iScience, Volume *23*

Supplemental Information

MS2-TRIBE Evaluates Both Protein-RNA Interactions and Nuclear Organization of Transcription by RNA Editing

Jeetayu Biswas, Reazur Rahman, Varun Gupta, Michael Rosbash, and Robert H. Singer

Supplemental Figures:

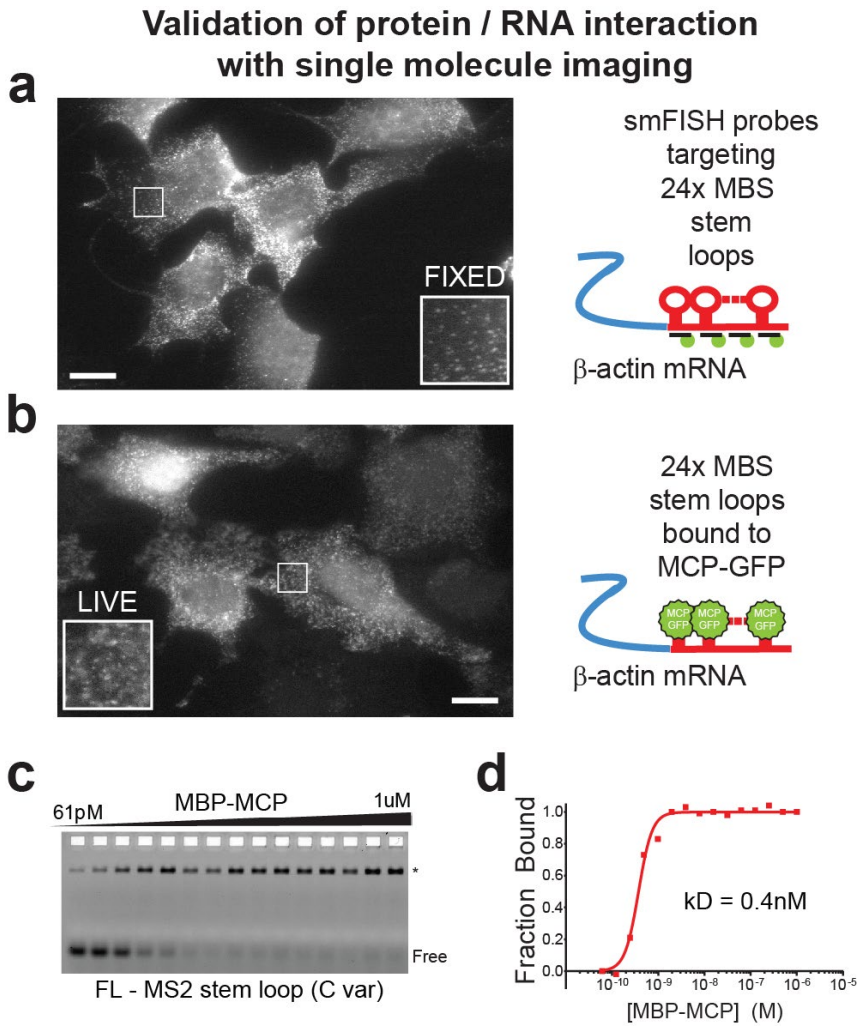


Figure S1. Single molecule imaging and biochemical characterization of the MCP MBS complex. Related to figure 1.

(a) Immortalized MEFs from the β-actin-MBS mouse were fixed and smFISH was performed using probes against the stem loop array. Single molecules of β-actin mRNA were visualized as bright, uniform intensity spots in the cytoplasm (insert), scale bar = 15μm.

(b) Live imaging of the immortalized β-actin MBS MEFs was performed by stably infecting the cells with MCP-GFP and FACS sorting the GFP positive populations, single molecules were again visualized and had uniform fluorescence intensity (insert), scale bar = 15μm.

(c) Representative electrophoretic mobility shift assay (EMSA) MS2 stem loop RNA. The filled triangle represents a 1:1 serial dilution of MCP. The RBPRNA complex (*) and free RNA (FREE) are labeled.

(d) Quantification and fit to the Hill equation of EMSA results for MCP (red solid line).

