Supplemental Material

Supplementary Materials and Methods

Oligonucleotides, siRNA duplexes and fluorescent hybridization probes

I. The following oligodeoxynucleotides were used for plasmid preparation, mutagenesis reactions and sequencing:

Plasmids:

RBPMS_PCR_for, ACGCGTCGACATGAACAACGGCGGCAAAGC;

RBPMS_PCR_rev, ATAGTTTAGCGGCCGCTCAGCAGAACTGACGGGAC;

RBPMS2_PCR_for,

AAAAAGCAGGCTTCATGAGCAACCTGAAGCCGGACGGCGA;

RBPMS2_PCR_rev, **AGAAAGCTGGGTG**TCAACAGAACTGACGGTAC;

Protein expression:

RBPMS_pET23a_SalI_for_FL, ACGCGTCGACCCATGAACAACGGCGGCAAAGC;

RBPMS_pET23a_NotI_rev_FL,

ATAGTTTAGCGGCCGCGCAGAACTGACGGGACTTC;

RBPMS_pET23a_NotI_rev_F_Cterm100, ATAGTTTAGCGGCCGCCTTAGCAAA

CTCTAGTCGTAGTG

RBPMS_pET23a_SalI_for_Nterm20, ACGCGTCGAC**CC**ATGGAGGAGGAGGTC

CGGACCCT

RBPMS_pET23a_NotI_rev_F_Cterm111, ATAGTTTAGCGGCCGC

TACGAGTTTGTTCTTGGCC

RBPMS_pET23a_NotI_rev_F_Cterm120, ATAGTTTAGCGGCCGC

CAGAGGAGTACTGGGGTTTGG

RBPMS_pET23a_NotI_rev_F_Cterm129, ATAGTTTAGCGGCCGCGGCAATGAA

CTGAGGTACAG

RBPMS_pET23a_NotI_rev_F_Cterm144, ATAGTTTAGCGGCCGCGCTACTGGG

GTAAAGTGC

RBPMS2_pET23a_mod_for_SalI, ACGCGTCGAC**CC**ATG

AGCAACCTGAAGCCGGAC;

RBPMS2_pET23a_mod_rev_NotI,

ATAGTTTAGCGGCCGCACAGAACTGACGGTACTTCC

RBPMS_pET28a_rev_NotI, ATAGTTTAGCGGCCGC **TCA** GCA GAA CTG ACG

GGA CTT C

II. The following oligoribonucleotides were used for RBPMS protein *in vitro* binding studies:

Mono-, di-, tri- nucleotide panel

AC_dinucleotide_repeat(9)

ACACACACACACACACAC;

AU_dinucleotide_repeat(9)

AUAUAUAUAUAUAUAUAU;

GC_dinucleotide_repeat(9)

GCGCGCGCGCGCGCGCGC;

GU_dinucleotide_repeat(9)

GUGUGUGUGUGUGUGUGU;

CU_dinucleotide_repeat(9)

CUCUCUCUCUCUCUCUCU;

AG_dinucleotide_repeat(9)

AGAGAGAGAGAGAGAGAG;

CU_dinucleotide_repeat(9)

CUCUCUCUCUCUCUCUCU

CAC_trinucleotide_repeat(6)

CACCACCACCACCACCAC

CAU_trinucleotide_repeat(6)

CAUCAUCAUCAUCAUCAU

AAU_trinucleotide_repeat(6)

AAUAAUAAUAAUAAUAAU;

AUU_trinucleotide_repeat(6)

AUUAUUAUUAUUAUUAUU;

GCC_trinucleotide_repeat(6)

GCCGCCGCCGCCGCCGCC;

GGC_trinucleotide_repeat(6)

GGCGGCGGCGGCGGCGGC;

GUU_trinucleotide_repeat(6)

GUUGUUGUUGUUGUUGUU;

GGU_trinucleotide_repeat(6)

GGUGGUGGUGGUGGUGGU;

AAAGGGGGGGGGGGGGG;

polyU(18);

polyA(18);

polyC(18);

AC_dinucleotide_repeat(3)

