Supplemental experimental procedures

Cloning

TRIBE component cDNA clones were acquired from various sources; Hrp48 (P48809) from Donald Rio; dFMR1 (AF305881) from Tom Jongens, dADAR from Robert Reenan (Palladino et al., 2000), NonA (clone RE58280, Drosophila Genomics Resource Centre, Bloomington, IN). Cytidine deaminases assayed were mouse APOBEC1 (BC003792), yeast CDD1 (Gene ID 850946), putative Drosophila cytidine deaminases CG5292 (FBgn0038491) and CG6951 (FBgn0036959), human APOBEC3A (Sharma et al., 2015) and plant QED1 and RARE1 (Wagoner et al., 2015). Mutations of the two RNA recognition motifs of Hrp48 (P48809, F50V, F52V and F139V, F141V) were modelled after mutations performed by Mayeda et al. (1994) on hnRNP A1.

Cell sorting

A detailed description of the cell sorting and RNA-seq library generation protocol is provided in Abruzzi et al. (2015). Briefly, between 20 and 50 fly brains were dissected in the presence of tetrodotoxin only (0.1uM), digested in L-cysteine-activated papain (Worthington) for 20 minutes, before manual dissociation using fire-polished pipette tips. Dissociated cells are allowed to settle in a large volume of sorting medium on Sylgard coated plates. Using a dissecting microscope fluorescently labelled neurons are then manually aspirated using micropipettes pulled from capillary tubes (World Precision Glass Capillaries #1B100-4, Sutter Instrument Company). Three rounds of sorting are performed until approximately 150-400 fluorescent cells are sorted into 50ul of Lysis/Binding buffer (Invitrogen; Dynabeads mRNA direct kit) and frozen at − 80 °C.

Generation of sequencing libraries from manually sorted neurons

Generation of sequencing libraries was performed as is described in Abruzzi et al. (2015). We estimate we gather between 200-500 pg of total or ~5 pg of mRNA from ~100 manually sorted neurons. mRNA is isolated on poly dT beads (Dynabeads mRNA direct kit; Invitrogen) using a scaled down version of the manufacturer's protocol, and concentrated by vacuum centrifugation

(SpeedVac RC1010, Jouan, Winchester, VA). A modified linear amplification method is then used to amplify this RNA; reverse transcription using random hexamers and dT primers containing T7 promoters generates a cDNA template that is used for one round of *in vitro* transcription (MEGAscript kit, Ambion). The cRNA is isolated (RNA MinElute column, Qiagen), concentrated by vacuum centrifugation, and used as input for the RNA Tru-Seq library generation kit (Illumina). Resulting libraries are somewhat 3' biased and approximately 20-60% of reads map to the fly genome (as discussed in Abruzzi et al. (2015)). A table of reads and mapped reads is supplied in the Table S3. All sequencing was performed on either an Illumina HiSeq or MiSeq.

Analysis of sequencing data for RNA editing.

Analysis of RNA sequencing data was performed as previously published Rodriguez et al. (2012), with some modifications. All raw sequencing data and identified RNA editing sites are available for download from NCBI Gene Expression Omnibus (http://www.ncbi.nlm.nih.gov/geo/), accession number (GSE78065). (Except yw gDNA which is in Series GSE37232, sample GSM914095).

In brief, sequencing reads are trimmed with Trimmomatics (Bolger et al., 2014) by 6bp at each end, additional low quality bases from either end are removed and reads with average quality score of 30 or greater are kept. Genomic DNA reads are aligned using bowtie2 (parameters: --sensitive) to the reference genome (Release 5/dm3) and RNA sequencing reads are aligned using tophat2 (Trapnell et al., 2009) (parameters: -m 1 -p 5 -g 2 -I 50000 --microexon-search --no-coverage-search).

