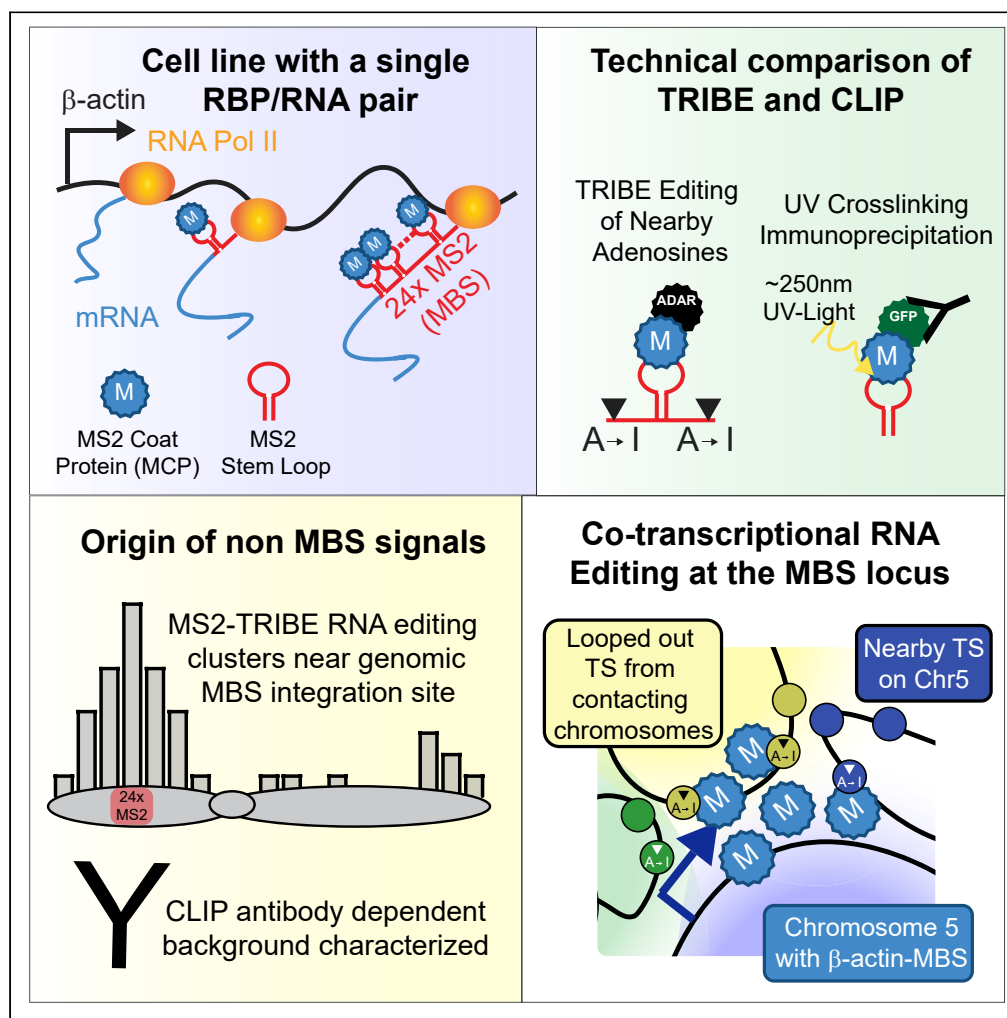# iScience

Article

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

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

robert.singer@einsteinmed.org

HIGHLIGHTS

Using the β-actin-MBS system, we identify the false-positives of TRIBE and CLIP

Mammalian TRIBE has fewer putative false-positives than CLIP

The non-MS2 targets for TRIBE are neighboring nascent transcripts

We identify a nuclear domain where nascent RNAs and physically interact

Article

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

Jeetayu Biswas,[1,5] Reazur Rahman,[2,5] Varun Gupta,[3] Michael Rosbash,[2] and Robert H. Singer[1,4,6,*]

SUMMARY

Both UV-cross-linking and immunoprecipitation (CLIP) and RNA editing (TRIBE) can identify the targets of RNA-binding proteins. To evaluate false-positives of CLIP and TRIBE, endogenous β-actin mRNA was tagged with MS2 stem loops, making it the only bona fide target mRNA for the MS2 capsid protein (MCP). CLIP and TRIBE detected β-actin, albeit with false-positives. False-positive CLIP signals were attributed to nonspecific antibody interactions. In contrast, putative false-positive TRIBE targets were genes spatially proximal to the β-actin gene. MCP-ADAR edited nearby nascent transcripts consistent with interchromosomal contacts observed in Hi-C. The identification of nascent contacts implies RNA regulatory proteins (e.g., splicing factors) associated with multiple nascent transcripts, forming domains of post-transcriptional activity. Repeating these results with an integrated inducible MS2 reporter indicated that MS2-TRIBE can be applied to a broad array of cells and transcripts to study spatial organization and nuclear RNA regulation.

INTRODUCTION

RNAs are regulated throughout all aspects of their lives (Vera et al., 2016). Their regulation by RNA-binding proteins (RBPs) provides spatiotemporal control over processing, export, localization, translation, and decay. RBP genes make up approximately 10% of all protein coding genes in the human genome (Gerstberger et al., 2014), and RBP families have coevolved with their RNA targets to provide control over functionally related transcripts (Hogan et al., 2015). Many RBPs contain more than one RNA-binding domain, with each domain contributing to the final specificity of the protein (Gerstberger et al., 2014). Accurately defining the targets of a given RBP has been a longstanding challenge in biology due to their complex and multivalent interactions.

In vitro approaches have determined the sequence specificity of single domains within RBPs. Approaches such as SELEX and "RNA bind-n-seq" (Lambert et al., 2014) have determined the consensus sequence preference for RBPs using short RNA fragments (<40 nucleotide [nt]). Complementary approaches have found consensus binding sites *in vivo,* using cross-linking and immunoprecipitation (CLIP).

Cell lysis before immunoprecipitation promotes adventitious interactions such as those between cytoplasmic RBPs and nuclear RNAs (Mili and Steitz, 2004). To overcome these limitations, CLIP sequencing and its derivatives use UV to irreversibly cross-link protein and RNA within cells (Ule et al., 2003). These interactions can be defined with single nucleotide resolution (Zhang and Darnell, 2011). However, CLIP relies on antibodies and suffers from high background due to nonspecific antibody-antigen interactions (Friedersdorf and Keene, 2014). Therefore, an orthogonal approach is important to obtain high-confidence targets for RBPs of interest. It should be highly specific, antibody independent, retain information about the total RNA present for a given species, and be limited in the amount of sample and its handling steps.

Two approaches have recently been developed, RNA editing in *Drosophila* ("TRIBE"; McMahon et al., 2016; Xu et al., 2018) and RNA tagging in yeast (Lapointe et al., 2015). Both approaches utilize enzymes fused to an RBP of interest to deposit marks on their RNA targets. We chose to focus on RNA editing because of its straightforward approach utilizing standard RNA sequencing (RNA-seq) library preparation.

Conspicuously missing from previous transcriptome-wide studies of RBPs is a positive control, which could resolve rigorously the extent of false-positives: the best-case scenario would be an RBP that recognizes

only a single transcript within the cell. Thus, any targets identified outside of the single transcript would be defined as false-positives. The specificity of the MCP-MBS system provides such a tool: its capsid protein (MCP) should only bind the MS2 stem loops (MBS) inserted into a gene of interest. We used cells derived from a mouse that has the MBS inserted into the 3′ UTR of the endogenous β-actin gene and used the MCP as a means to evaluate various approaches to identifying targets of RBPs (Lionnet et al., 2011).

To validate this approach, both CLIP and TRIBE were performed on cells containing the MCP-MBS system. The single β-actin mRNA labeled with the MCP-MBS system allowed for unambiguous determination of background or false-positives for each technique. Nonspecific antibody interactions contributed significantly to MCP-CLIP signals, whereas there were fewer non-β-actin target mRNAs with MCP-TRIBE (Tables S1 and S2). Moreover, these reproducible, non-random targets correlated with their proximity to the β-actin locus on the chromosome rather than manifesting some association with β-actin mRNA. This result led to the unexpected conclusion that MCP-ADAR had edited nearby transcribing RNAs. The tethered editing enzyme can therefore define a nuclear transcriptional domain with implications for how the transcriptome may be organized in the nucleus.

RESULTS

MCP-ADAR Edits Targets Consistent with Binding the β-Actin-MBS

We evaluated both CLIP and TRIBE with an RBP that should recognize a single target within the transcriptome. We used mouse fibroblasts containing 24xMBS integrated into the endogenous β-actin locus (Figure 1A, Lionnet et al., 2011). mRNA was imaged with single-molecule fluorescence *in situ* hybridization (smFISH) (Figure S1A), and live cells were imaged with MCP-GFP (Figure S1B). Both approaches demonstrated a sufficient signal-to-noise ratio for single molecule detection. The sub-nanomolar K_d of MCP for the β-actin-MBS *in vitro* (Figures S1C and S1D) was consistent with the high specificity of MCP for its stem-loop target *in vivo*.

To determine if the tight binding of MCP to the β-actin-MBS array would lead to RNA editing of the β-actin transcript, MCP fused to ADAR was introduced into cells containing the β-actin-MBS mRNA. We also performed CLIP on cells stably expressing MCP-GFP and β-actin-MBS. In addition, the MCP consensus binding sequence was utilized to perform an *in silico* search for MS2 binding sites across the transcriptome (Tutucci et al., 2018). All three approaches were successful in predicting MCP binding to the target sites (Figures 1B, S2A, and S2B).

Reproducible MCP-CLIP and MCP-TRIBE Define MCP Binding to β-Actin-MBS

In both approaches, the β-actin-MBS array was the top ranked site (by total peak height in CLIP or by total number of editing sites in TRIBE, [Tables S1 and S2]). For example, it had a TRIBE signal 4× greater than the next most highly edited transcript in the transcriptome (Table S1). This robust binding was expected because the β-actin-MBS array has the highest density of MCP binding motifs of any transcript in the transcriptome.

There were also 139 other transcripts identified with TRIBE and 105 with CLIP. One possibility is that they contained stem-loop structure resembling MS2 and were due to direct binding of MCP. To address this possibility, transcripts were examined for the presence of the MCP consensus sequence and the RNAs detected by MCP-TRIBE were compared with those cross-linked by MCP-CLIP.

Only two of the 139 MCP-TRIBE transcripts showed an MCP consensus sequence, and only three transcripts were shared between MCP-CLIP and MCP-TRIBE in addition to β-actin (Figure 2A, Tables S1, S2, S3). The three transcripts shared between MCP-CLIP and TRIBE showed no resemblance to either the sequence or structure of MBS (Table S3), suggesting that they were coincidental. Therefore, β-actin is the only transcript identified using all approaches (TRIBE, CLIP and consensus sequence).

Off-Target Characterization for CLIP and TRIBE

To characterize these targets further, we performed the CLIP and TRIBE without MCP or without the stem loops. CLIP of cells without MCP-GFP generated nearly 200 significant peaks, reproducing 40% of the peaks present in cells with MCP-GFP (Figure 2B). These peaks most likely reflect nonspecific antibody interactions. In support of this interpretation, there was a 5-fold increase in transcript targets when polyclonal antibodies rather than monoclonal antibodies were used (Biswas et al., unpublished data).