ACACAC;

ACACACACA;

AC_dinucleotide_repeat(6)

ACACACACACAC;

ACACACACACACACA;

PAR-CLIP synthetic oligoribonucleotides used for EMSAs

NDUFA6_wt: GAGCACAAAUAAACUCACUAU;

NDUFA6_mut1: GAGCACAAAUAAACUC**C**CUAU;

NDUFA6_mut2: GAGC**C**CAAAUAAACUCACUAU;

NDUFA6_mut3: GAGCACAAAUA**C**ACUCACUAU;

NDUFA6_mut4: GAGCACAAA**C**A**C**AACUCACUAU;

NDUFA6_mut5: GAGCACA**C**A**C**AAACUCACUAU;

NDUFA6_mut6: GAGC**C**CAAAUAAACUC**C**CUAU;

NDUFA6_mut7: CACAAAUACACUCAC;

NDUFA6_mut8: CACAAAUAAACUCAC;

ETF1_wt: AGGCACUUUCAACUUCACUCC;

ETF1_mut: AGGC**C**CUUUCAACUUC**C**CUCC;

SRM_wt: CCCUCACUCACCAAACACGUG;

SRM_mut: CCCUC**C**CUCACCAAAC**C**CGUG;

UBE2V1_wt: UACUCACUCACUCUCCACCAG;

UBE2V1_mut: UACUC**C**CUCACUCUCCCCCAG;

BCLAF1_4SU_1: UUCA(4SU)CACAUCCCU(4SU)CACGAA

(residues in bold were mutated in the synthetic RNAs; CAC motif is underlined)

Pre-annealed siRNA duplexes targeting RBPMS were purchased from Applied Biosystems:

R2 (s21729), GCCUUCCUCUGGAUAUCAAtt sense,

UUGAUAUCCAGAGGAAGGCca antisense;

R3 (s21730), GGCUAUGAGGGUUCUCUUAtt sense,

UAAGAGAACCCUCAUAGCCct, antisense.

Fluorescent hybridization probes:

Poly(A): TtTtTtTtTtTtTtTtTtT

28S rRNA: GGTtCCtCtCGtACTGa

(residues in small caps denote LNA modification)

Supplementary Figure Legends

Supplementary Figure 1: RBPMS and RBPMS2 expression in adult human tissues and cell lines. A. Adult human tissues. Illumina BodyMap2 poly(A) RNAseq data from 16 normal human tissues (2x50 nt, unstranded) downladed from ArrayExpress (accession number E-MTAB-513). Reads were aligned to Hg19 assembly using TopHat and

ENSEMBL and gene expression was quantified using HTseq-count. RPKM-type expression levels were calculated by normalizing for mRNA length and for the total number of ENSEMBL-mapped reads. **B. Cell lines.** RNAseq data from 7 cell lines was downloaded from the UCSC ENCODE repository, and subsequently processed similarly to the BodyMap2 data above. A549 (epithelial cell line derived from lung carcinoma tissue), GM12878 (B-lymphocyte, lymphoblastoid, International HapMap Project – CEPH/Utah – European Caucasian Epstein-Barr Virus), H1-hESC (embryonic stem cells), HeLa-S3 (cervical carcinoma), HepG2 (hepatocellular carcinoma), HSMM (skeletal muscle myoblasts), HUVEC (umbilical vein endothelial cells).

Supplementary Figure 2: A. Distribution of genomic location of PAR-CLIP clusters with ≥0.5 T-to-C conversion specificity for library A and B. **B.** Distribution of genomic location of PAR-CLIP reads in clusters with ≥0.5 T-to-C conversion specificity for library A and B. **C.** Overlap of genes containing PAR-CLIP clusters with ≥0.5 T-to-C conversion specificity between library A and B. **D.** Distribution of genomic location of reads in RBPMS binding sites. **E.** Distribution of reads in RBPMS binding sites in library A and B with ≥ 0.5 T-to-C conversion specificity.