PCR duplicates were not removed for editing analysis. Following re-analysis of a select number of data sets there was only a small change in the number or identity of editing sites and essentially no change in the overlap with CLIP data. The gtf file used throughout the analysis was downloaded from iGenomes (via Illumina, dm3 build). The editing analysis pipeline used RefSeq annotation of genes, which were downloaded from UCSC genome browser. Only uniquely mapped reads are considered for editing analysis. A table of raw and mapped reads is included in Table S3. As number of editing events detected will change with sequencing depth, within each experiment each sample was sequenced to roughly the same depth. Initial

experiments with Hrp48 TRIBE in S2 cells are an exception, experimental samples were sequenced to ~200 million reads to determine if many events were not being detected with less coverage. 70-90 edit sites were detected in wild type cells and cells expressing the dADARcd alone, these are included in Table S4 and on GEO (these likely represent SNPs, PCR errors, endogenous editing sites and may include sites edited by the dADARcd alone).
Only events with minimum 20 reads and 10% editing were considered to be an editing events, gDNA coverage and uniformity of nucleotide identity were also required to avoid inclusion of single nucleotide polymorphisms. RNA editing data is converted to bedgraph file format for display, where the editing percentage (# edit bases/ # total bases) is the height of the displayed black bar. RNA editing data is manipulated as bed/bedgraph format using bedtools.

HITS-CLIP

The CLIP libraries of Hrp48 were constructed as described previously with slight modification (Cho et al., 2012). S2 cells or head powder of CS flies were irradiated twice for 300mJ/cm^2 at 254nm (CL-1600 Ultraviolet crosslinker). For RNaseA treatment, 40µl of 40ng/ml RNaseA was added to 1ml of lysate. For Immunoprecipitation of Hrp48, dynabeads protein A (Invitrogen) and rabbit anti-Hrp48 antibody (a gift from Donald Rio, (Hammond et al., 1997)) were used. For Immunoprecipitation of Hrp48-ADAR-V5, dynabeads protein A (Invitrogen), rabbit anti-mouse IgG antibody (Millipore, 06-371) and mouse monoclonal anti-V5 antibody (Invitrogen, 46-1157) were used. Following linkers and primers were used for making libraries. 3' linker (RA3), /5rApp/TGGAATTCTCGGGTGCCAAGG/3ddC/; 5' linker (RA5-N4), rGrUrUrCrArGrArGrUrUrCrUrArCrArGrUrCrCrGrArCrGrArUrCrNrNrNrN; RNA RT Primer (RTP), GCCTTGGCACCCGAGAATTCCA; RNA PCR Primer (RP1, forward primer), AATGATACGGCGACCACCGAGATCTACACGTTCAGAGTTCTACAGTCCGA; RNA PCR Primer-Index 7 (RPI7, reverse primer), CAAGCAGAAGACGGCATACGAGATGATCTGGTGACTGGAGTTCCTTGGCACCCGAG AATTCCA; RNA PCR Primer-Index 8 (RPI8, reverse primer), CAAGCAGAAGACGGCATACGAGATTCAAGTGTGACTGGAGTTCCTTGGCACCCGAG AATTCCA; RNA PCR Primer-Index 9 (RPI9, reverse primer),

CAAGCAGAAGACGGCATACGAGATCTGATCGTGACTGGAGTTCCTTGGCACCCGAG
AATTCCA.

CLIP, motif analysis and gene ontology

CLIP sequence reads were aligned using Novoalign, (novoalign -t 85 -l 23 -s 1 -r None). Subsequent bioinformatic analysis of CLIP data was performed as described in Moore et al. (2014) and using the CLIPper algorithm (available at https://github.com/YeoLab/clipper) with default settings. To calculate a per gene CLIP enrichment value we locally normalized the peak heights of each CLIP region to the mRNA-seq reads extracted for that region, and summed these values across each gene (using CLIPper peaks). We performed motif analysis using MEME ((Bailey et al., 2009), version 4.10.0) (parameters: meme -minw 6 -maxw 10 -maxsize 1000000 -dna -nmotifs 5 -mod zoops) around significant CLIP sites as defined by the Darnell pipeline (Moore et al., 2014). MEME analysis for TRIBE data was performed by inputting the region around the editing site. Gene ontology analyses were performed using DAVID (Huang da et al., 2009), TRIBE target genes (in FBgn format) were analyzed against a background of all genes expressed in the fly brain (>2 fpkm, ~ 8000 genes). Gene expression levels were quantified using Cufflinks2.

RNA structure analysis

RNA structure folding with UNAFOLD was carried out on flanking sequences near TRIBE editing sites and CLIP binding sites identified by CIMS analysis (Cross Linking induced Mutation Site) that lack editing by TRIBE. A flanking region of 250nt both 5' and 3' of the site (501nt in total) was folded with UNAFold parameters (hybrid-ss-min --suffix DAT --mfold=5,8,200 --noisolate), base pairing was counted in the predicted minimum free energy (MFE) and predicted suboptimal structures within $\Delta\Delta G$=5Kcal/mol of the MFE [1] for each site and five nucleotides on either side of the sites to create a profile of double strandedness. All profiles for each site are averaged to create a plot (mean +/- SEM, n = 17 TRIBE editing sites) for the TRIBE sites and CIMS CLIP sites.