In contrast, MCP-TRIBE in cells without the β-actin-MBS reproduced only 13% of the MBS peaks (Figure 2C), and there were no transcripts with editing levels comparable to the 27 β-actin-MBS sites, e.g., more than 4
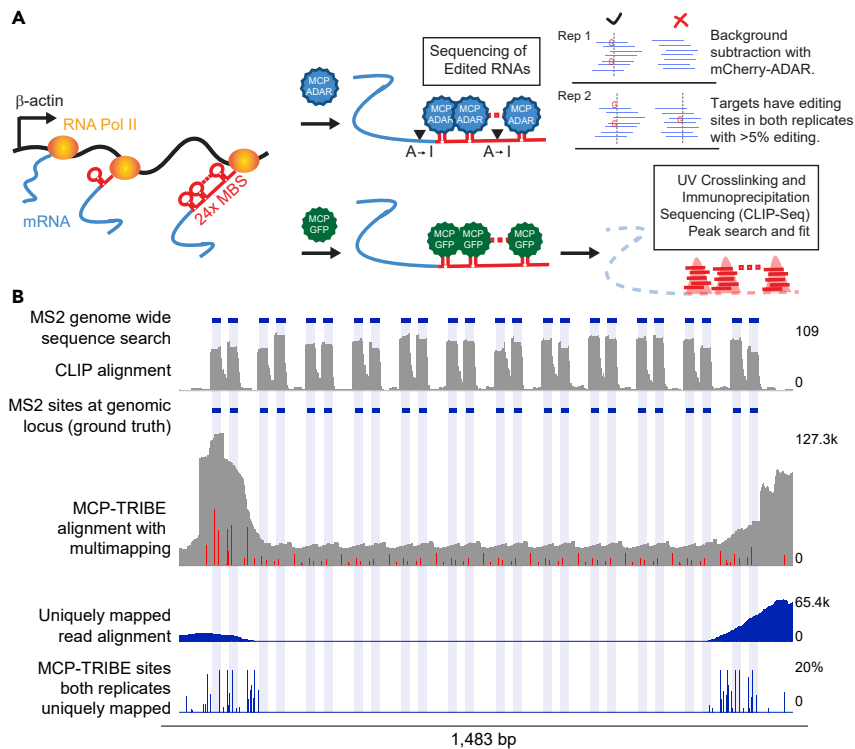
Figure 1. Transcriptome-wide Studies Show CLIP Peaks and TRIBE Editing at β-Actin-MBS

(A) Schematic of MS2 stem loops that have been integrated into the endogenous β-actin locus being actively transcribed by RNA polymerase II. (Right arrows, top) Schematic of MCP-ADAR binding to MS2 stem loops and editing adjacent nucleotides from adenosine to inosine (which is then recognized as guanosine upon cDNA conversion). After background subtraction, editing sites that are conserved across both replicates are used for further analysis. (Right arrows, bottom) Schematic of MCP-GFP binding to MS2 stem loops. UV cross-linking and immunoprecipitation (CLIP) isolates RNA fragments (red horizontal bars), which are then computationally analyzed and fit to peaks (red peaks).

(B) β-Actin gene, focusing on the MBS array showing (from top to bottom) the following. MS2 genome-wide sequence search: bioinformatic search with blue horizontal bars depicting discovered sites containing the MS2 consensus sequences. CLIP alignment: CLIP sequencing reads that remain after mapping to the genome are depicted in gray (scale bar to right). MS2 sites at genomic locus (ground truth): blue horizontal bars represent the known location of the MS2 stem loops. MCP-TRIBE alignment with multimapping: reads are depicted in gray (scale bar for number of reads on right); editing sites are depicted as red bars. Uniquely mapped read alignment: mRNA coverage without multimapping depicted in blue (scale bar for number of reads on right). MCP-TRIBE sites' both replicates uniquely mapped: editing events as indicated by dark blue bars where height corresponds to the average editing percentage across both replicates at that nucleotide (scale to right). Light blue shading indicates location of the stem-loop nucleotides.

editing sites (Figure 2D). Interestingly TRIBE showed far fewer off-target peaks compared with CLIP; only 28 transcripts in the entire transcriptome had at least two editing sites.

As the MCP β-actin-MBS TRIBE target transcripts were highly reproducible, did not have an MCP consensus sequence, were not found in MCP-CLIP, and were dependent on the presence of the MBS array (they did not appear in MCP-ADAR-expressing cells without MS2 stem loops), we considered explanations for these particular off-target transcripts.

Origin of MCP-TRIBE Off-Target Signals Can Be Correlated with the Genetic Locus

One possibility for these reproducible transcripts detected by TRIBE was that they could be spatially proximal to the β-actin-MBS-bound ADAR. For instance, if the mRNAs were packaged with β-actin mRNAs in a granule, the mRNAs might come close enough together to facilitate *trans*-editing. If this were the case, we would expect the transcripts to share a common feature, such as a consensus sequence similar to the β-actin mRNA zipcode (Patel et al., 2012). However, interrogation of the targets did not find any common consensus sequence. In addition, gene ontology analysis failed to find transcripts that code for proteins
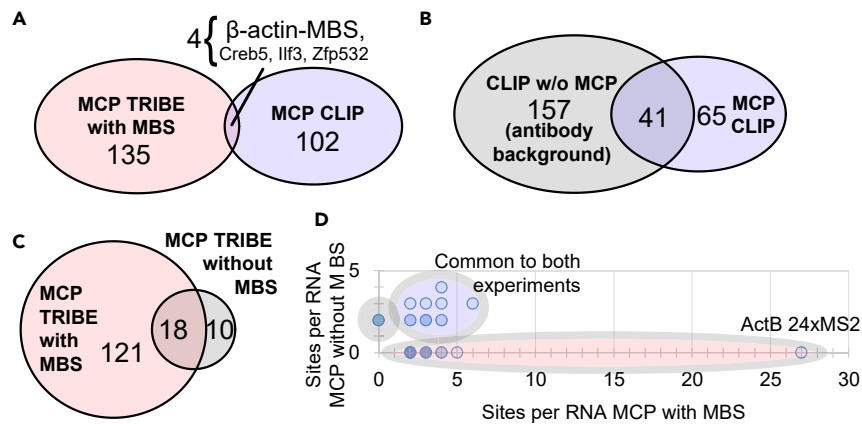
Figure 2. Comparison of MCP-CLIP and MCP-TRIBE Targets

(A) Venn diagram showing significant peaks discovered in MCP-CLIP (blue oval) and MCP-TRIBE (red oval) recognize different target pools. Four RNAs overlap between both techniques (intersection with line to names of RNAs), with only one RNA containing evidence of the MS2 consensus sequence (β-actin-MBS).

(B) Venn diagram showing intersection of RNAs found by MCP-CLIP (blue oval) and CLIP without MCP (background CLIP, gray oval). 41/106 significant CLIP peaks are reproduced in the cells that do not express MCP, representing antibody background.

(C) Venn diagram showing intersection of RNAs found with MCP-MBS-TRIBE (red circle) and MCP without MBS TRIBE (background TRIBE, blue circle).

(D) Dot plot of data from (C). Individual genes from MCP-TRIBE in cells with β-actin-MBS plotted on x axis (red circle), MCP-TRIBE in cells without β-actin-MBS plotted on y axis (blue circle). The RNAs found in both experiments are plotted on the diagonal (purple circle).

related to β-actin's role in the cytoplasm (e.g., cell motility). We then considered the possibility that the targets we identified could be close enough together to undergo *trans*-editing in the nucleus, perhaps in some nuclear domain. To this end, the off-target transcripts edited by MCP-TRIBE were interrogated to determine if β-actin-MBS mRNA could be within spatial proximity within the nucleus (Figure 3A).

Surprisingly, many of these transcripts that were edited were on chromosome five, the same chromosome as the β-actin-MBS. Chromosome five genes were edited significantly more often than expected in MCP-TRIBE data (Figure 3B), but not in TRIBE without MBS or MCP-CLIP. This suggested that most of the RNAs edited by β-actin-MBS TRIBE were originating not due to direct MCP binding to those RNAs but rather due to indirect editing by ADAR.

Spatial interactions would be expected to affect regions proximal to the β-actin locus more than distal parts of chromosome five. Indeed, higher numbers of chromosome editing sites were nearer to the β-actin-MBS array (Figure 3C). After addressing *cis*-contacts on chromosome five, transcripts from other chromosomes were queried to determine if *trans*-interactions with chromosome five could explain some of the other identified targets. Previously published Hi-C inter-chromosomal contacts from MEFs found chromosome five to have enriched contacts with chromosomes, one, two, four, five through nine, and eleven and depleted for contacts with chromosomes three, twelve through sixteene, eighteen, nineteen, and X (Battulin et al., 2015). The most enriched chromosome contacts in this dataset were two, seven, and eleven with chromosome three being notable as an example of a large chromosome with depleted contacts. Consistent with these Hi-C data, edited RNAs originated more frequently from chromosomes predicted to have more chromosome five contacts (Figures 3D, 3F, and S3A–S3C) and less frequently from chromosome three. The latter also indicated that chromosome size was not a major confounding factor.