Supplementary Figure 3: Analysis of replicate RBPMS PAR-CLIP libraries. A. Scatter plot and Spearman correlation of cumulative number of reads from all clusters per gene, as well as total clusters per gene, between the two libraries. **B.** Clusters identified in both libraries display a higher T-to-C conversion specificity compared to clusters

identified only in one library. **C.** Concordance of number of reads belonging to RBPMS binding sites identified in both libraries.

Supplementary Figure 4: Position of RBPMS binding sites across genomic regions. Density plots representing density of sites (y-axis) for each specific distance between binding sites (x axis, in nt). N represents number of clusters found in the nucleotide range indicated by the x-axis. RBPMS binding site distribution in different genomic locations (red line). Median (black), interquartile range (dark blue), interdecile range (light blue), and the outliers (grey lines) are denoted for the background.

Supplementary Figure 5: RBPMS binding specificity. A. Supershift analysis, through addition of anti-HA and anti-FLAG antibodies, confirmed that the shift of the $(AC)_9$ RNA migration was indeed due to binding of the FLAG-HA-tagged protein. Addition of anti-HA and anti-FLAG antibodies did not result in a shift of the $(AAU)_6$ RNA migration. **B.** Binding specificity for the $(AC)_9$ RNA was not altered by the FLAG-HA or His_6-tag used for protein purification.

Supplementary Figure 6: PARalyzer clusters used as the basis for synthetic RNAs utilized in EMSAs. CAC motifs are illustrated in red, while residues substituted for 4SU are depicted in green.

Supplementary Figure 7: *In vitro* crosslinking with recombinant RBPMS and synthetic radiolabeled RNA containing two 4SU residues flanking the tandem CAC motif based on

the BCLAF1 binding cluster. Bands were separated on SDS-PAGE and visualized using the phosphorimager, revealing a major and minor band at ~20 and ~40 kDa for the crosslinked full-length RBPMS protein and ~30 and ~55 kDa, for the N-terminal truncated RBPMS protein on SDS-PAGE, but only one band at ~15 and ~12 kDa for the two C-terminal truncated RBPMS proteins. Protein concentration as indicated, RNA at 10 nM. Bottom panel represents Coomassie stained 15% SDS-PAGE.

Supplementary Figure 8: **Effect of RBPMS knockdown and overexpression on target stability. A. RBPMS knockdown.** Western Blotting demonstrates successful knockdown (KD) of the FLAG-HA-tagged RBPMS overexpressing HEK293 cell line using a final 25 nM siRNA concentration (R2, R3, or mixture) and lipofectamine RNAimax after 72 h. The response of transcripts categorized by the average number of RBPMS binding sites (colored lines) was assessed after RBPMS knockdown using siRNA R2 relative to mock transfection. The fold change in expression was corrected for siRNA off-target effects and transcript length as described in Methods (also see below); similar results were obtained using siRNA R3 (data not shown). **B. Detection and correction of siRNA off-target effects.** We detected R2 siRNA off-target effects using miReduce (antisense R2 siRNA strand) and corrected for this effect categorizing genes based on the presence of one or multiple 7mer or 7mer1A seed matches (plot on left); plot on right shows the correction (see Methods for description). **C. RBPMS overexpression.** The response of transcripts categorized by the average number of RBPMS binding sites (colored lines) was assessed after RBPMS overexpression after 72 h doxycycline induction. Results are shown for the poly(A) RNAseq analysis (see

Methods for description). Similar trend, but smaller magnitude effects, were obtained by microarray analysis (data not shown).

Supplementary Figure 9: RBPMS localizes to cytoplasmic granules after arsenite administration. For assay conditions and description see Methods. **A.** Evaluation of RBPMS localization after 100 and 400 μM 30 min arsenite administration in HEK293 cells inducibly expressing FLAG-HA-tagged RBPMS (with and without doxycycline induction). RBPMS cytoplasmic granules co-localize with poly(A) cytoplasmic granules. **B.** Evaluation of RBPMS localization after 0 and 100 μM 4SU administration. **C.** Evaluation of G3BP1 localization, a known cytoplasmic stress granule marker, after 100 and 400 μM 30 min arsenite administration using HEK293 cells inducibly expressing FLAG-HA-tagged G3BP1 (with and without doxycycline induction). G3BP1 cytoplasmic granules co-localize with poly(A) cytoplasmic granules. **D.** Evaluation of G3BP1 localization after 0 and 100 μM 4SU administration.