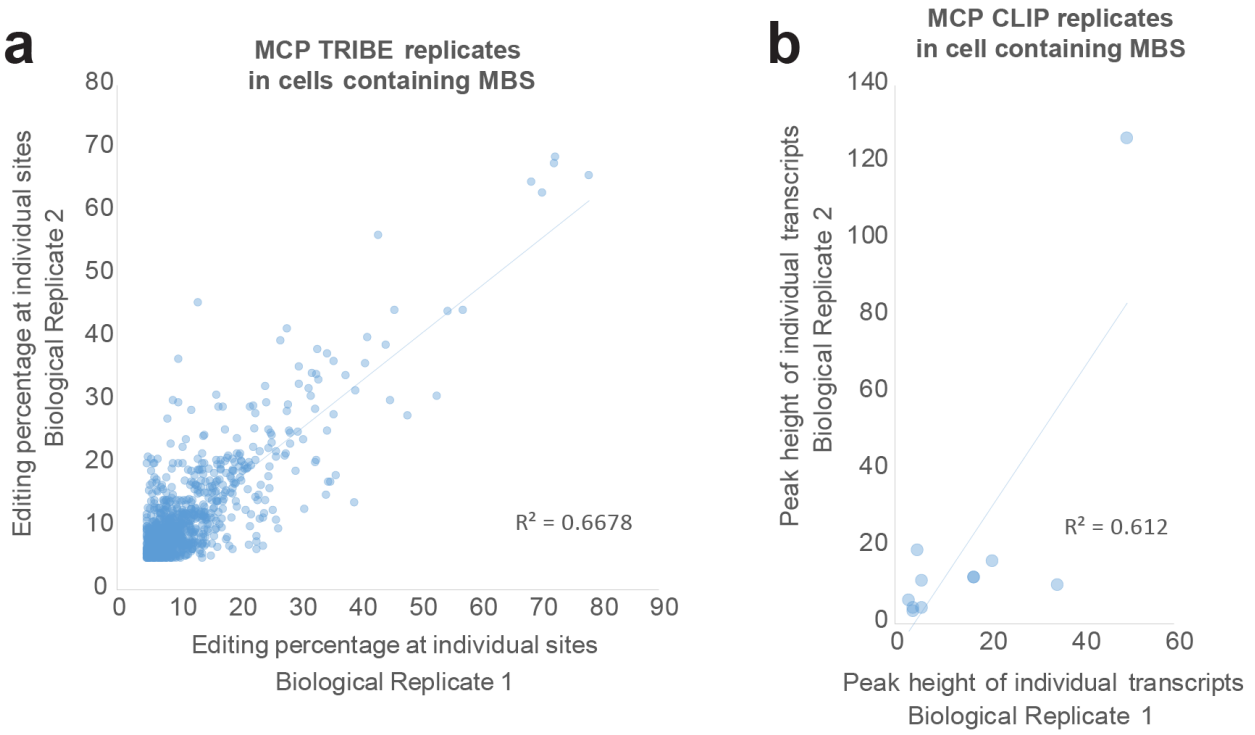
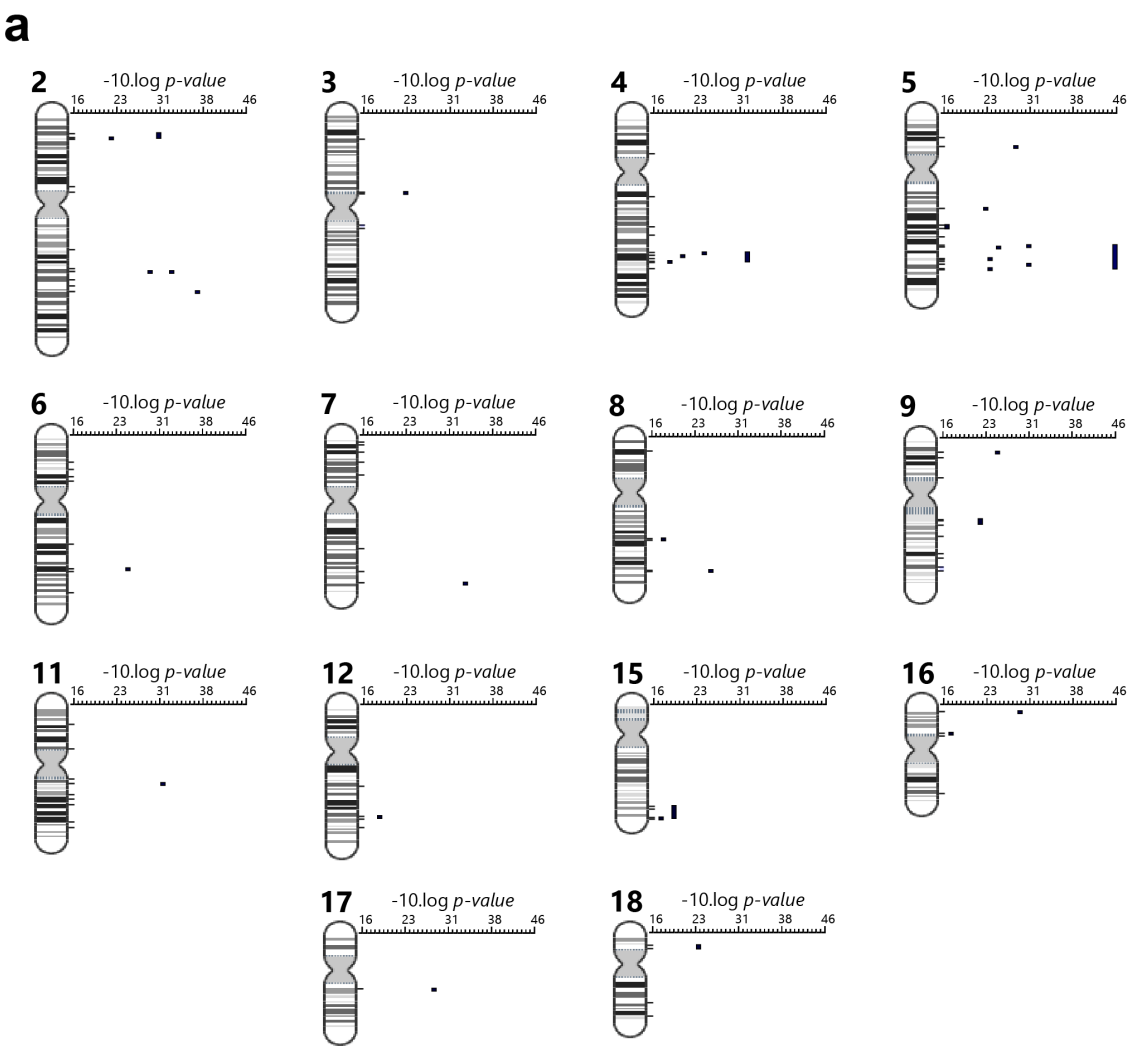
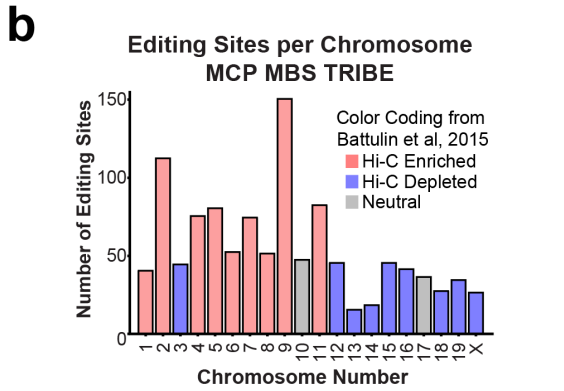


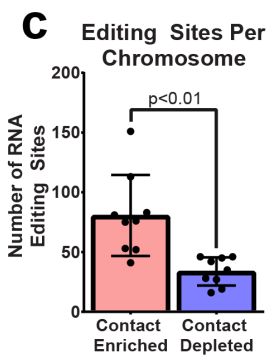
Figure S2. Correlation between MCP-TRIBE and MCP-CLIP biological replicates. Related to figure 1 and 2.

(a) Correlation of RNA editing events across MCP-TRIBE biological replicates. X axis is editing percentage of nucleotides from the replicate one, Y axis is editing percentage of nucleotides from replicate two. Individual edited nucleotides are represented as blue dots. Data fit to linear regression (blue line) and used to calculate R^2.