We postulated that if editing of transcripts was occurring on nascent pre-mRNAs, including those transcribed from other chromosomes, they should all be enriched in intronic sequences. Indeed, chromosome five transcripts proximal to the MS2 insertion had more intronic editing sites as did transcripts on contacted chromosomes (Figures 3E and 3G). To validate our observations independently, we used published data provided by the intron seq-FISH approach (Shah et al., 2018). The genes that were identified as being in proximity to the β-actin-MBS locus also had their transcription sites in spatial proximity. Of the RNAs
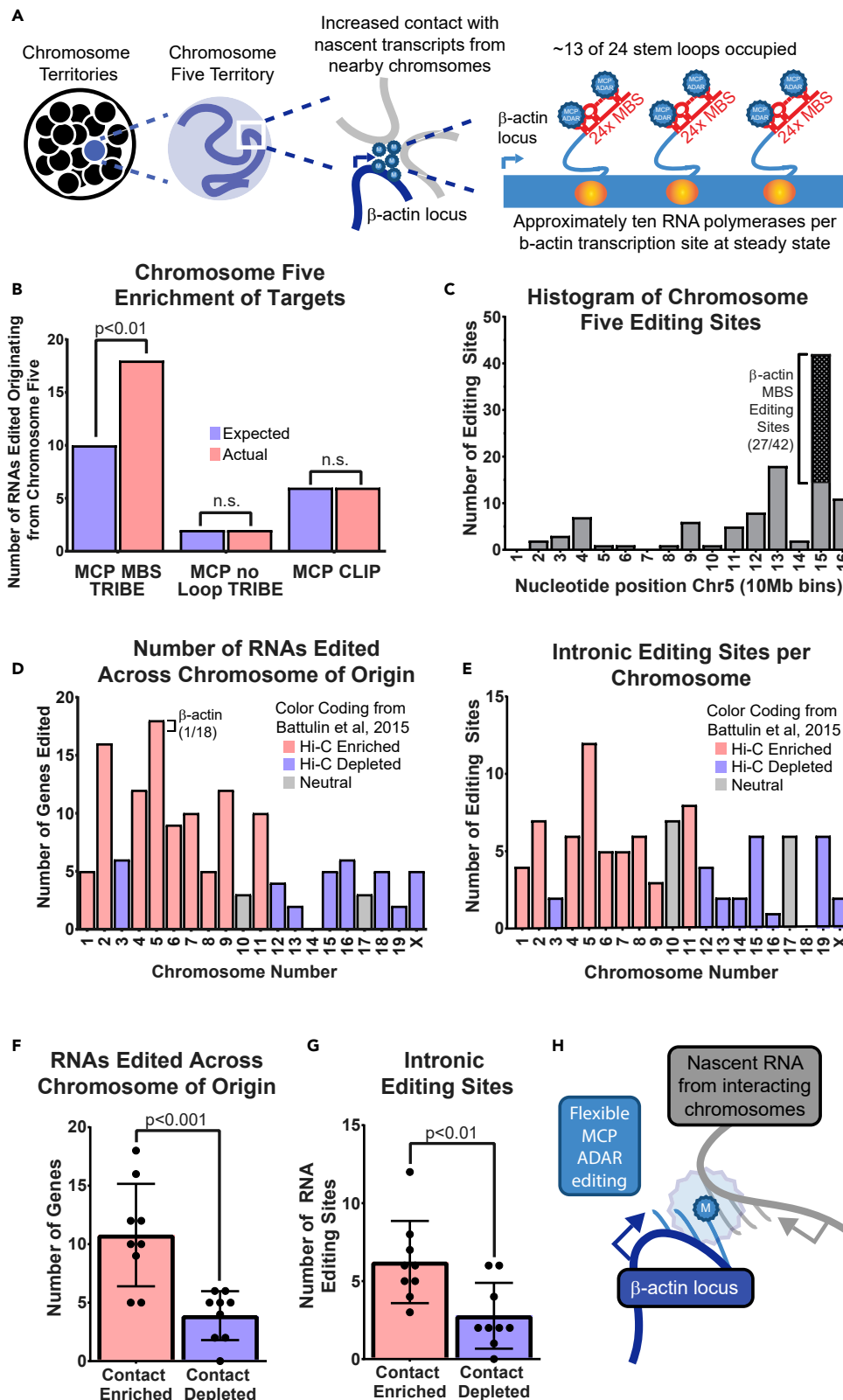
A

Chromosome Territories

Chromosome Five Territory

Increased contact with nascent transcripts from nearby chromosomes

β-actin locus

~13 of 24 stem loops occupied

β-actin locus

Approximately ten RNA polymerases per b-actin transcription site at steady state

B Chromosome Five Enrichment of Targets

Number of RNAs Edited Originating from Chromosome Five

p<0.01

Expected
Actual

n.s. n.s.

MCP MBS TRIBE
MCP no Loop TRIBE
MCP CLIP

C Histogram of Chromosome Five Editing Sites

Number of Editing Sites

β-actin MBS Editing Sites (27/42)

Nucleotide position Chr5 (10Mb bins)

D Number of RNAs Edited Across Chromosome of Origin

Number of Genes Edited

β-actin (1/18)

Color Coding from Battulin et al, 2015
Hi-C Enriched
Hi-C Depleted
Neutral

Chromosome Number

E Intronic Editing Sites per Chromosome

Number of Editing Sites

Color Coding from Battulin et al, 2015
Hi-C Enriched
Hi-C Depleted
Neutral

Chromosome Number

F RNAs Edited Across Chromosome of Origin

Number of Genes

p<0.001

Contact Enriched
Contact Depleted

G Intronic Editing Sites

Number of RNA Editing Sites

p<0.01

Contact Enriched
Contact Depleted

H

Nascent RNA from interacting chromosomes

Flexible MCP ADAR editing

M

β-actin locus

Figure 3. Increased Presence of Editing Sites on Chromosome Five and Increased Density of Editing around the β-Actin Locus

(A) Schematic of cell nucleus (large circle) containing individual chromosomes within their territories (smaller black circles with chromosome five highlighted in blue). Enlarged view of chromosome five territory showing a snapshot of DNA (blue line), as well as region of flexible movement (blue circle). Enlarged view of β-actin gene locus, showing transcription of β-actin mRNA and loading of MCP-ADAR near the site of transcription and chromatin contacts. Schematic showing increased loading of MCP-ADAR onto actively transcribing RNA.

(B) Bar graph showing enrichment (red) versus expected (blue) number of sites for chromosome five across experiments. Two-tailed chi-square test was used to determine significance for each group.

(C) Bar graph showing number of editing sites (y axis) along chromosome five (x axis, 10-Mb bins). Editing at the bin containing the β-actin-MBS is represented as a hashed box.

(D) Bar graph showing number of RNAs edited 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.

(E) Number of intronic 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.

(F) Data in (D) 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 (blue). Unpaired, two-tailed t test was used to determine significance.

(G) Data in (E) 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 (blue). Unpaired, two-tailed t test was used to determine significance.

(H) Schematic of proposed model for chromosome contacts. Transcription sites (TS, angled arrows) often loop away from chromosome of origin and intermingle with other active loci. The β-actin-MBS (blue star) recruits MCP-ADAR at the time of transcription and marks nearby active loci.

that were edited by MCP-TRIBE, 85 of 86 transcription sites interacted with chromosome five as determined by the reported seq-FISH results (within 500 nm). Strikingly, 56 of the 100 genes found to be transcribing RNA by intron seq-FISH closest to the β-actin locus on chromosome five were MCP-TRIBE targets (Figure S3D). This substantial correlation indicated that edited RNAs in the MCP-TRIBE experiment resulted from intra- and inter-chromosomal contacts with the transcripts from the β-actin-MBS locus.

Hence, editing could be a proximity labeling approach that can define a nuclear domain wherein transcripts and their processing factors spatially interact. We analyzed Hi-C data and determined that chromosomes likely to be near the β-actin-MBS locus also correlated with the chromosomes that we found to contain edited nascent transcripts (Figures 3H and S3A). Notably, as technical advantage, in contrast to chromosome conformation capture, where high sequencing depth is required to achieve megabase resolution, RNA editing is able to detect the likelihood of inter-chromosomal interactions with single gene resolution (Figures S3A, S5A–S5C, and S6).

Extending the Utility of MS2-TRIBE Using a Randomly Integrated, Inducible MBS Cassette

To explore the generality of this approach, we asked if similar results on chromatin contacts could be obtained within other cell lines containing an integrated MBS cassette. By using human cancer cells and a randomly integrated reporter cassette rather than an endogenous targeted gene, possible biases that were present due to the cell type or integration were mitigated. We previously published the use of a human osteosarcoma cell line (U2OS-2-6-3) containing an inducible MBS array integrated into chromosome one (specifically near the1p36 locus, Figure 4A) (Janicki et al., 2004).

When transfected with the TRIBE constructs, RNA containing the MBS was edited in MCP-ADAR samples and not in the mCherry controls (Figure 4B). We found that genes proximal to the MBS transcription site on chromosome one were more highly edited, consistent with *cis*-chromosomal transcripts being identified by MS2-TRIBE (Figures 4C and 4D). When plotted along the length of chromosome one, the majority of edited transcripts occurred near the known 1p36 integration locus, validating that the signal originated from transcription of the MBS. The signal appeared to broaden along the chromosome at 48 h after induction, relative to 24 h, possibly reflecting increased RNA-RNA interactions over time.

When looking at editing sites across all chromosomes, the second highest abundance of RNA editing occurred from genes on chromosome 17 (Figure 4E). Two possibilities exist for the chromosome 17 edits. There could be a translocation between chromosome 1 and 17 leading to the presence of the MS2 array
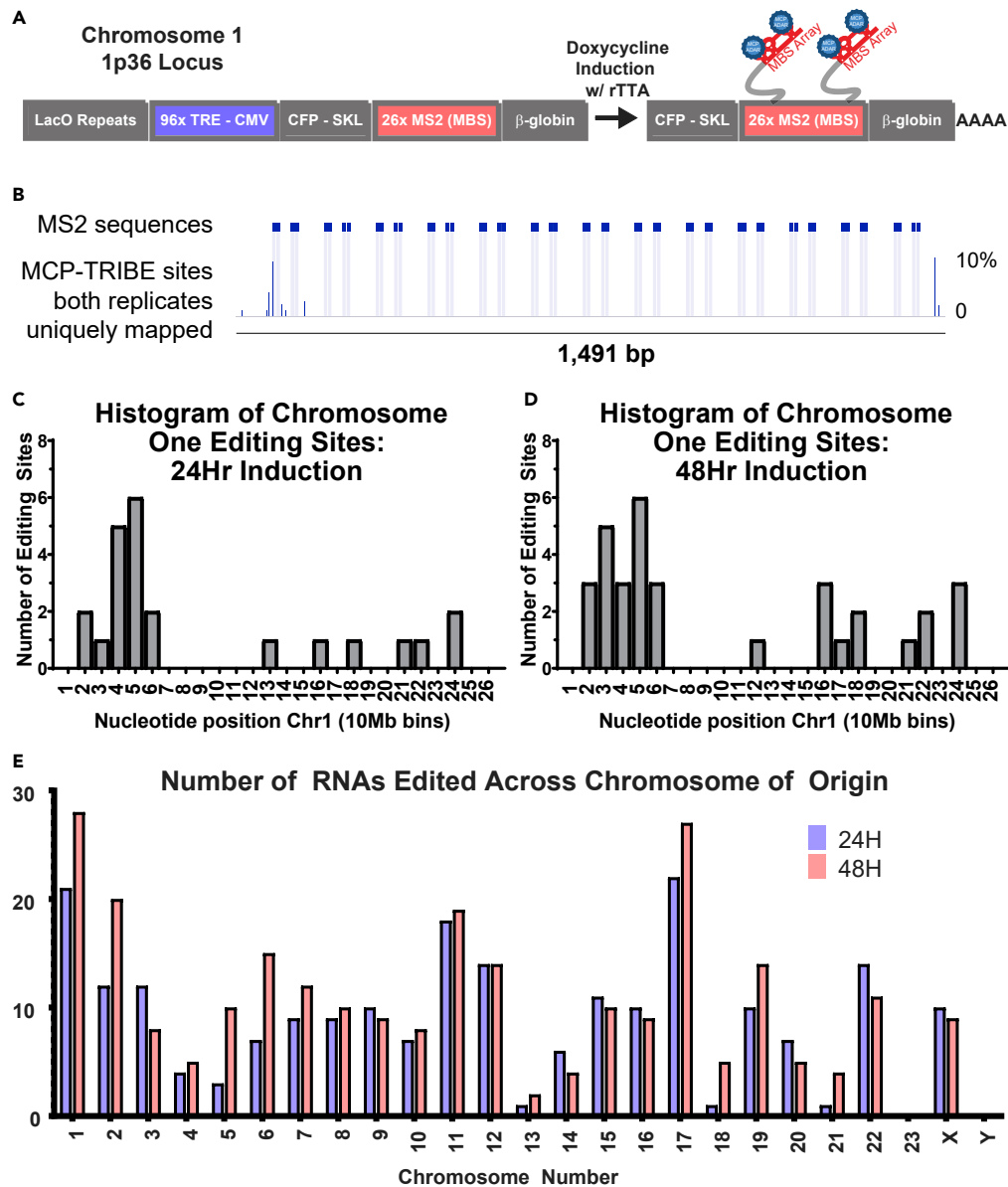
Figure 4. Increased Presence of Editing Sites on Chromosome One and Increased Density of Editing around the MBS Reporter Locus

(A) Schematic of the reporter locus on chromosome 1p36. Lac operator repeats (LacO) are present upstream of 96 repeats of the tet response element (TRE); when bound by reverse tet transactivator, the cytomegalovirus (CMV) promoter is inactive. Upon doxycycline induction the RNA containing the MBS array becomes actively transcribed (right arrow) at which point, MCP-ADAR (blue circles) can bind to the RNA at the site of transcription.