Supplementary Table legends:

Supplementary Table 1: PAR-CLIP summary read statistics for library A and library B, shown on top (reads filtered for ambiguous calls and length (<20 nt)). Distribution of reads, clusters and genes across genomic locations, shown below. Columns for each annotation category reflect total number of reads, T-to-C fraction, conversion specificity, fraction of reads that end in G, redundant sequence fraction, redundant copy fraction, average cluster length.

Supplementary Table 2: PARalyzer cluster definition and cluster location, characteristics, repeat and motif content. Cluster genomic location (1-4); alignment category (5); gene name (6); annotation category (7); cluster ID (8); cluster sequence (9); cluster read count (10); conversion location count, number of conversion locations (11); conversion event count, number of conversion events (12); non conversion event count (13); T-to-C fraction, T-to-C conversion fraction (14-15); conversion specificity (16); fraction of reads that end in G (17); redundant sequence fraction (number of sequences with multiple copies out of the number of different sequences) (18); redundant copy fraction (number of redundant read copies out of the total number of read copies) (19); number of unique reads (20); number of distinct conversion events: none, other than T-to-C, T-to-C (21 through 23); RRE definition, CAC or NA (24); RRE count (25); presence of cluster in either single or replicate libraries indicated accordingly, assessing clusters across libraries with full overlap (26). Results shown for library A and library B. Tab C provides the information from both libraries defining the RBPMS binding sites.

Supplementary Table 3: Gene Ontology (GO) analysis of the genes containing RBPMS binding sites.

Supplementary Table 4: Gel filtration results for His_6-tagged RBPMS deletion proteins.

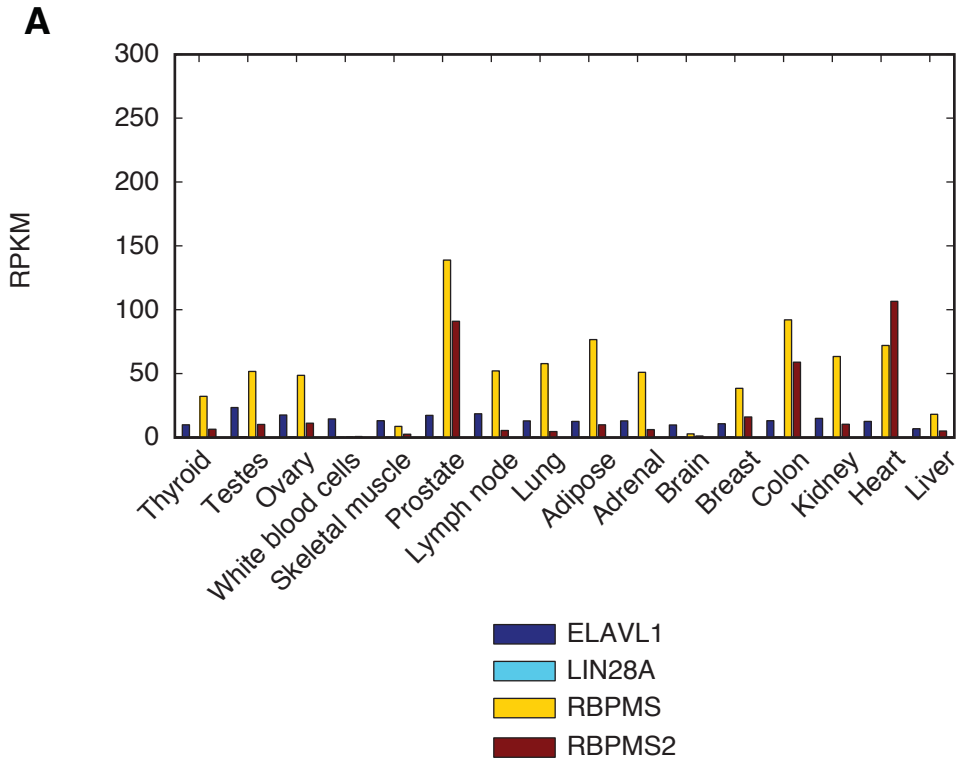
Supplementary Table 5: miReduce analysis. 7-mers identified to be over-represented in up-regulated genes after RBPMS knockdown (top), and down-regulated genes after RBPMS overexpression (bottom).