(b) Correlation of CLIP peak heights across MCP-CLIP biological replicates. X axis is editing peak height from the replicate one, Y axis is the peak height from replicate two. Individual transcripts are represented as blue dots. Data fit to linear regression (blue line) and used to calculate R^2.


a
2
3
4
5
6
7
8
9
11
12
15
16
17
18
-10.log p-value
16
23
31
38
46


b
Editing Sites per Chromosome
MCP MBS TRIBE
Number of Editing Sites
150
100
50
0
Color Coding from
Battulin et al, 2015
Hi-C Enriched
Hi-C Depleted
Neutral
1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18 19 X
Chromosome Number


c
Editing Sites Per
Chromosome
Number of RNA
Editing Sites
200
150
100
50
0
p<0.01
Contact
Enriched
Contact
Depleted

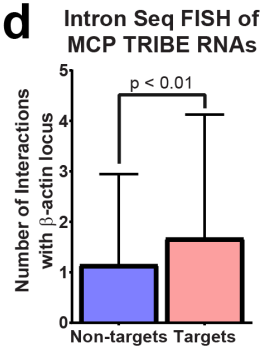

d
Intron Seq FISH of
MCP TRIBE RNAs
Number of Interactions
with β-actin locus
5
4
3
2
1
0
p < 0.01
Non-targets
Targets

Figure S3. MCP TRIBE shows significantly edited regions of DNA corresponding to chromosome contacts. Related to figure 3.

(a) Positional gene enrichment of MCP TRIBE results in β-actin-MBS cells. To left are chromosome numbers with diagrams beneath. Significantly enriched regions are displayed as black boxes from left to right, in order of increasing significance (-10 log pvalue).
(b) Bar graph showing number of editing sites organized by chromosome of origin. Chromosomes are colored by Hi-C contacts with Chromosome five (Data from (Battulin et al., 2015)). Chromosomes enriched for contacts are colored in red, chromosomes depleted in contacts are colored in blue.
(c) Data in B represented as individual points, mean of the data summarized as a bar graph with error bars representing standard deviation. Chromosomes were grouped by their relationship to Chromosome five in MEFs, either into contact enriched (red) or contact (depleted). Unpaired, two tailed t-test was used to determine significance.
(d) Bar graph showing average number of contacts (as calculated by seq-FISH interaction matrix) for MCP-TRIBE targets vs non targets with the region surrounding β-actin-MBS, error bars represent standard deviation. Unpaired, two tailed t-test was used to determine significance.

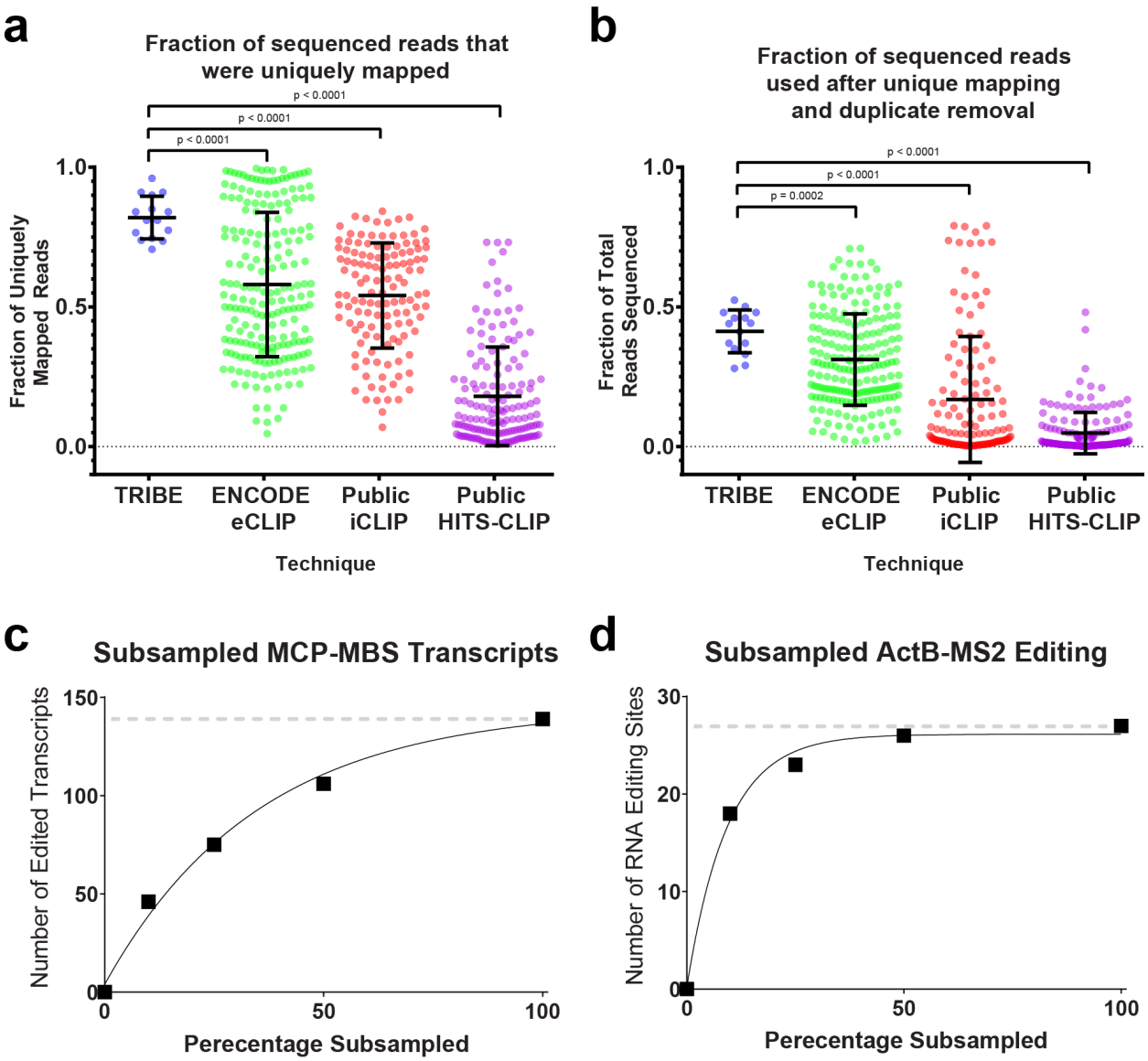


Figure S4. TRIBE libraries are more complex and more efficiently sequenced than CLIP libraries, allowing discovery of targets to be saturated. Related to figures 2, 3 and 4.