(B) Reporter MBS gene, focusing on the MBS array showing (from top to bottom) the following. MS2 sites at genomic locus (ground truth): blue horizontal bars represent the known location of the MS2 stem loops. MCP-TRIBE sites: editing events as indicated by dark blue bars where height corresponds to the average editing percentage across both replicates at that nucleotide (scale to right). Light blue shading indicates location of the stem-loop nucleotides.

(C) Bar graph showing number of editing sites (y axis) along chromosome one (x axis, 10-Mb bins) at 24 h post-doxycycline induction.

(D) Bar graph showing number of editing sites (y axis) along chromosome one (x axis, 10-Mb bins) at 48 h post-doxycycline induction.

Figure 4. *Continued*

(E) Bar graph showing 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 (B). Gray bars representing cells with no doxycycline induction (low expression), blue bars represent 24 h of doxycycline induction, and red bars represent 48 h of doxycycline induction.

physically close to the chromosome 17 transcripts. Alternatively, this could reflect transient chromosome interactions that are preferred. Further experiments with karyotyping the cell line may clarify this.

To determine if MS2-TRIBE was dependent on high expression of the stem-loop arrays we used several conditions to induce their expression. Cells transfected with reverse tetraycline transactivator (rtTA) but without doxycycline repression showed 5–6× higher levels of RNA expression than untransfected cells, similar to previously observed leaky expression of tet-inducible systems (Raj et al., 2006). Cells could then be strongly induced 10× further using 1 μM doxycycline (Figures S5A and S5B). Together, these experiments allowed us to test the dynamic range of MS2-TRIBE and determine that it provided similar results for both modestly and highly expressed transcripts.

TRIBE editing results from continuous exposure of ADAR to the target of interest. Therefore TRIBE modification of transcripts is different from methods that rely on fixation to capture a snapshot of mRNA associations. To determine whether MS2-TRIBE signal changed over time, mRNA expression was induced with doxycycline for either 24 or 48 h before RNA-seq library preparation. We found no significant difference in overall editing between 24 and 48 h of induction, suggesting that both conditions had reached steady state (Figures S5A and S5B) and that differential expression of the reporter mRNA was not be a confounding factor.

We investigated inter-chromosomal contacts over time. There was a modest 12% increase in the total number of genes edited at the 48 h time point; chromosomes one and seventeen increased the number of genes edited by 33% and 27% respectively. This suggested that integration of chromosome contacts over time provided higher editing signal of TRIBE (Figures 4E and S5C, Table S4).

DISCUSSION

By adapting TRIBE to mammalian cells and converting it to a single plasmid system, a number of advantages were evident. Mammalian TRIBE is an antibody independent approach that can be performed on 1000× fewer cells (samples were generated from fewer than 2,000 cells) than CLIP (2–20 million cells required) or chromosome conformation capture-based approaches (requiring >10 million cells). The sequencing depth required for TRIBE is similar to either CLIP or chromatin conformation capture variants, but by relying on a minimum number of processing steps for RNA-seq library preparation, TRIBE leads to more efficient sequencing (Figures S4A and S4B). By retaining the sequence complexity of the cellular transcriptome, fewer PCR cycles are required for library amplification, thus leading to increased sequencing fidelity. Significantly more sequenced TRIBE reads are usable than the most efficient CLIP approaches (Figures S4A and S4B).

During the process of gel excision, CLIP tags are limited in size due to the additional mass of the protein-RNA complex, 100 nt of RNA add approximately or 32 kDa mass to the protein-RNA smear. Post processing, most CLIP tags are ~50 nt; in contrast, TRIBE can utilize longer RNA-seq reads (150 nt here), which increases its ability to uniquely map reads (Figure S4A). This allows TRIBE to discover the entire repertoire of RNA targets, as evidenced by saturated target discovery and RNA editing (Figures S4B and S4C). In addition, the molecular biology required for TRIBE is radiation free and limited to cell sorting, RNA isolation, and standard RNA library preparation, processes that are routine in many laboratories and can be completed in as few as 3 days.

Definition of Background

Until recently, limited alternative approaches to CLIP existed. One prior approach was pertinent to the study of splicing factors. In this context, CLIP was assisted by correlating binding sites with sites of alternative splicing, as determined by RNA-seq. Ultimately, multiple transcriptome-wide approaches should be integrated to evaluate the targets of RBPs (Herzog et al., 2020; Lapointe et al., 2018; Rahman et al., 2018).

Prior work comparing CLIP with RNA tagging in yeast found ~50% overlap between the two approaches (Lapointe et al., 2015, 2018). Similar percentages were found when comparing CLIP and TRIBE in *Drosophila* (McMahon et al., 2016). However, most comparisons have been performed on RBPs that bind to a significant percentage of cell transcripts (often >50%), thus confounding the value of percentage overlap. The large number of bound targets also makes it challenging to identify the origin of these differences. By using the MCP-MBS system we were able to perform CLIP and TRIBE with only one true target within the cell. The small number of discovered targets in each technique (<200 of 15,000 transcripts) avoided spurious overlap and readily identified overlap as well as false-positives.

Prior work showed that UV cross-linking occurred at the site of RBP binding to RNA. We investigated the origin of off-target CLIP signal using multiple methods. We were unable to find the same RNAs when assayed by TRIBE (Figure 2A), and off-target RNAs discovered by CLIP did not contain matches to the well-defined MS2 consensus sequence (Table S3). Finally, for CLIP, the use of control cells not expressing MCP reveals that nonspecific RNA binding to antibodies contributes to CLIP signal. A similar issue has been described in chromatin immunoprecipitation sequencing data from both *Drosophila* and yeast, leading to the presence of "phantom peaks" that are present even when the antigen is not (Jain et al., 2015; Teytelman et al., 2013).

There are a number of advantages to CLIP such as the nucleotide-level identification of RBP binding sites and its application to endogenous targets and tissues. CLIP has clear merits that support its widespread adaptation and use by the scientific community. However, it is necessary to consider antibody-dependent background issues that can dominate the CLIP data of RBPs. These findings provide a comparison of relative strengths and limitations of the two techniques using a definitive positive control experiment. Defining a single positive control, and the insights gained from this, provides a standard test for any methodology for characterizing their signal and noise components.

Considerations and Limitations when Using MS2-TRIBE

The small number of MCP-TRIBE targets and higher signal to noise allowed us to discover the cause of off-target transcripts. When multiple copies of MCP-ADAR were tethered to a highly transcribed 24xMS2 locus, editing also affected nearby transcripts, both on its own chromosome and on nearby chromosomes. This is likely assisted by frequent transcription of both the β-actin gene and our reporter gene (Kalo et al., 2015). The high levels of transcription may promote assembly of "transcription hubs" and further promote inter-chromosomal interactions. As 13 molecules of MCP are on average loaded onto each molecule of mRNA containing the MBS reporter (Wu et al., 2012) this likely enhances the amount of RNA editing of nearby transcripts.

The 24xMS2 stem loops in the MBS create high local concentration of ADAR loaded onto the nascent chains. We do not expect that this will affect the use of TRIBE to define binding sites when fused to an RBP. This is based on two observations from our MS2-TRIBE data, the first being that (as detailed in Figure 4D) the direct interactions of ADAR with the stem loop containing RNA causes far more editing sites (27 *cis*-editing sites for β-actin-MBS) than those in *trans* (the majority having one or two sites). Consistent with this finding, we also observe that when performing TRIBE with other well-known RBPs such as ZBP1, its targets have a high number of RNA editing sites, far more than could be garnered from *trans* interactions (Biswas et al., 2020, Biswas et al., in preparation). As many sequence-specific RBPs bind with a 1:1 stoichiometry (Wu et al., 2015) we anticipate that a single ADAR enzyme will not be able to act with high efficiency upon nearby targets when fused to RBPs that are not highly multiplexed on their targets. This is further supported by a paucity of RNA edits when the MBS scaffold is absent. When MCP-ADAR is introduced into cells but cannot multiplex given the lack of MBS, the background of TRIBE is far lower than CLIP. Future experiments may modify the number of MS2 repeats on the MBS to characterize the sensitivity of a single enzyme.

Identification of Chromatin Contacts and Transcription Domains

Editing that spans inter-chromosomal transcription sites would only occur if the transcribing loci from nearby chromosomes shared intimate contacts with the β-actin nascent chains. Inter-chromosomal contacts have been identified by Hi-C (Battulin et al., 2015; Lieberman-Aiden et al., 2009) and further supported by transcriptome-wide RNA FISH studies in MEFs that observed nascent transcripts looping away from the chromosomal DNA (on average 0.8 ± 1.1 μm away) (Shah et al., 2018). This implies a possible intermingling with other nascent transcripts even from multiple chromosomes (Shah et al., 2018).

Owing to the limited resolution of ligation-based approaches (Hi-C) and optical microscopy (intron seq-FISH), specific interactions at the nucleotide level could not be previously resolved. In addition, the requirement for edited RNA focuses on loci that are simultaneously transcribing and should reduce contributions from many other sources of DNA/DNA interactions. This functional filter may simplify the extensive inter-chromosomal contacts that have been observed with other approaches.

Recently, proteins that interact with the β-actin-MBS were profiled using MCP fused to the biotin ligase BirA. After controlling for background, nuclear proteins as well as chromatin components were discovered to interact with β-actin-MBS, suggesting similar co-transcriptional RNA-protein interactions when the MBS array is used (Mukherjee et al., 2019). However, TRIBE differs in the mechanism of transcriptional marking: the enzyme is affixed to the transcript rather than an interaction that occurs by chance diffusion of a reactive intermediate. Therefore, the transcripts must be in actual physical contact rather than simply nearby.

Regulatory Considerations

The fact that nascent transcripts intermingle makes it likely that nearby transcripts can share factors; for instance, the effective concentration of splicing factors would be increased by the proximity of the transcripts undergoing splicing. This should significantly increase reaction rates, an explanation for why *in vivo* splicing is so efficient. ADAR may possibly edit excised introns, but as these are short-lived compared with the exonic sequences, it is more likely that the editing occurs before or during splicing. Effects on splicing rates may be testable by assaying a splicing reporter placed near highly transcribed and rapidly spliced genes.

The distance from the MCP (PDB: 2BU1) bound to the stem loops to the editing pocket of the ADAR (PDB: 5ED1) is only a few nanometers, indicating that transcripts come very close to one another. However, nascent chains may explore a much larger volume so they eventually come into contact with each other; this may occur when nascent chains of different transcripts are looped into a constrained volume. Calculations using the MS2 stem loop, the MBS, and β-actin-MBS suggest that this distance be can up to 200 nm away from where the nascent chain is anchored at the transcription site (Figure S7). In either case, collisions of these RNA strands do not result in entanglements and examples of *trans*-splicing are rare. This could best be accomplished if the nascent chains assume a minimal spherical volume rather than a linear dimension, so editing would then occur where these surface regions of exposed nucleotides come into contact.