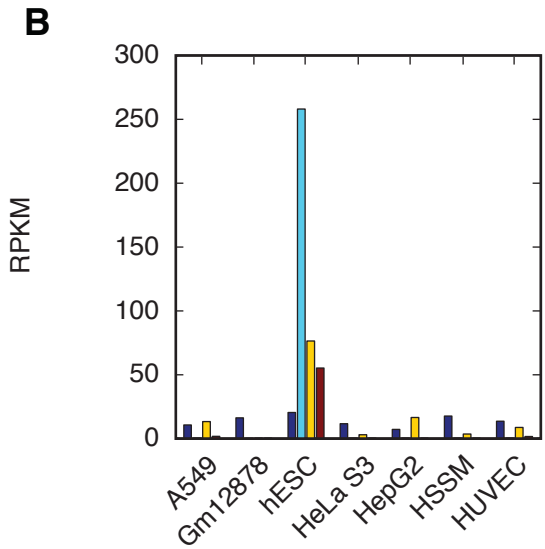
Supplementary Table 6: Differential exon usage after RBPMS overexpression in HEK293 cells. Generalized linear model identified statistically significant (adjusted p-value $\leq$0.1) differential exon usage while controlling for differences at the level of gene expression, comparing poly(A) RNAseq expression in RBPMS overexpressing HEK293 cells with uninduced cells. Tab B shows distance of closest RBPMS binding site from the splice site. Genes with binding sites within 100 nt from the splice site are highlighted in yellow.

Supplementary Table 7: PARalyzer parameters used for analysis.

Supplementary Table 8: Datasets used for RBPMS family expression during mouse embryogenesis, germ cell development and mouse adult tissues.