(a) Percentage of reads in FASTQ file that were uniquely mapped to the genome. Individual points represent individual experiments or biological replicates and error bars represent standard deviation of the mean. P values calculated using two-tailed Welch's t-test. Processed CLIP data from (Van Nostrand et al., 2016).

(b) Percentage of reads in FASTQ file that are retained after complete processing. Major RNA seq processing steps include unique mapping and PCR duplicate removal. Individual points represent individual experiments or biological replicates and error bars represent standard deviation of the mean. P values calculated using two-tailed Welch's t-test. Processed CLIP data from (Van Nostrand et al., 2016).

(c) Scatter plot showing number of edited transcripts as a function of random subsampling from the aligned reads. X axis, subsampled percentages (100%, 50%, 25%, 10%, 0%). Y axis, number of

RNAs in the transcriptome that were edited by MCP-TRIBE (black boxes). Solid black line represents nonlinear fit of the data (one phase association).

(d) Scatter plot showing number of RNA editing sites on β-actin-MBS transcript a function of random subsampling from the aligned reads. X axis, subsampled percentages (100%, 50%, 25%, 10%, 0%). Y axis, number of RNAs in the transcriptome that were edited by MCP-TRIBE (black boxes). Solid black line represents nonlinear fit of the data (one phase association).

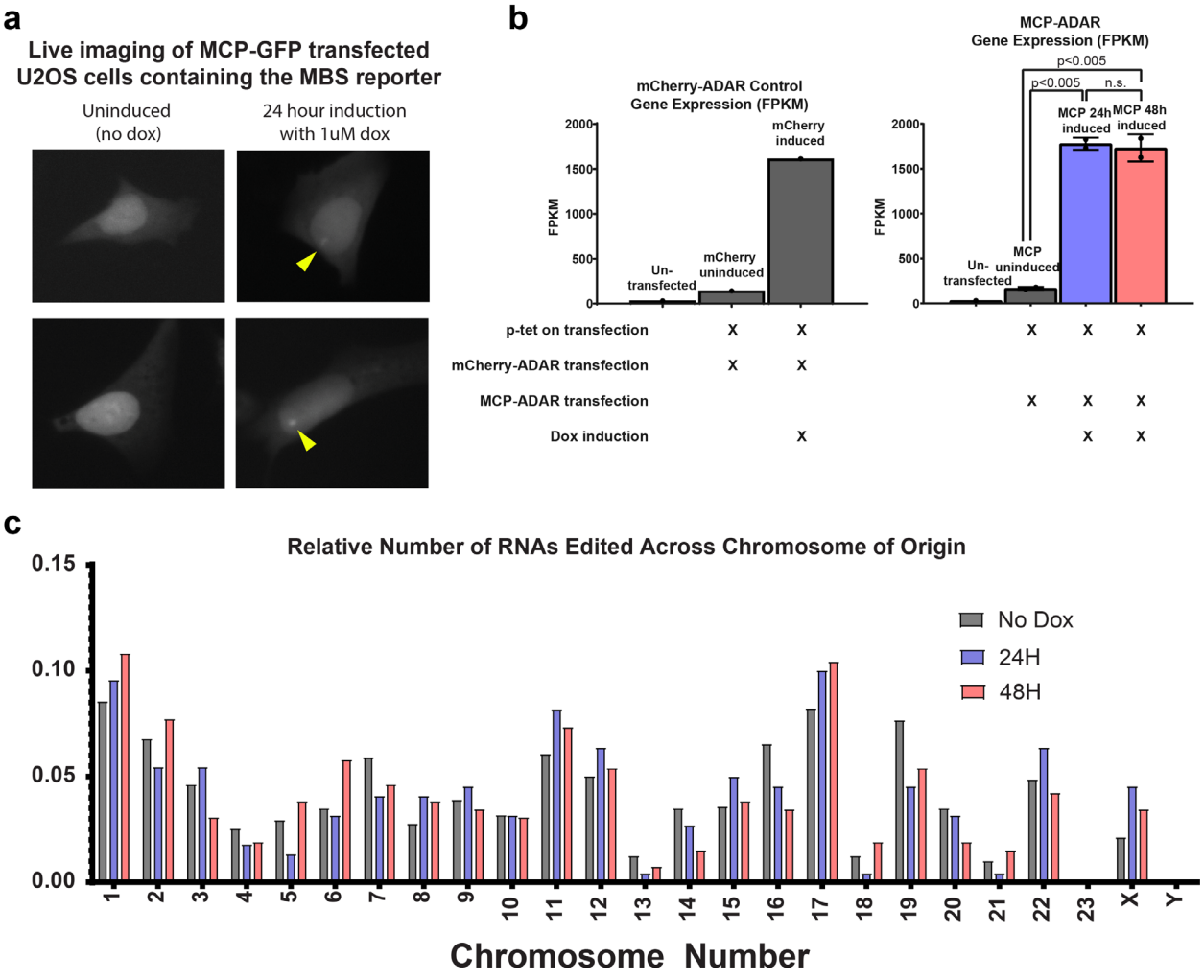


Figure S5. Gene expression changes during induction of MBS reporter RNA into U2OS cells show that inter-chromosomal contacts are preserved at a range of expression levels. Related to figure 4.

(a) Live imaging of the immortalized U2OS cells containing the stably integrated reporter MBS array. Cells were transiently transfected with ptet-on and MCP-GFP, cells were either kept in DMEM without doxycycline (left) or media supplemented with 1uM doxycycline (right). Upon induction, bright transcription sites (yellow arrows) were visualized.

(b) Gene expression was quantified and normalized to sequenced depth post PCR duplicate removal. Gene counts were then normalized as FPKM values. Different conditions for transfection and doxycycline treatment depicted as X symbols below each bar. Expression from each biological replicate was averaged, error bars representing standard deviation of two samples. Unpaired, two tailed t-test was used to determine significance.

(c) Bar graph showing relative number of genes edited (y-axis) organized by chromosome of origin (x-axis). Relative number of editing sites calculated by normalizing genes edited per chromosome to the total number of genes edited in the experiment. Chromosomes are colored by experiment and colors correspond to panel B. Gray bars representing cells with no doxycycline induction (low expression), blue bars represent 24 hours of doxycycline induction and red bars represent 48 hours of doxycycline induction.

