 48 hours after transfection for adequate expression of GFP before sorting with BD FACSAria™ II machine for GFP positive cells. Total RNA was extracted from the sorted cells with TRIzol™ LS reagent. TRIBE protein expression was induced with copper sulfate 24 hours before FACS sorting. Expression of all fusion proteins was assayed by transient expression in S2 cells and western blot against V5 tag (Invitrogen, 46-1157). TRIBE stable cell lines used in this paper are the same as the original TRIBE paper[5].

Standard Illumina TruSeq® RNA library Kit was used to construct RNA-seq library from S2 cells. All libraries were sequencing by Illumina NextSeq® 500 sequencing system using NextSeq® High Output Kit v2 (75 cycles). Each sample were covered by ~20 million raw reads.

RNA-editing Analysis

The criteria for RNA editing events were: 1) The nucleotide is covered by a minimum of 20 reads in each replicate; 2) More than 80% of genomic DNA reads at this nucleotide is A with zero G (use the reverse complement if annotated gene is in the reverse strand); 3) A minimum of 10% G is observed at this site in mRNA (or C for the reverse strand). Genomic DNA of S2 cell is sequenced to identify and exclude possible polymorphism on DNA level. RNA sequencing data were analyzed as previously described[5,10], with minor modifications. Background editing sites found in samples expressing Hyper-ADARcd alone were subtracted from the TRIBE identified editing sites. Overlap of editing sites from two datasets was identified using "bedtools intersect" with parameters "-f 0.9 -r".

Data availability

The accession number for the raw sequencing data and processed RNA-editing tracks reported in this paper is NCBI GEO: GSE102814. The scripts used in the analysis are available on GitHub (https://github.com/rosbashlab/TRIBE).

10 Rodriguez, J., Menet, J. S. & Rosbash, M. Nascent-seq indicates widespread cotranscriptional RNA editing in Drosophila. *Molecular cell* **47**, 27-37, doi:10.1016/j.molcel.2012.05.002 (2012).