Because the ADAR is an enzyme, it requires some time to find and modify the appropriate adenosine. This has been estimated to be about 24 s (Kuttan and Bass, 2012). Therefore, ADAR must stay in contact with its substrate longer than would be required for CLIP, where cross-linking is instantaneous. This would provide a filter for more persistent interactions and possibly more physiologically relevant ones (McMahon et al., 2016). In addition, the integration of RNA editing signal over time is an advantage compared with techniques requiring cellular fixation. Expression of RBP fusions (such as in MS2-TRIBE and also in other recently reported methods, Medina-Muñoz et al., 2019) may allow for RBPs to bind to a larger proportion of their targets when integrated over time, leading to more efficient identification of targets.

The discovery of chromatin contacts by MCP-TRIBE suggests a number of future directions. Large nuclear foci consisting of nascent transcripts have been previously observed (Fay et al., 1997), and more recent studies have proposed a transcriptional hub model where multiple transcribing loci spatially organize (Cho et al., 2018; Hnisz et al., 2017). Several theories have been proposed for the function of transcriptional hubs; in particular these data support the role of transcription factors in hub formation (Liu and Tjian, 2018). Consistent with a transcription factor-driven process for RNA-RNA clustering, half of the MCP-TRIBE targets contain upstream binding sites for serum response factor (Roider et al., 2009; Table S5). Further work will determine if additional parameters such as processing and splicing rate are correlated across contacts and if this coordination is mediated by specific transcription factors.

Transcription hubs change during stem cell differentiation and have been shown to contain multiple copies of both mediator and RNA polymerase II (Cho et al., 2018). To date, it is still not known which genes are actively transcribing within these transcriptional hubs. The highly multivalent nature of proteins within the hub, e.g., Mediator, makes it an attractive target for future studies of chromatin contacts with TRIBE. In addition, the discovery of enhancer RNAs makes them an attractive target for MS2 tagging and subsequent TRIBE to determine a list of functional RNA targets by their possible co-editing. The development of

MS2-tagged gene libraries and CRISPR-mediated integration will further allow chromatin contacts to be studied by MCP-TRIBE.

Limitations of the Study

Here we show that MS2-TRIBE provides a standard for determining false-positives across different techniques and show that it has a low false-positive rate compared with CLIP. The low levels of RNA editing in cells without the β-actin-MBS reveal that the observed editing of interacting RNAs near the MBS requires these binding sites for the MCP-ADAR. The use of the MS2 approach has both strengths and limitations. It provides a benchmark for approaches such as CLIP (Ule et al., 2003), Hi-C (Lieberman-Aiden et al., 2009), TSA sequencing (Chen et al., 2018; Zhang et al., 2020), and APEX sequencing (Fazal et al., 2019; Padrón et al., 2019), and it uniquely can interrogate RNA-RNA contacts originating from a single locus. A limitation of the approach is the requirement for MBS to be integrated into a site of interest. Thus, MS2-TRIBE cannot be applied to patient samples to ascertain important protein RNA contacts related to disease, as can be done with CLIP (Luna et al., 2017). However, the genetic engineering required to perform MS2-TRIBE on cultured cells has been made easier with the widespread adaptation of CRISPR insertions (Spille et al., 2019) and coding sequence tagging where hundreds of lines have been made to express MS2 stem loops (Sheinberger et al., 2017; Wan et al., 2019). In addition, a mouse expressing the β-actin gene homozygously tagged with MS2 loops provides a resource for these studies *in vivo* (Lionnet et al., 2011) and hence the transcriptional environment around the actin locus in a variety of tissues.

TRIBE when fused to an RBP of interest can complement CLIP (Ule et al., 2003) to determine RBP binding targets both *cis* and *trans*. Because TRIBE represents a proximity approach to editing, the location of the *cis*-editing sites correlates with the location of RBP binding, but does not provide the exact nucleotides for RBP binding as does CLIP. Additional applications of MS2-TRIBE are being developed; for instance, MS2-based RNA editing within single cells (Rodriques et al., 2020) provides a way to interrogate mRNA age by assessing the increase of edits with time and may be used to determine how *cis*-editing sites evolve with changes in developmental state.

Resource Availability

Lead Contact

Further information and requests for resources and reagents should be directed to and will be fulfilled by the Lead Contact, Dr. Robert H. Singer (robert.singer@einsteinmed.org).

Material Availability

Plasmids encoding mCherry-ADAR (Addgene plasmid #154786) and MCP-ADAR (Addgene plasmid #154787) are available at Addgene.

Data and Code Availability

All raw sequencing data and identified RNA editing sites have been deposited in NCBI's Gene Expression Omnibus (https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE152855) and are accessible through accession number GSE152855. Code used to process and analyze the data is publically available at (https://github.com/rosbashlab/HyperTRIBE/). All other relevant data are available from the authors upon request.

METHODS

All methods can be found in the accompanying Transparent Methods supplemental file.

SUPPLEMENTAL INFORMATION

Supplemental Information can be found online at https://doi.org/10.1016/j.isci.2020.101318.

ACKNOWLEDGMENTS

AUTHOR CONTRIBUTIONS

Conceptualization, J.B. and R.H.S.; Methodology, J.B., V.G., and R.R.; Investigation, J.B., V.G., and R.R.; Writing – Original Draft, J.B. and R.H.S.; Writing – Review & Editing, J.B., R.R., M.R., and R.H.S.; Funding Acquisition, J.B., R.H.S., and M.R.; Resources, M.R. and R.H.S.; Supervision, M.R. and R.H.S.

DECLARATION OF INTERESTS

The authors have no competing interests.

REFERENCES

Battulin, N., Fishman, V.S., Mazur, A.M., Pomaznoy, M., Khabarova, A.A., Afonnikov, D.A., Prokhortchouk, E.B., and Serov, O.L. (2015). Comparison of the three-dimensional organization of sperm and fibroblast genomes using the Hi-C approach. Genome Biol. *16*, 77.

Biswas, J., Nunez, L., Das, S., Yoon, Y.J., Eliscovich, C., and Singer, R.H. (2020). Zipcode binding protein 1 (ZBP1; IGF2BP1): a model for sequence-specific RNA regulation. Cold Spring Harb. Symp. Quant. Biol. 039396.

Chen, Y., Zhang, Y., Wang, Y., Zhang, L., Brinkman, E.K., Adam, S.A., Goldman, R., van Steensel, B., Ma, J., and Belmont, A.S. (2018). Mapping 3D genome organization relative to nuclear compartments using TSA-Seq as a cytological ruler. J. Cell Biol. *217*, 4025–4048.

Cho, W.-K., Spille, J.-H., Hecht, M., Lee, C., Li, C., Grube, V., and Cisse, I.I. (2018). Mediator and RNA polymerase II clusters associate in transcription-dependent condensates. Science *361*, 412–415.

Fay, F.S., Taneja, K.L., Shenoy, S., Lifshitz, L., and Singer, R.H. (1997). Quantitative digital analysis of diffuse and concentrated nuclear distributions of nascent transcripts, SC35 and poly(A). Exp. Cell Res. *231*, 27–37.

Fazal, F.M., Han, S., Parker, K.R., Kaewsapsak, P., Xu, J., Boettiger, A.N., Chang, H.Y., and Ting, A.Y. (2019). Atlas of subcellular RNA localization revealed by APEX-Seq. Cell *178*, 473–490.e26.

Friedersdorf, M.B., and Keene, J.D. (2014). Advancing the functional utility of PAR-CLIP by quantifying background binding to mRNAs and lncRNAs. Genome Biol. *15*, R2.

Gerstberger, S., Hafner, M., and Tuschl, T. (2014). A census of human RNA-binding proteins. Nat. Rev. Genet. *15*, 829–845.

Herzog, J.J., Xu, W., Deshpande, M., Rahman, R., Suib, H., Rodal, A.A., Rosbash, M., and Paradis, S. (2020). TDP-43 dysfunction restricts dendritic complexity by inhibiting CREB activation and altering gene expression. Proc. Natl. Acad. Sci. U S A *117*, 11760–11769.

Hnisz, D., Shrinivas, K., Young, R.A., Chakraborty, A.K., and Sharp, P.A. (2017). A phase separation model predicts key features of transcriptional control. Cell *169*, 13–23.

Hogan, G.J., Brown, P.O., and Herschlag, D. (2015). Evolutionary conservation and diversification of Puf RNA binding proteins and their mRNA targets. PLoS Biol. *13*, e1002307.

Jain, D., Baldi, S., Zabel, A., Straub, T., and Becker, P.B. (2015). Active promoters give rise to false positive "Phantom Peaks" in ChIP-seq experiments. Nucleic Acids Res. *43*, 6959–6968.

Janicki, S.M., Tsukamoto, T., Salghetti, S.E., Tansey, W.P., Sachidanandam, R., Prasanth, K.V., Ried, T., Shav-Tal, Y., Bertrand, E., Singer, R.H., et al. (2004). From silencing to gene expression: real-time analysis in single cells. Cell *116*, 683–698.

Kalo, A., Kanter, I., Shraga, A., Sheinberger, J., Tzemach, H., Kinor, N., Singer, R.H., Lionnet, T., and Shav-Tal, Y. (2015). Cellular levels of signaling factors are sensed by β-actin alleles to modulate transcriptional pulse intensity. Cell Rep. *11*, 419–432.

Kuttan, A., and Bass, B.L. (2012). Mechanistic insights into editing-site specificity of ADARs. Proc. Natl. Acad. Sci. U S A *109*, E3295–E3304.

Lambert, N., Robertson, A., Jangi, M., McGeary, S., Sharp, P.A., and Burge, C.B. (2014). RNA Bind-n-Seq: quantitative assessment of the sequence and structural binding specificity of RNA binding proteins. Mol. Cell *54*, 887–900.

Lapointe, C.P., Wilinski, D., Saunders, H.A.J., and Wickens, M. (2015). Protein-RNA networks revealed through covalent RNA marks. Nat. Methods *12*, 1163–1170.

Lapointe, C.P., Stefely, J.A., Jochem, A., Hutchins, P.D., Wilson, G.M., Kwiecien, N.W., Coon, J.J., Wickens, M., and Pagliarini, D.J. (2018). Multi-omics reveal specific targets of the RNA-binding protein Puf3p and its orchestration of mitochondrial biogenesis. Cell Syst. *6*, 125–135.e6.

Lieberman-Aiden, E., van Berkum, N.L., Williams, L., Imakaev, M., Ragoczy, T., Telling, A., Amit, I., Lajoie, B.R., Sabo, P.J., Dorschner, M.O., et al. (2009). Comprehensive mapping of long range interactions reveals folding principles of the human genome. Science *326*, 289–293.

Lionnet, T., Czaplinski, K., Darzacq, X., Shav-Tal, Y., Wells, A.L., Chao, J.A., Park, H.Y., de Turris, V., Lopez-Jones, M., and Singer, R.H. (2011). A transgenic mouse for in vivo detection of endogenous labeled mRNA. Nat. Methods *8*, 165–170.