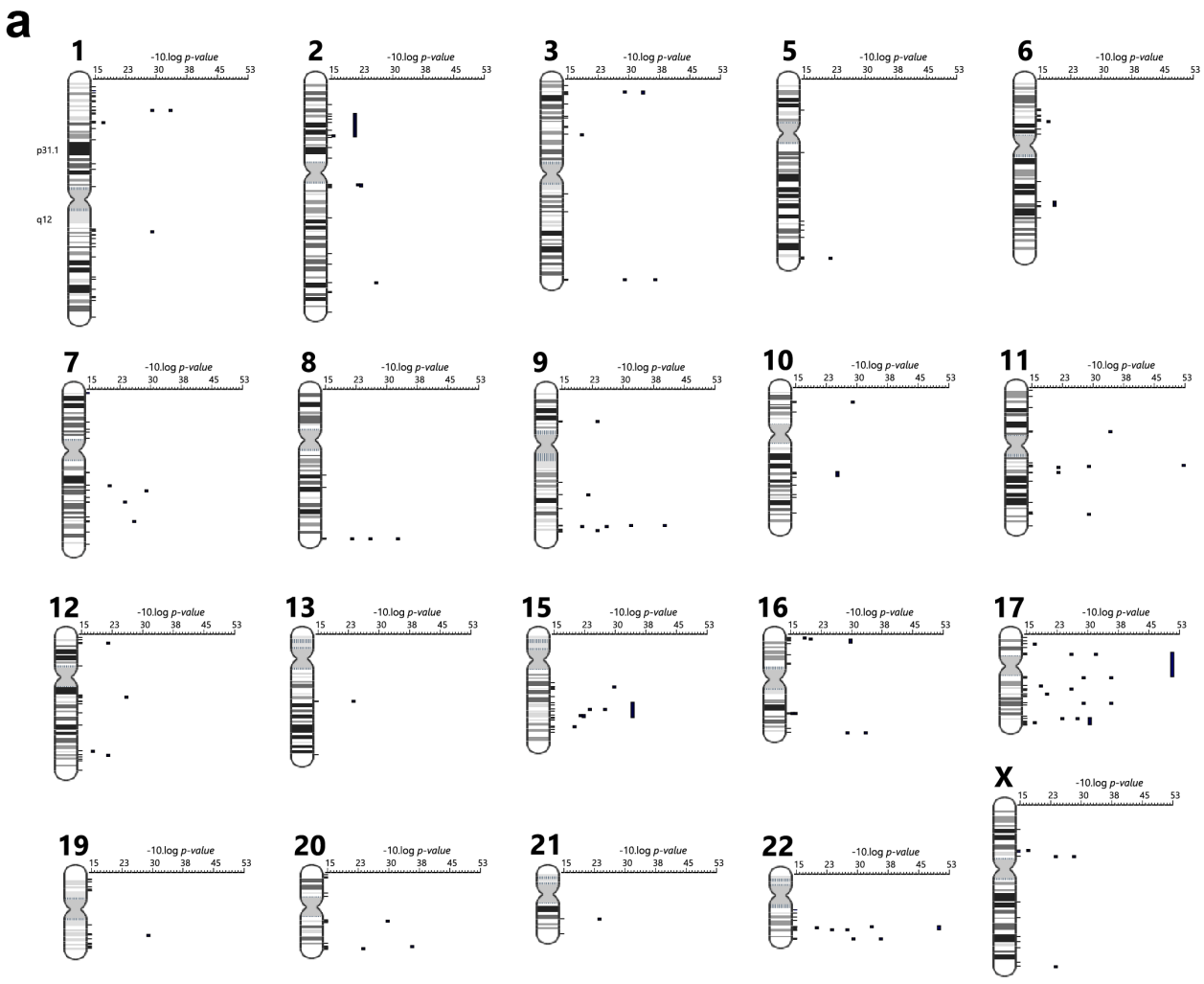


Figure S6. Positional gene enrichment shows high inter chromosomal contacts with chromosome 17. Related to figure 4.

(a) Positional gene enrichment of MCP TRIBE results in reporter MBS U2OS cells. To left are chromosome numbers with diagrams beneath. Significantly enriched regions are displayed as black boxes from left to right, in order of increasing significance (-10 log pvalue).