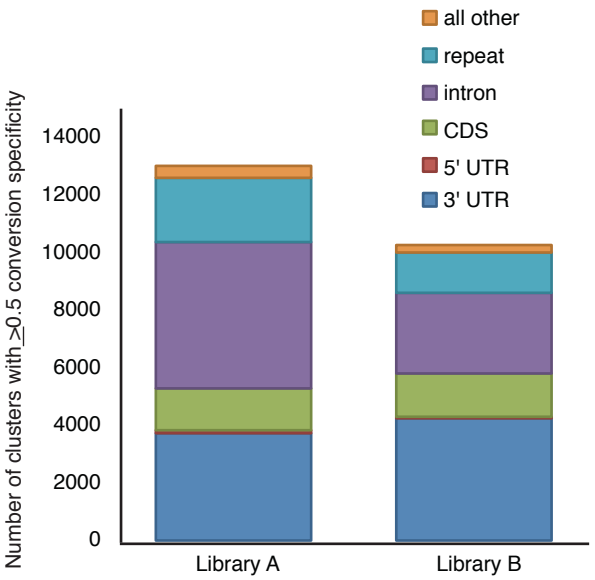
Farazi_Supplementary Figure 1



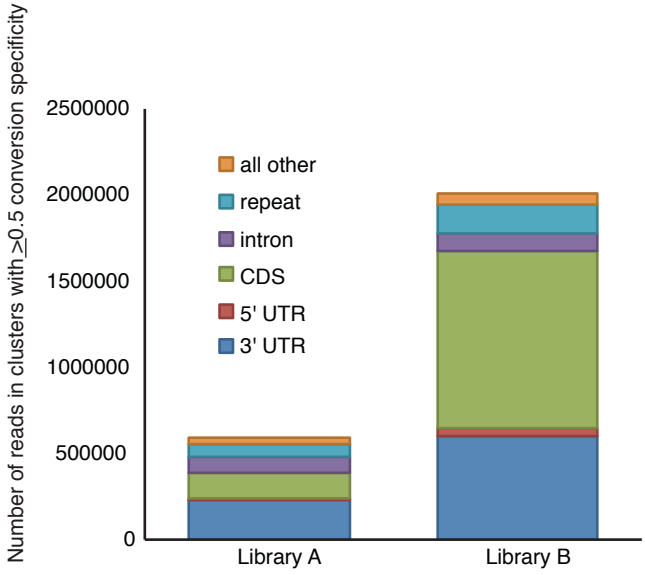


Farazi_Supplementary Figure 2

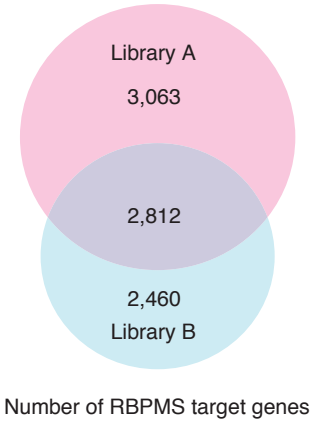
A



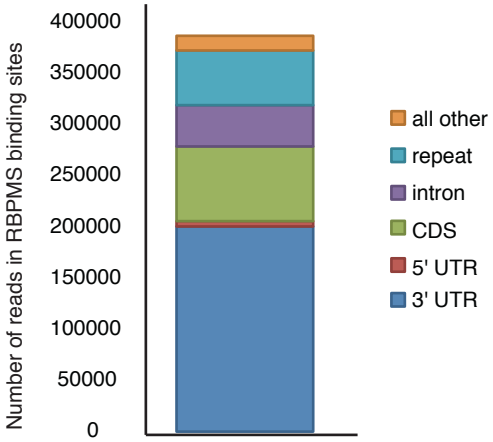
B



C

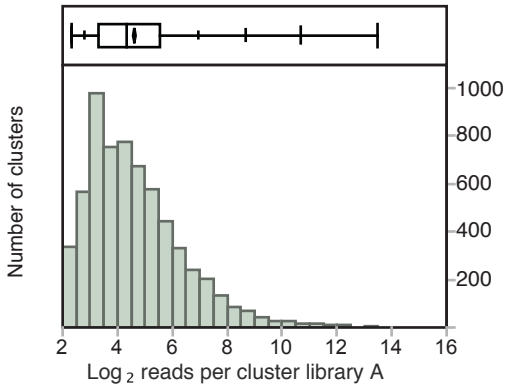


D



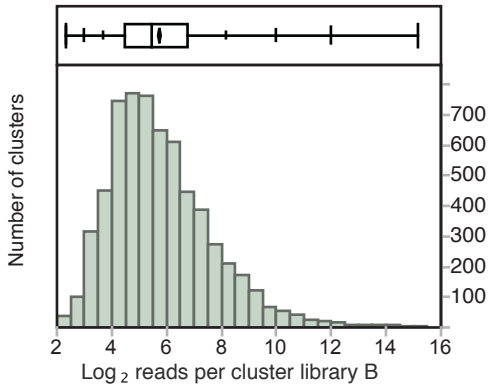
E