Liu, Z., and Tjian, R. (2018). Visualizing transcription factor dynamics in living cells. J. Cell Biol. *217*, 1181–1191.

Luna, J.M., Barajas, J.M., Teng, K., Sun, H.-L., Moore, M.J., Rice, C.M., Darnell, R.B., and Ghoshal, K. (2017). Argonaute CLIP defines a deregulated miR-122 bound transcriptome that correlates with patient survival in human liver cancer. Mol. Cell *67*, 400–410.e7.

McMahon, A.C., Rahman, R., Jin, H., Shen, J.L., Fieldsend, A., Luo, W., and Rosbash, M. (2016). TRIBE: Hijacking an RNA-editing enzyme to identify cell-specific targets of RNA-binding proteins. Cell *165*, 742–753.

Medina-Muñoz, H.C., Lapointe, C.P., Porter, D.F., and Wickens, M. (2019). Records of RNA localization through covalent tagging. BioRxiv. https://doi.org/10.1101/785816.

Mili, S., and Steitz, J.A. (2004). Evidence for reassociation of RNA-binding proteins after cell lysis: implications for the interpretation of immunoprecipitation analyses. RNA *10*, 1692–1694.

Mukherjee, J., Hermesh, O., Eliscovich, C., Nalpas, N., Franz-Wachtel, M., Maček, B., and Jansen, R.-P. (2019). β-Actin mRNA interactome mapping by proximity biotinylation. Proc. Natl. Acad. Sci. U S A *116*, 12863–12872.

Padrón, A., Iwasaki, S., and Ingolia, N.T. (2019). Proximity RNA labeling by APEX-seq reveals the organization of translation initiation complexes and repressive RNA granules. Mol. Cell *75*, 875–887.e5.

Patel, V.L., Mitra, S., Harris, R., Buxbaum, A.R., Lionnet, T., Brenowitz, M., Girvin, M., Levy, M., Almo, S.C., Singer, R.H., et al. (2012). Spatial arrangement of an RNA zipcode identifies mRNAs under post-transcriptional control. Genes Dev. *26*, 43–53.

Rahman, R., Xu, W., Jin, H., and Rosbash, M. (2018). Identification of RNA-binding protein targets with HyperTRIBE. Nat. Protoc. *13*, 1829.

Raj, A., Peskin, C.S., Tranchina, D., Vargas, D.Y., and Tyagi, S. (2006). Stochastic mRNA synthesis in mammalian cells. PLoS Biol. *4*, e309.

Rodrigues, S.G., Chen, L.M., Liu, S., Zhong, E.D., Scherrer, J.R., Boyden, E.S., and Chen, F. (2020). Recording the age of RNA with deamination. BioRxiv. https://doi.org/10.1101/2020.02.08. 939983.

Roider, H.G., Manke, T., O'Keeffe, S., Vingron, M., and Haas, S.A. (2009). PASTAA: identifying transcription factors associated with sets of co-regulated genes. Bioinform. Oxf. Engl. *25*, 435–442.

Shah, S., Takei, Y., Zhou, W., Lubeck, E., Yun, J., Eng, C.-H.L., Koulena, N., Cronin, C., Karp, C., Liaw, E.J., et al. (2018). Dynamics and spatial genomics of the nascent transcriptome by intron seqFISH. Cell *174*, 363–376.e16.

Sheinberger, J., Hochberg, H., Lavi, E., Kanter, I., Avivi, S., Reinitz, G., Schwed, A., Aizler, Y., Varon, E., Kinor, N., et al. (2017). CD-tagging-MS2: detecting allelic expression of endogenous mRNAs and their protein products in single cells. Biol. Methods Protoc. *2*, 1–14.

Spille, J.-H., Hecht, M., Grube, V., Cho, W.-K., Lee, C., and Cissé, I.I. (2019). A CRISPR/Cas9 platform for MS2-labelling of single mRNA in live stem cells. Methods San Diego Calif. *153*, 35–45.

Teytelman, L., Thurtle, D.M., Rine, J., and Oudenaarden, A.van (2013). Highly expressed loci are vulnerable to misleading ChIP localization of multiple unrelated proteins. Proc. Natl. Acad. Sci. U S A *110*, 18602–18607.

Tutucci, E., Vera, M., Biswas, J., Garcia, J., Parker, R., and Singer, R.H. (2018). An improved MS2 system for accurate reporting of the mRNA life cycle. Nat. Methods *15*, 81–89.

Ule, J., Jensen, K.B., Ruggiu, M., Mele, A., Ule, A., and Darnell, R.B. (2003). CLIP Identifies nova-regulated RNA networks in the brain. Science *302*, 1212–1215.

Vera, M., Biswas, J., Senecal, A., Singer, R.H., and Park, H.Y. (2016). Single-cell and single-molecule analysis of gene expression regulation. Annu. Rev. Genet. *50*, 267–291.

Wan, Y., Anastasakis, D.G., Rodriguez, J., Palangat, M., Gudla, P., Zaki, G., Tandon, M., Pegoraro, G., Chow, C.C., Hafner, M., et al. (2019). Dynamic Imaging of Nascent RNA Reveals General Principles of Transcription Dynamics and Stochastic Splice Site Selection (Social Science Research Network).

Wu, B., Chao, J.A., and Singer, R.H. (2012). Fluorescence fluctuation spectroscopy enables quantitative imaging of single mRNAs in living cells. Biophys. J. *102*, 2936–2944.

Wu, B., Buxbaum, A.R., Katz, Z.B., Yoon, Y.J., and Singer, R.H. (2015). Quantifying protein-mRNA interactions in single live cells. Cell *162*, 211–220.

Xu, W., Rahman, R., and Rosbash, M. (2018). Mechanistic implications of enhanced editing by a HyperTRIBE RNA-binding protein. RNA *24*, 173–182.

Zhang, C., and Darnell, R.B. (2011). Mapping *in vivo* protein-RNA interactions at single-nucleotide resolution from HITS-CLIP data. Nat. Biotechnol. *29*, 607–614.

Zhang, L., Zhang, Y., Chen, Y., Gholamalamdari, O., Wang, Y., Ma, J., and Belmont, A.S. (2020). TSA-Seq reveals a largely "hardwired" genome organization relative to nuclear speckles with small position changes tightly correlated with gene expression changes. BioRxiv, 607–614.

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

**Validation of protein / RNA interaction
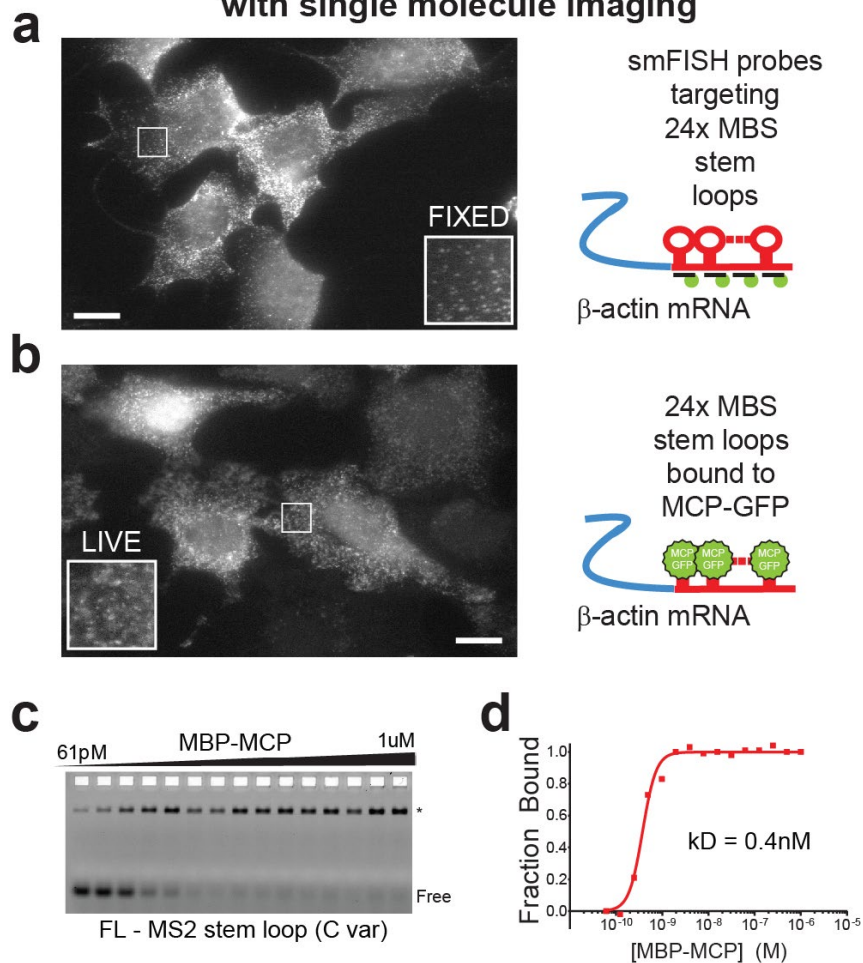with single molecule imaging**

**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).

a MCP TRIBE replicates in cells containing MBS

Y axis: Editing percentage at individual sites, Biological Replicate 2

X axis: Editing percentage at individual sites, Biological Replicate 1

$R^2 = 0.6678$

b MCP CLIP replicates in cell containing MBS

Y axis: Peak height of individual transcripts, Biological Replicate 2

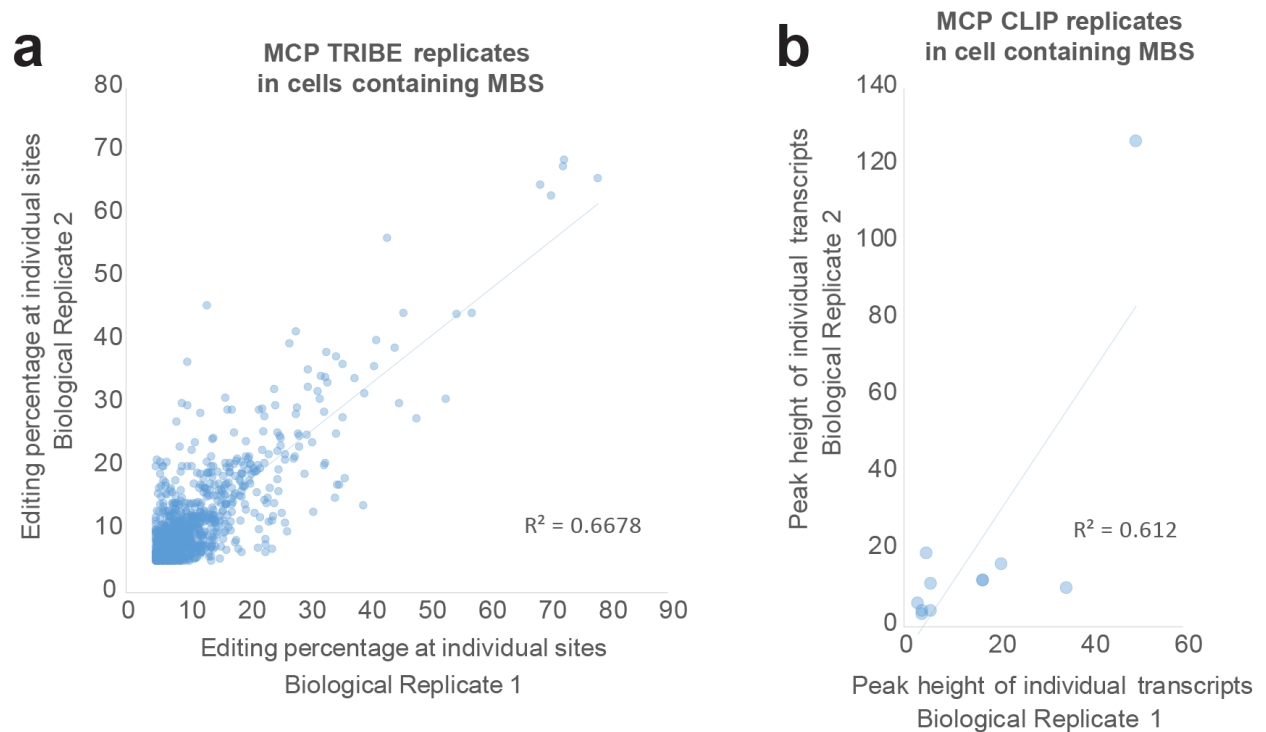X axis: Peak height of individual transcripts, Biological Replicate 1