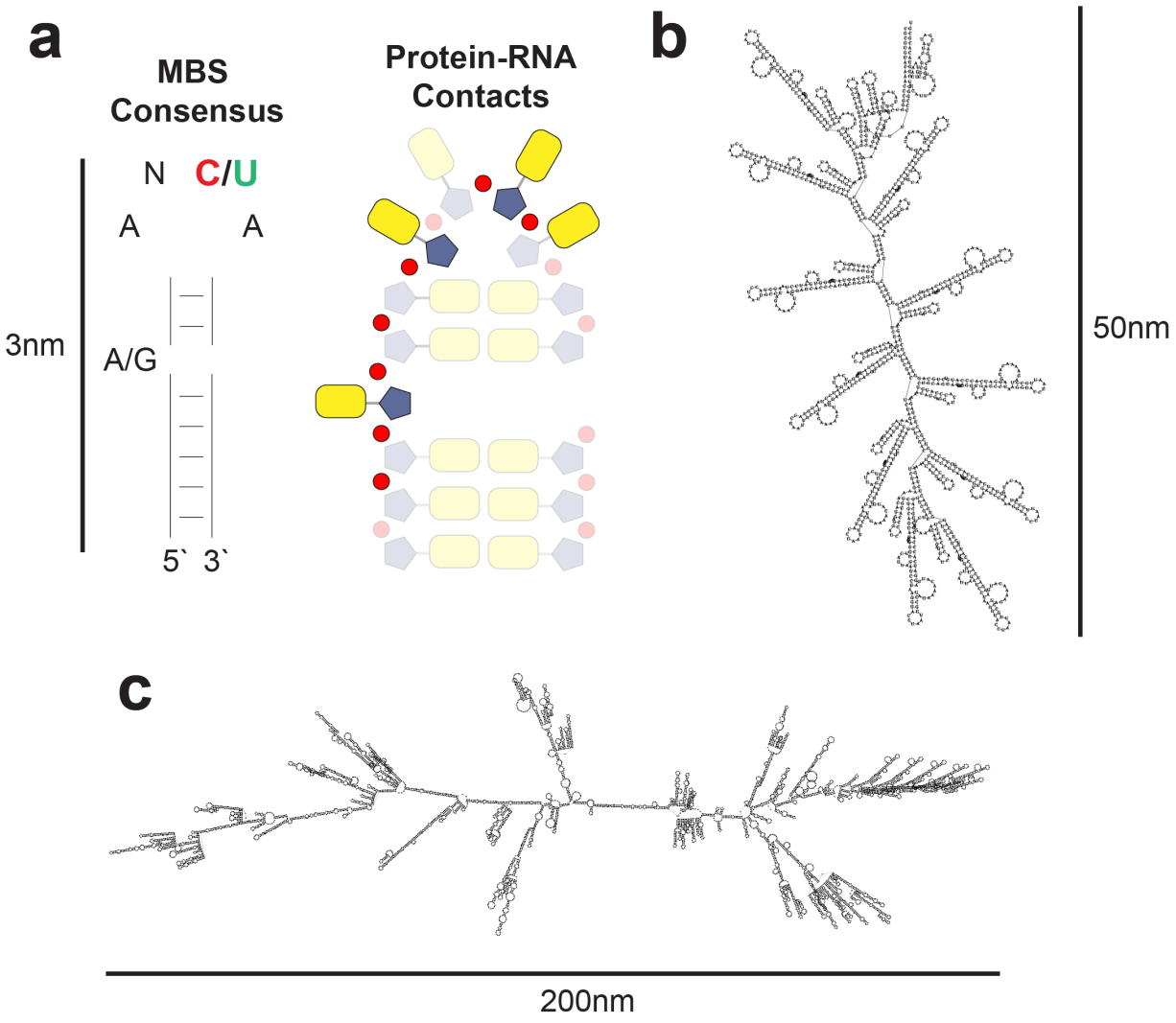


Figure S7. Estimation of RNA sizes for MS2 size, MBS and β-actin-MBS. Related to figures 3 and 4.

(a) Distance from base of MS2 stem to top of loop is approximately 3nm. The nucleotide consensus sequence for MCP recognition on left (Tutucci et al., 2018). Protein RNA contacts depicted as darker shaded shapes on right where yellow boxes represent nucleobases, red circles represent phosphate and blue pentagons represent ribose sugars. The distance and contacts were based on the co-crystal structure of MCP bound to the MS2 stem loops, PDB: 2BU1.

(b) Folded and energy minimized RNA secondary structure of the MBS (24x MS2 stem loops). Predictions were made using Vienna RNA (Hofacker, 2003).

(c) Folded and energy minimized RNA secondary structure of β-actin-MBS mRNA. Predictions were made using Vienna RNA (Hofacker, 2003).

Transparent Methods:

Modifications from previously published TRIBE protocol

When adapting the technique from Drosophila (McMahon et al., 2016; Rahman et al., 2018) a number of changes were made. The first being that the two plasmid system required for TRIBE (one plasmid carrying the RBP fused to ADAR, the other carrying GFP as a marker) was combined onto a single multicistronic vector by using the p2A system. The equimolar expression for the upstream RBP-ADAR and downstream GFP cistrons, allowed for accurate quantification RBP expression levels during FACS sorting (Lo et al., 2015). These same clonal lines were also used as background controls for CLIP studies, highlighting antibody specific contributions to downstream CLIP analysis. To further increase the signal to noise of RNA editing, two dADAR point mutations were used. The hyperediting mutant of dADAR (E488Q) had previously been shown to be 10x more active (Kuttan and Bass, 2012) and had less sequence bias (Xu et al., 2018). To further increase signal, S458 was synonymized to stop auto inactivation of RNA editing (Palladino et al., 2000).

Plasmid construction

tdMCP-stdGFP (Addgene # 98916) was modified to generated the vectors for this study. The first GFP was excised and replaced with dADAR E488Q with synonymized serine 458 codon (synthesized by IDT). A p2A sequence was added to the end of dADAR by PCR. Once this vector was made, MCP was replaced by mCherry and served as a control.

Cell culture and transfections

Primary MEFs were isolated from E14 β-actin-MBS embryos containing 24x MS2 stem loops in the endogenous β-actin 3' UTR (Katz et al., 2012). Primary MEFs were then immortalized by transient transfection with a plasmid expressing SV40 large T antigen (Addgene #21826). Single cell clones were then isolated by limiting dilution. MEFs were continuously cultured in DMEM (4.5g/L, Corning) supplemented with penn/strep (Gibco) and 10% FBS (Atlanta Biologicals). Transient transfection with TRIBE plasmids was performed using JetPrime (Polyplus) as per manufacturer instructions 12 hours before fluorescence activated cell sorting (FACS).

Previously published human U2OS cells (Janicki et al., 2004) containing a stably integrated reporter RNA were cultured continuously in DMEM (4.5g/L, Corning) supplemented with penn/strep (Gibco) and 10% FBS (Atlanta Biologicals). Two days prior to experiments cells were transiently transfected with designated plasmids (mCherry-ADAR, MCP-ADAR, pTet-on) using JetPrime (Polyplus) as per manufacturer instructions. At noted timepoints (24 hours or 48 hours prior to FACS sorting), media was supplemented with 1uM doxycycline to induce expression of the integrated reporter construct.

Fluorescence Activated Cell Sorting (FACS)

Cells were trypsinized and suspended in sorting buffer (DPBS, 1% BSA supplemented with DAPI) and passed through a single cell mesh filter. FACS was performed by selecting for GFP positive DAPI negative single cells on a BD Aria II instrument. At least 1,000 GFP positive cells were directly FACS sorted into 800uL Trizol (Invitrogen) and RNA isolation was performed as per manufacturer's instructions.

RNA isolation and library prep from FACS sorted cells

GFP+ DAPI- cells were FACS sorted directly into Trizol and RNA was extracted, quantified with qubit HS RNA assay and integrity was visualized with a RNA pico chip (Agilent Bioanalyzer). Samples with no apparent degradation were used to prepare paired end, stranded, libraries (NEB Ultra II Directional RNA Library Prep Kit with poly A RNA isolation module). At least 10ng of RNA input per sample (as quantified

by qubit RNA HS assay). After polyA selection and adapter saserreeddligation, final libraries were amplified with 1-2 rounds of PCR less than the manufacturers recommendation and checked on the bioanalyzer for appropriate size distribution. Library concentrations were then quantified by qubit dsDNA HS assay, qPCR and Agilent bioanalyzer (the latter two of which were performed by Novogene and used for multiplex library normalization prior to sequencing).