Notes on analysis of RNA editing data

All analysis of TRIBE data performed here are binary; either a gene has a TRIBE editing site or not, i.e., the mRNA is a target or not. We could have employed two possible metrics to rank the 'strength' of RBP targets, the number of editing sites in a gene and the extent to which they are edited (editing percentage). However, we believe that such metrics should be interpreted with caution. First, longer genes obviously have more capacity for harboring more editing sites. Second, editing percentage may be affected by the propensity of the target region to participate in double stranded structures (see Discussion) as well as by the strength of interaction between an RBP and its target.

The number of sites caused by a particular RBP-TRIBE protein will presumably vary quite widely so it is difficult to recommend a specific sequencing depth. Factors that could affect the number of editing events include the strength of binding of the endogenous RBP, the strength of binding of the TRIBE-RBP, the structure preference of the RBP, the expression level of the TRIBE-RBP, in addition sequencing depth effects the number of sites detected. The deeper the sequencing depth the more editing sites will be detected. For our experiments the deepest we have sequenced is ~200 million mapped reads (Hrp48-TRIBE, s2 cells), however, in our experience 20-30 million mapped reads can be sufficient (Fig S5). We recommend that within an experiment each sample is sequenced to roughly the same depth.

Supplemental References

Abruzzi K, Chen X, Nagoshi E, Zadina A, Rosbash M (2015) RNA-seq profiling of small numbers of Drosophila neurons. Methods in enzymology 551:369-386.

Bailey TL, Boden M, Buske FA, Frith M, Grant CE, Clementi L, Ren J, Li WW, Noble WS (2009) MEME SUITE: tools for motif discovery and searching. Nucleic acids research 37:W202-208.

Bashirullah A, Halsell SR, Cooperstock RL, Kloc M, Karaiskakis A, Fisher WW, Fu W, Hamilton JK, Etkin LD, Lipshitz HD (1999) Joint action of two RNA degradation

pathways controls the timing of maternal transcript elimination at the midblastula transition in Drosophila melanogaster. The EMBO journal 18:2610-2620.
Blanchette M, Green RE, MacArthur S, Brooks AN, Brenner SE, Eisen MB, Rio DC (2009) Genome-wide analysis of alternative pre-mRNA splicing and RNA-binding specificities of the Drosophila hnRNP A/B family members. Molecular cell 33:438-449.
Bolger AM, Lohse M, Usadel B (2014) Trimmomatic: a flexible trimmer for Illumina sequence data. Bioinformatics (Oxford, England) 30:2114-2120.
Cho J, Chang H, Kwon SC, Kim B, Kim Y, Choe J, Ha M, Kim YK, Kim VN (2012) LIN28A is a suppressor of ER-associated translation in embryonic stem cells. Cell 151:765-777.
Hammond LE, Rudner DZ, Kanaar R, Rio DC (1997) Mutations in the hrp48 gene, which encodes a Drosophila heterogeneous nuclear ribonucleoprotein particle protein, cause lethality and developmental defects and affect P-element third-intron splicing in vivo. Molecular and cellular biology 17:7260-7267.
Huang da W, Sherman BT, Lempicki RA (2009) Systematic and integrative analysis of large gene lists using DAVID bioinformatics resources. Nature protocols 4:44-57.
Moore MJ, Zhang C, Gantman EC, Mele A, Darnell JC, Darnell RB (2014) Mapping Argonaute and conventional RNA-binding protein interactions with RNA at single-nucleotide resolution using HITS-CLIP and CIMS analysis. Nature protocols 9:263-293.
Nelson MR, Luo H, Vari HK, Cox BJ, Simmonds AJ, Krause HM, Lipshitz HD, Smibert CA (2007) A multiprotein complex that mediates translational enhancement in Drosophila. The Journal of biological chemistry 282:34031-34038.
Rodriguez J, Menet JS, Rosbash M (2012) Nascent-seq indicates widespread cotranscriptional RNA editing in Drosophila. Molecular cell 47:27-37.
Trapnell C, Pachter L, Salzberg SL (2009) TopHat: discovering splice junctions with RNA-Seq. Bioinformatics (Oxford, England) 25:1105-1111.