$R^2 = 0.612$

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.
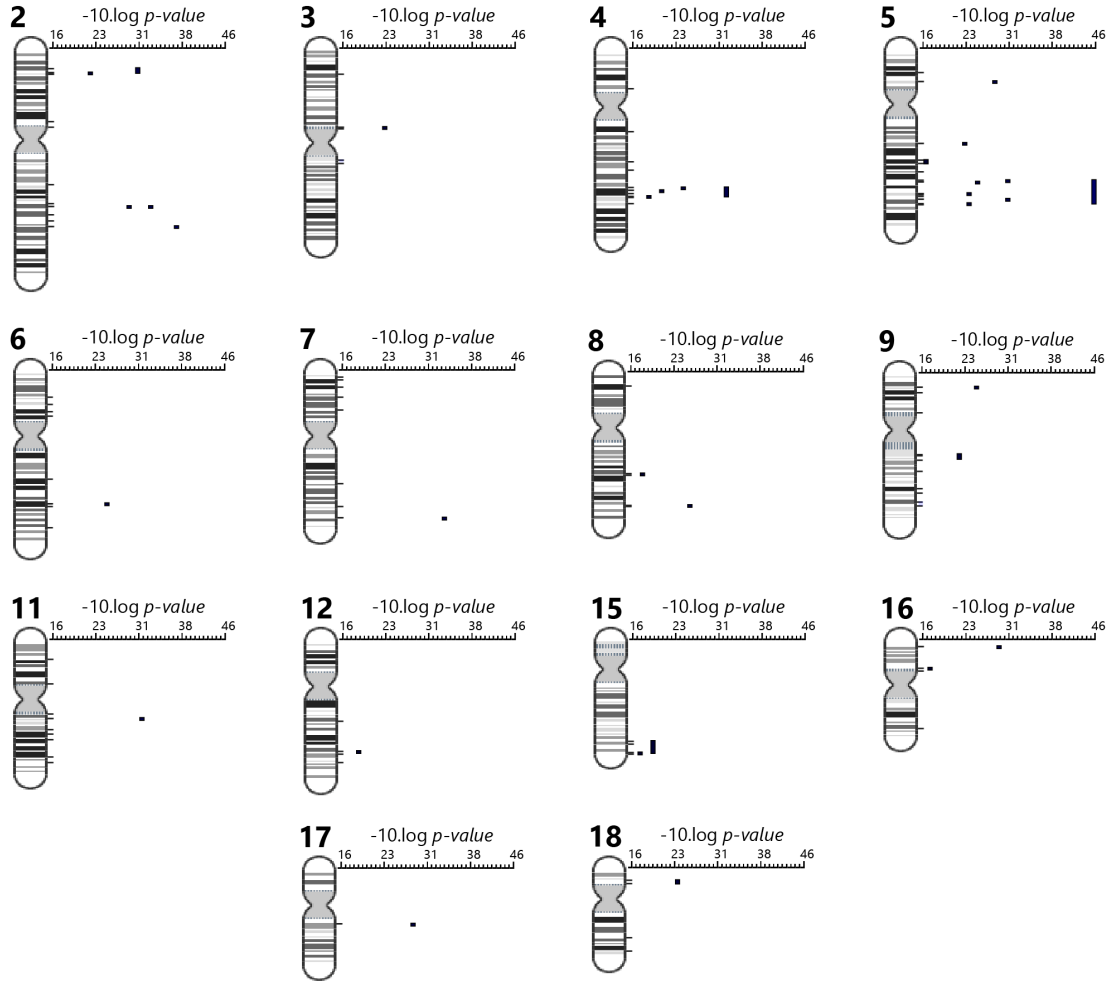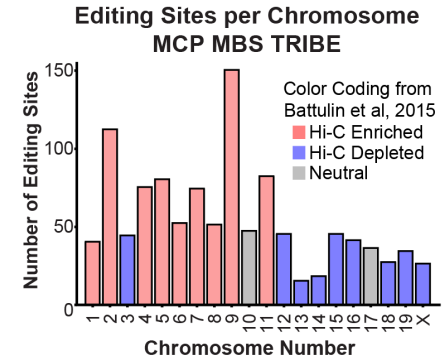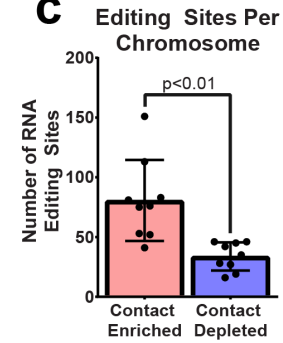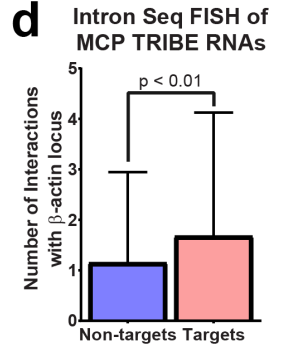
a

b Editing Sites per Chromosome MCP MBS TRIBE

Color Coding from Battulin et al, 2015
Hi-C Enriched
Hi-C Depleted
Neutral

c Editing Sites Per Chromosome

p<0.01

d Intron Seq FISH of MCP TRIBE RNAs

p < 0.01

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.

a Fraction of sequenced reads that were uniquely mapped

b Fraction of sequenced reads used after unique mapping and duplicate removal

c Subsampled MCP-MBS Transcripts
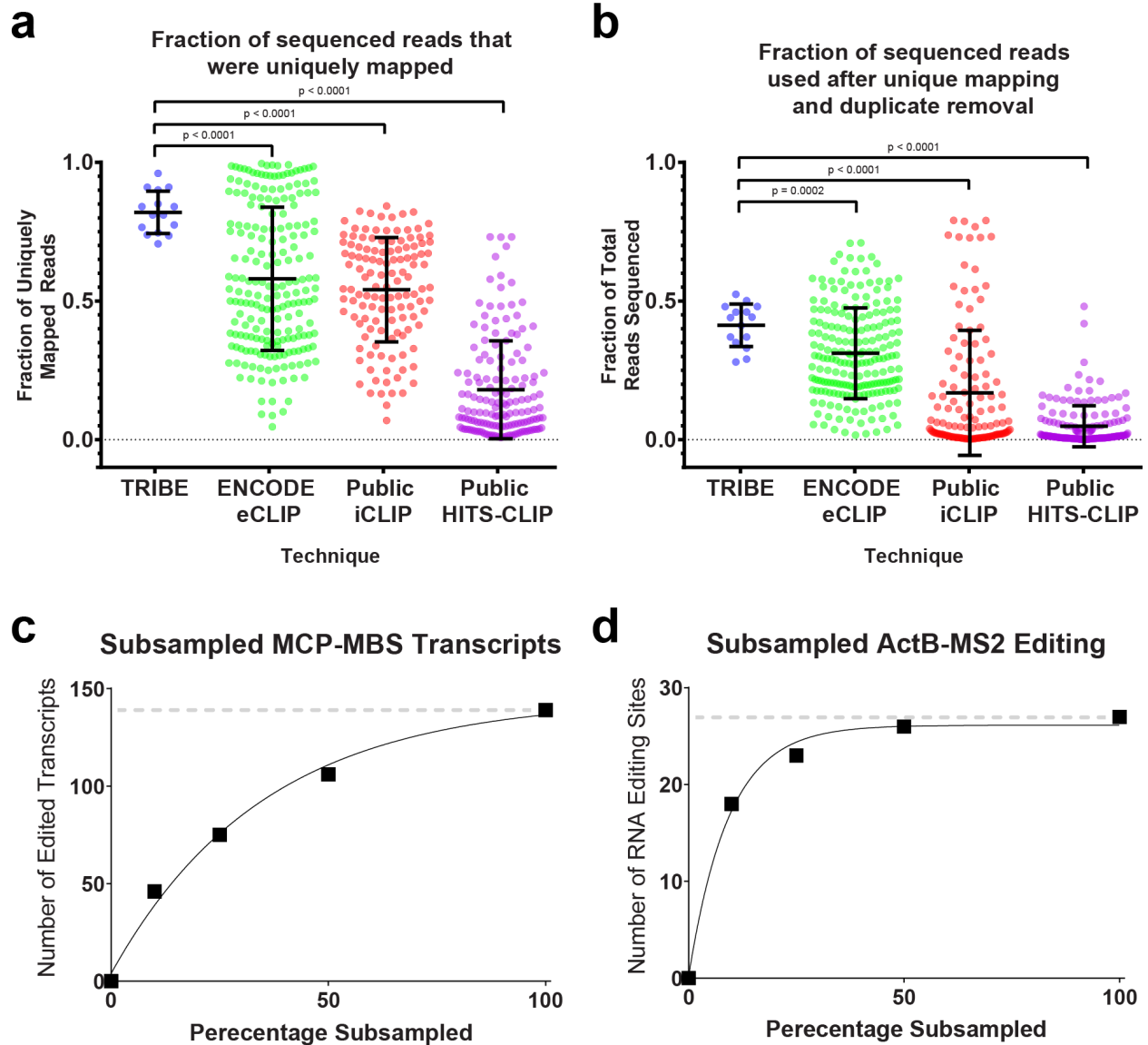
d Subsampled ActB-MS2 Editing

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

5

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).

a

Live imaging of MCP-GFP transfected U2OS cells containing the MBS reporter

Uninduced (no dox)

24 hour induction with 1uM dox

b

mCherry-ADAR Control Gene Expression (FPKM)