Computational search for MS2 sequence motifs

Find motif, an operation built into IGV was utilized to discover MCP binding motifs acrosss mouse (mm10 + β-actin-MBS contig) genomes using the following regex pattern: [AG][ACGT][ACGT]A[ACGT][CT]A. Motif searches were performed one chromosome at a time and then individual BED concatenated into a single file with all genome wide motif coordinates. BED files containing a genome wide map of the motifs found were intersected with the corresponding GTF file (mm10) to associate the motifs with gene symbols. Motifs that intersected with exons in the GTF file were then used to define putative targets of each RNA binding protein. Each consensus sequence containing RNA was associated with its FPKM value from the cell type of interest. Expressed transcripts were defined as those with an FPKM >1. Transcripts that fell below this value were considered non expressed and excluded from further analysis.

Sequencing and analysis of TRIBE data

Samples were then combined and multiplexed across multiple lanes for HiSeq 4000 runs (performed by Novogene). Paired end sequencing was performed with 150bp reads at a depth of at least 27.5Gb per sample (this corresponded to at least 40 million sequenced read pairs per sample). The data was analyzed following the publicly available computational pipeline [https://github.com/rosbashlab/HyperTRIBE/] developed here (Rahman et al., 2018) with a few modifications described in this section. The github repository has been updated to incorporate the modifications. Reads from sequencing libraries are trimmed and aligned to the transcriptome, which was modified to include either an additional MBS sequence at the mm10 β-actin locus or at the hg38 reporter gene locus. Each modified locus was added as a separate contig. Subsequently, the nucleotide frequency at each position in the transcriptome is recorded from aligned reads. For each nucleotide in the transcriptome, the TRIBE RNA nucleotide frequency is compared with the wild type mRNA library (wtRNA) nucleotide frequency to identify RNA editing sites. For unambiguous identification of RNA editing sites above background editing, the frequency of A is greater than 80% and the frequency of G is 0.5% in the wtRNA and the frequency of G is greater than 0 in the TRIBE RNA (using the reverse complement if annotated gene is in the reverse strand). There are two modifications compared to the previous version of the TRIBE pipeline. Modifying the nucleotide frequency of G in wtRNA from 0 to 0.5% allows very low level sequence heterogeneity in wtRNA and does not disrupt the identification of legitimate editing sites in deeply sequenced libraries. We used an editing threshold of at least 5% in each replicate, as lowering the threshold from 10% did not significantly alter the percentage of TRIBE target genes with consensus sequence overlap. The TRIBE edit sites are required to be present in both replicates and any sites that overlap with catalytic Adar alone edit sites with at least 1% editing are removed. Finally, a list of transcripts is created by tabulating all transcript that contain edit sites.

CLIP protocol (From (Cho et al., 2012)) with the following adaptations

15cm dishes of MEFs (at ~70% confluency) were washed with ice cold DPBS, after aspiration and addition of 12mL ice cold DPBS, cells were irradiated with 300mJ/cm of 254nm UV light (BioRad GS Gene Linker). After irridation, each 15cm dish was scraped and collected into a 15mL falcon tube. Cells were pelleted at 500g for 5 minutes and then the pellets were stored at -80 until use.

For Immunoprecipitations, a 1:1 mix of monoclonal Anti-GFP antibodies (Sigma Aldrich / Roche 11814460001) were used with protein G dynabeads (4ug of antibodies with 50uL of beads, Thermo Fisher Scientific). All washes were performed using a magnetic stand.

Sequencing and analysis of CLIP data:

CLIP libraries were sequenced with 75bp SE reads on NextSeq500 using the high output mode. Both replicates were concatenated before peak calling. Both traditional peak calling with Clip Tool Kit (CTK) (Shah et al., 2017) [https://zhanglab.c2b2.columbia.edu/index.php/Standard/BrdU-CLIP_data_analysis_using_CTK] as well as CIMS analysis (Moore et al., 2014) were performed.

Interchromosomal Hi-C Analysis

The comparison to MEF Hi-C data to determine interchromosomal enrichment / depletion done using two different data sets. Chromosome five contacts with all other chromosomes were considered and data was binned into three groups (enriched, depleted or neutral).

Both data sets were mapped using the iterative correction outlined (Imakaev et al., 2012) and gave consistent results. The first data set utilized previously published data (Battulin et al., 2015) and was analyzed by hiclib software. A separate MEF Hi-C experiment was analyzed using the Juicer suite of tools. Interchromosomal contacts were isolated from a .hic file containing high quality mappings (>30).

The expected number of interactions between each chromosome pair (chromosomes i and j for example) was calculated (Lieberman-Aiden et al., 2009). Reads originating from between two different chromosomes (interchromosomal) were considered.

The fraction of interchromosomal reads coming from chromosome i * fraction of interchromosomal reads coming from chromosome j * total number of interchromosomal reads

The expected number of interactions represents chances of a read randomly occurring between chromosome i AND chromosome j based on the total number of interchromosmal reads generated.

Enrichment was then calculated by taking the number of interactions observed between chromosomes i and j and dividing by the expected value. Chromosomes enriched above the average for chromosome five were considered enriched and those below average were considered depleted.

Intron seqFISH data analysis

Previously published intron seq-FISH data was utilized to determine how often transcription sites contact with one another (Shah et al., 2018). In brief, the pairwise distances between each transcription site was calculated. The pairwise contact frequency was determined by counting the number of times two transcription sites were within 500nm of each other. This contact frequency was used to generate a pairwise interaction matrix. To analyze the matrix, the 100 transcripts adjacent to the β-actin locus were isolated as one set of the pairwise interaction. TRIBE targets were then isolated as the other end of the of pairwise interactions, the non TRIBE targets were used as a control group. The average contact frequencies between the targets and non targets were compared.

Cellular Seeding and Fixation

Immortalized MEFs containing MS2-tagged β-actin mRNA were seeded on a 18mm #1 cover glass in a 12-well cell culture plate and grown for at least 24 h in DMEM (4.5g/L, Corning) supplemented with penn/strep (Gibco) and 10% FBS (Atlanta Biologicals). Cells were washed three times with PBS-M (1x PBS supplemented with 5mM $MgCl_2$), fixed for 10 min with 4% paraformaldehyde in PBS-M at room temperature, washed three times in PBS-M and then quenched in 50 mM glycine in PBS-M. Permeabilization was then performed with 0.1% Triton X-100 (Thermo Fisher Scientific) in 1× PBS-M for 10 min at room temperature.