MCP-ADAR Gene Expression (FPKM)

	mCherry-ADAR Control	MCP-ADAR
p-tet on transfection	X X	X X X
mCherry-ADAR transfection	X X	
MCP-ADAR transfection		X X X
Dox induction	X	X X

c

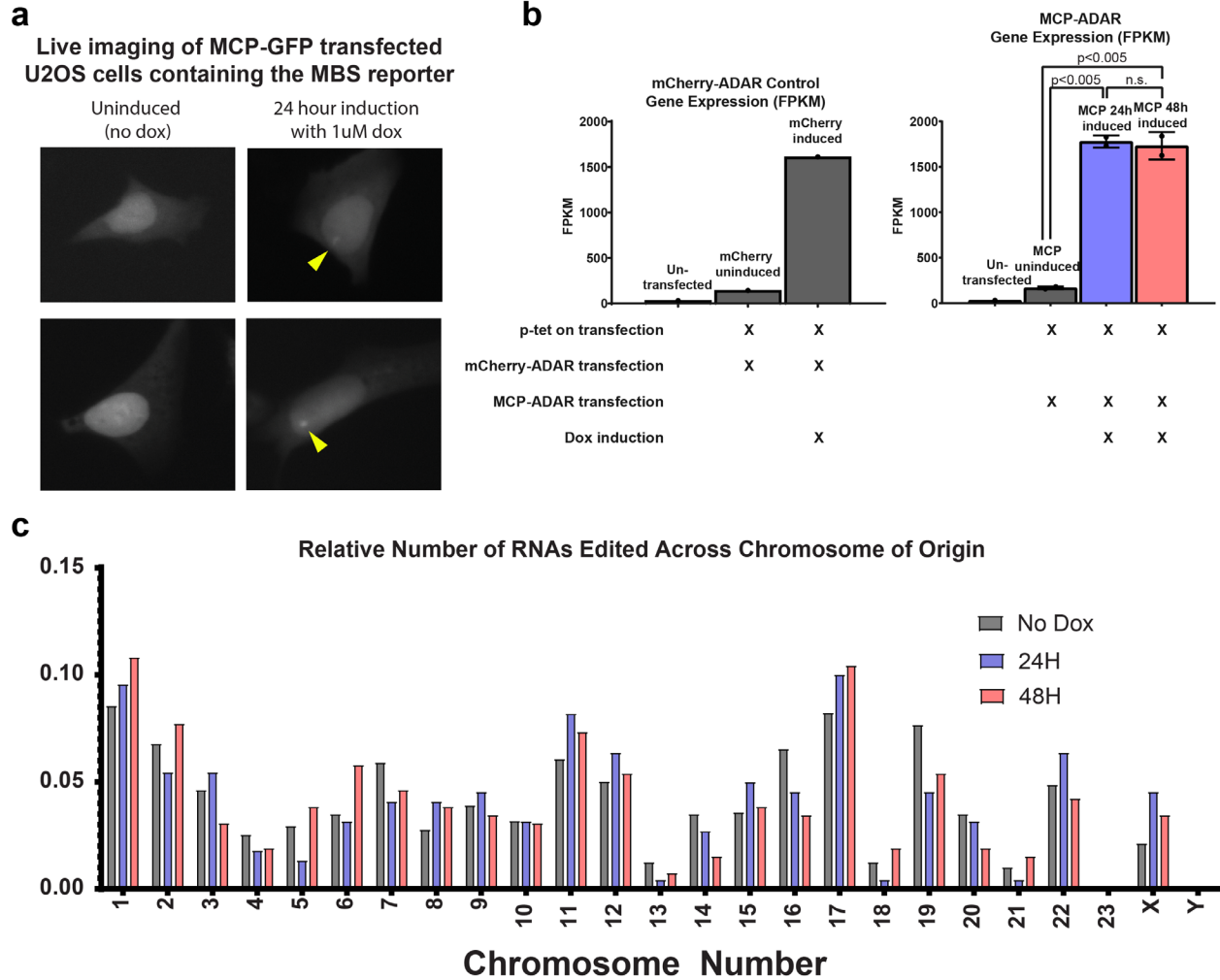Relative Number of RNAs Edited Across Chromosome of Origin

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.
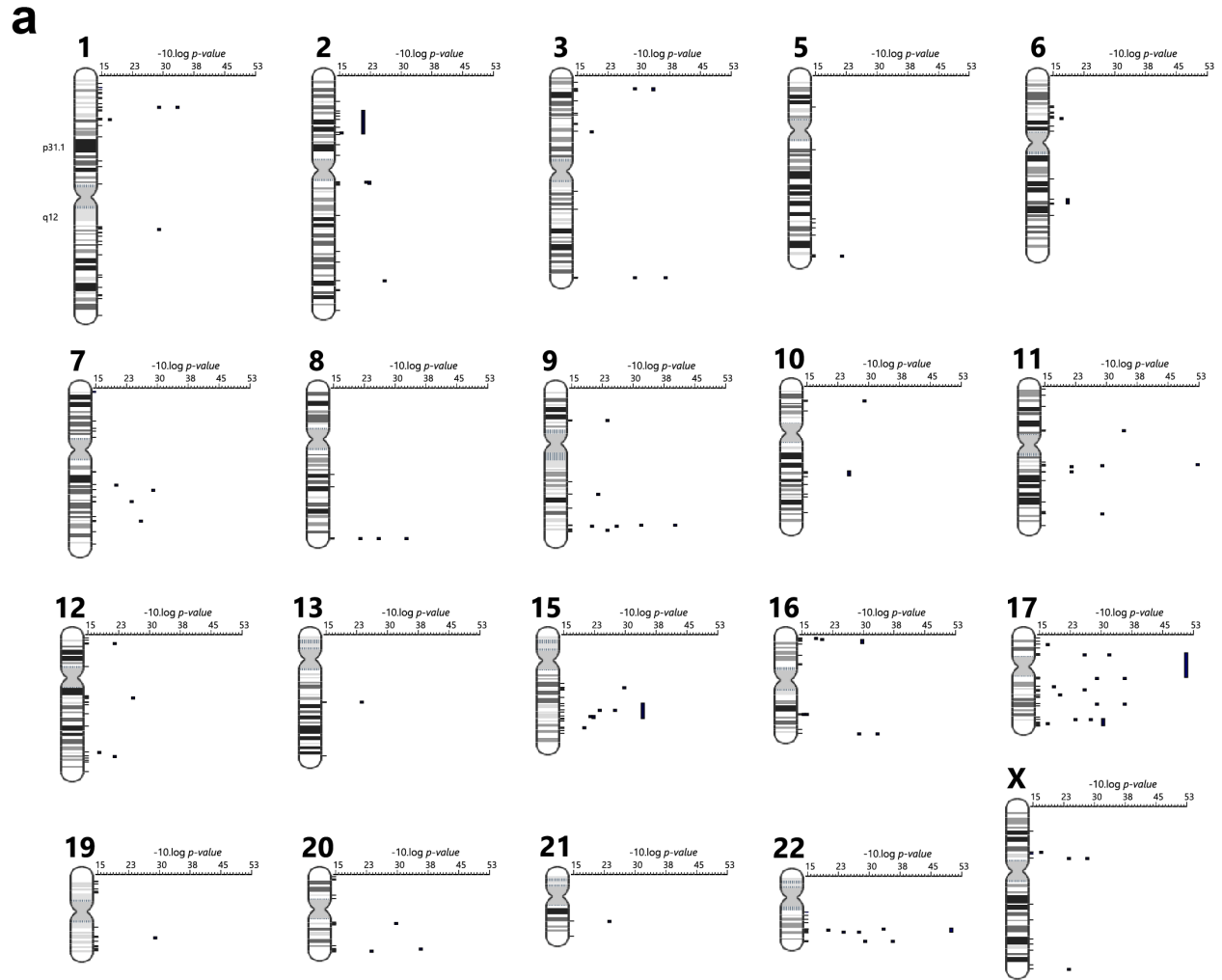
7

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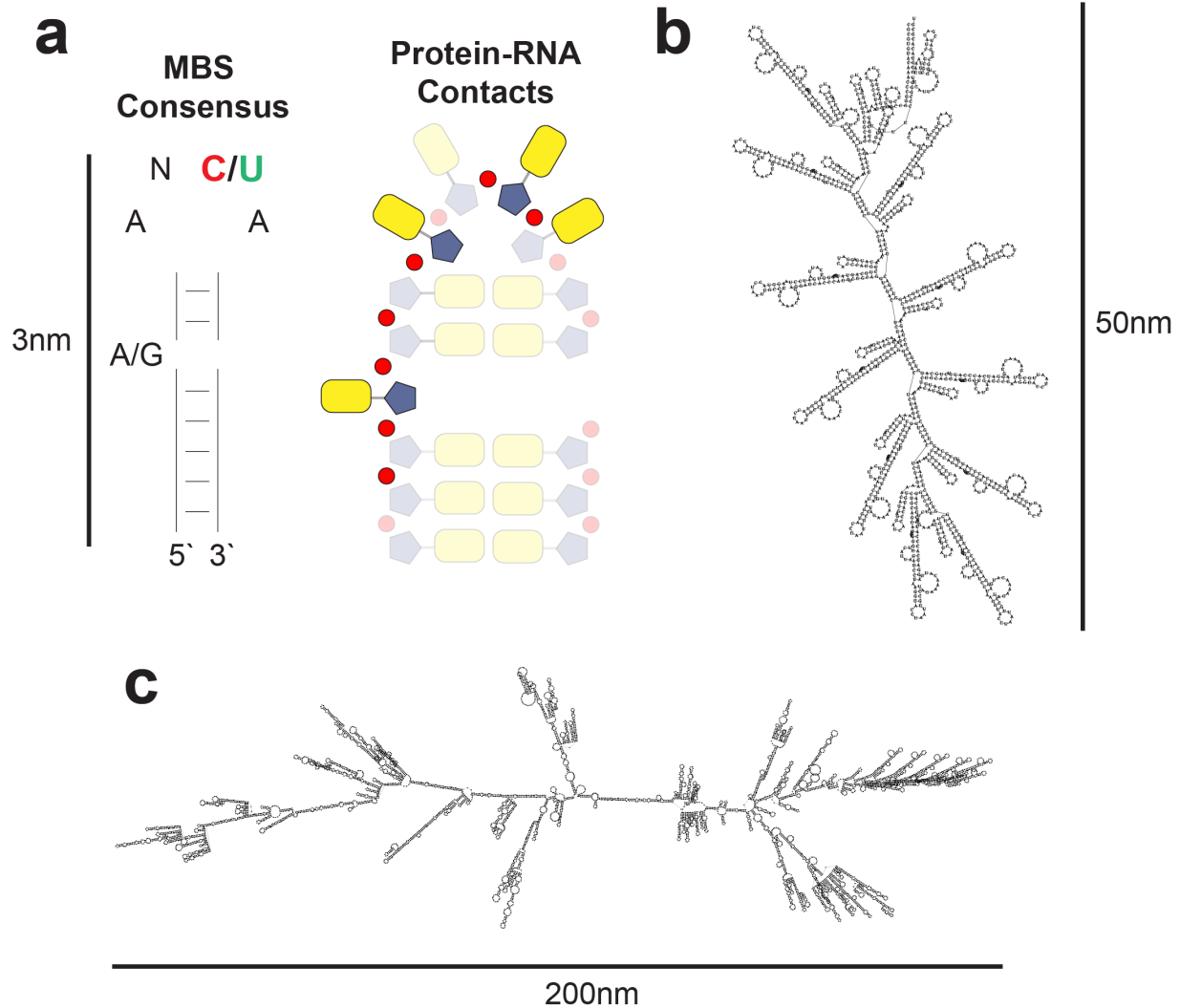
Figure S7. Estimation of RNA sizes for MS2 size, MBS and β-actin-MBS. Related to figures 3 and 4.

(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.