Fixed cell imaging - smFISH

β-actin-MBS probes were used to detect MBS cassette present in β-actin mRNA 3′-UTR. The probes used (MS2_LK20 - tttctagagtcgacctgcag, MS2_LK51-1 - ctaggcaattaggtaccttag, MS2_LK51-2 - ctaatgaacccgggaatactg) were 50mers of ssDNA bearing each 4–5 fluorophores (as previously described (Lionnet et al., 2011)). After cellular seeding and fixation (above), coverslips were washed with PBS and then incubated in 30% (vol/vol) formamide, 2x SSC in RNase-free water for 1 hour at room temperature, followed by incubation for 3 hours at 37 °C with 10-ng of probes in hybridization buffer containing 30% formamide, 1 mg/mL E. coli tRNA, 10% dextran sulfate, 20 mg/mL BSA, 2x SSC, 2 mM vanadyl ribonucleoside complex, and 10 U/mL SUPERase-In (Ambion) in RNase-free water. After incubation, cells were washed twice with in 30% formamide in 2x SSC RNase-free water for 10 min at 37 °C. Following three quick washes in 2x SSC, cells were mounted using ProLong Diamond Antifade Reagent with DAPI (Life Technologies). Images were taken using an upright, wide-field Olympus BX-63 Microscope equipped with a SuperApochromatic 60x/1.35 N.A. Olympus Objective (UPLSAPO60XO), an X-Cite 120 PC Lamp (EXFO), an ORCA-R2 Digital Interline CCD Camera (C10600-10B; Hamamatsu) and zero-pixel shift filter sets: DAPI-5060C-Zero, FITC-5050A-Zero, Cy3-4040C-Zero, and Cy5-4040C-Zero from Semrock. Metamorph software (Molecular Devices) was used for controlling microscope automation and image acquisition.

Live cell imaging

Immortalized MEFs containing MS2-tagged β-actin mRNA and stably expressing MCP-GFP were seeded on a 18mm #1 cover glass in a 12-well cell culture plate and grown for at least 24h in DMEM (4.5g/L, Corning) supplemented with penn/strep (Gibco) and 10% FBS (Atlanta Biologicals). After 24 hours the cells were placed into pre-warmed live cell imaging media (Lebovitz 15, Life Technologies supplemented with 10% FBS, Atlanta Biologics). Images were taken using an inverted, wide-field Olympus IX-81 Microscope equipped with a 60×/1.4 N.A. Olympus Objective, an X-Cite 120 PC Lamp (EXFO), a Sensicam QE cooled CCD camera, and fluorescence filter sets: DAPI, CFP, FITC, Cy3, Cy5 and YFP. IP Lab software (BD Biosciences) was used for controlling microscope and image acquisition.

Electrophoretic Mobility Shift Assays (EMSAs)

Electrophoretic mobility shift assay. Single MS2 stem–loop RNAs with 5′ fluorescein modification (Dharmacon) were deprotected at 60C for 30 minutes using 100 mM acetic acid adjusted to a pH of 3.4–3.8 using TEMED. Prior to the experiment, RNA stocks were heated to 70 °C for 5 min then snap cooled on ice. 100 pM RNAs were incubated at room temperature for 3 h with two-fold dilutions of MCP in 10 mM Tris, 100 mM NaCl, 0.1 mM EDTA, 0.01 mg/mL tRNA, 50 μg/mL heparin and 0.01% IGEPAL CA630. Complexes were then run using 5% native PAGE in 0.5× TBE and visualized using the Typhoon 9400 variable mode laser scanner (GE Healthcare).

Supplemental References:

Cho, J., Chang, H., Kwon, S.C., Kim, B., Kim, Y., Choe, J., Ha, M., Kim, Y.K., and Kim, V.N. (2012). LIN28A Is a Suppressor of ER-Associated Translation in Embryonic Stem Cells. Cell *151*, 765–777.

Hofacker, I.L. (2003). Vienna RNA secondary structure server. Nucleic Acids Res. *31*, 3429–3431.

Imakaev, M., Fudenberg, G., McCord, R.P., Naumova, N., Goloborodko, A., Lajoie, B.R., Dekker, J., and Mirny, L.A. (2012). Iterative correction of Hi-C data reveals hallmarks of chromosome organization. Nat. Methods *9*, 999–1003.

Katz, Z.B., Wells, A.L., Park, H.Y., Wu, B., Shenoy, S.M., and Singer, R.H. (2012). β-Actin mRNA compartmentalization enhances focal adhesion stability and directs cell migration. Genes Dev. *26*, 1885–1890.

Lo, C.-A., Kays, I., Emran, F., Lin, T.-J., Cvetkovska, V., and Chen, B.E. (2015). Quantification of Protein Levels in Single Living Cells. Cell Rep. *13*, 2634–2644.

Moore, M.J., Zhang, C., Gantman, E.C., Mele, A., Darnell, J.C., and Darnell, R.B. (2014). Mapping Argonaute and conventional RNA-binding protein interactions with RNA at single-nucleotide resolution using HITS-CLIP and CIMS analysis. Nat. Protoc. *9*, 263–293.

Palladino, M.J., Keegan, L.P., O'Connell, M.A., and Reenan, R.A. (2000). dADAR, a Drosophila double-stranded RNA-specific adenosine deaminase is highly developmentally regulated and is itself a target for RNA editing. RNA N. Y. N *6*, 1004–1018.

Shah, A., Qian, Y., Weyn-Vanhentenryck, S.M., and Zhang, C. (2017). CLIP Tool Kit (CTK): a flexible and robust pipeline to analyze CLIP sequencing data. Bioinformatics *33*, 566–567.

Van Nostrand, E.L., Pratt, G.A., Shishkin, A.A., Gelboin-Burkhart, C., Fang, M.Y., Sundararaman, B., Blue, S.M., Nguyen, T.B., Surka, C., Elkins, K., et al. (2016). Robust transcriptome-wide discovery of RNA-binding protein binding sites with enhanced CLIP (eCLIP). Nat. Methods *13*, 508–514.