RBPMS binding sites



Summary statistics

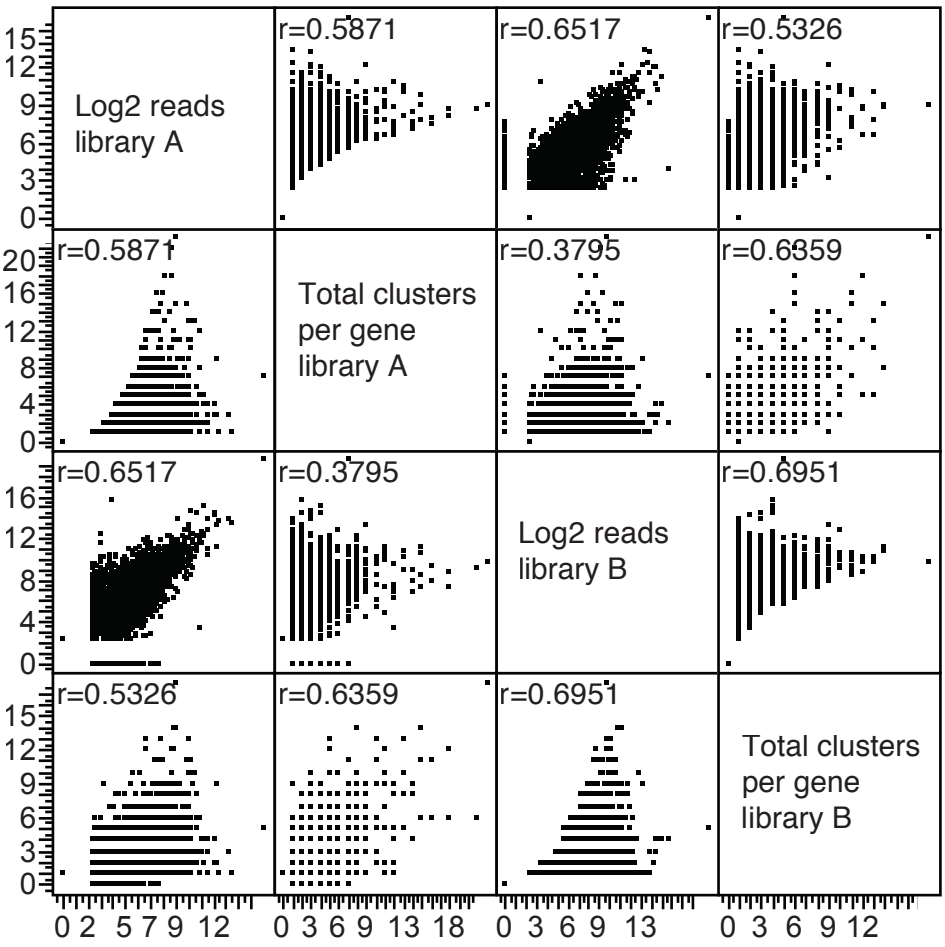
Mean	4.60
Std Deviation	1.68
Std Error Mean	0.02
Upper 95% Mean	4.65
Lower 95% Mean	4.56
N	6,207



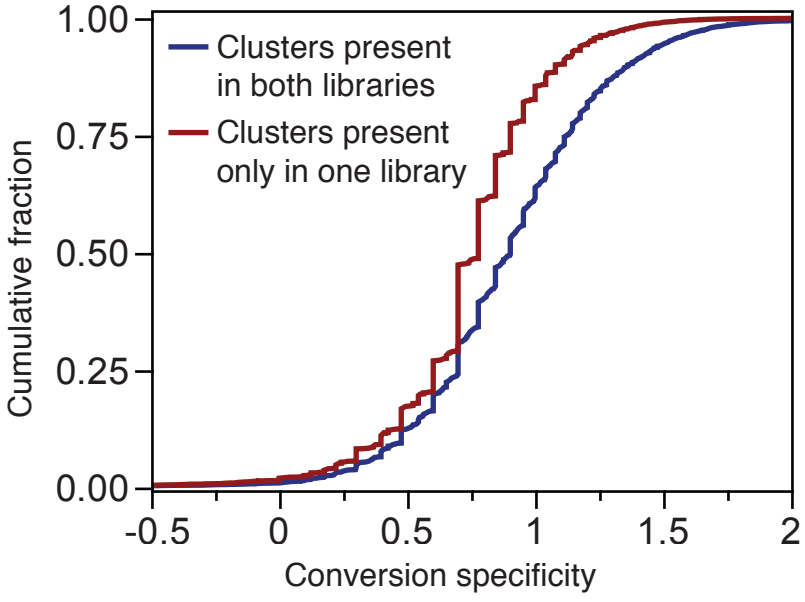
Summary statistics

Mean	5.75
Std Deviation	1.81
Std Error Mean	0.02
Upper 95% Mean	5.79
Lower 95% Mean	5.70
N	6,207

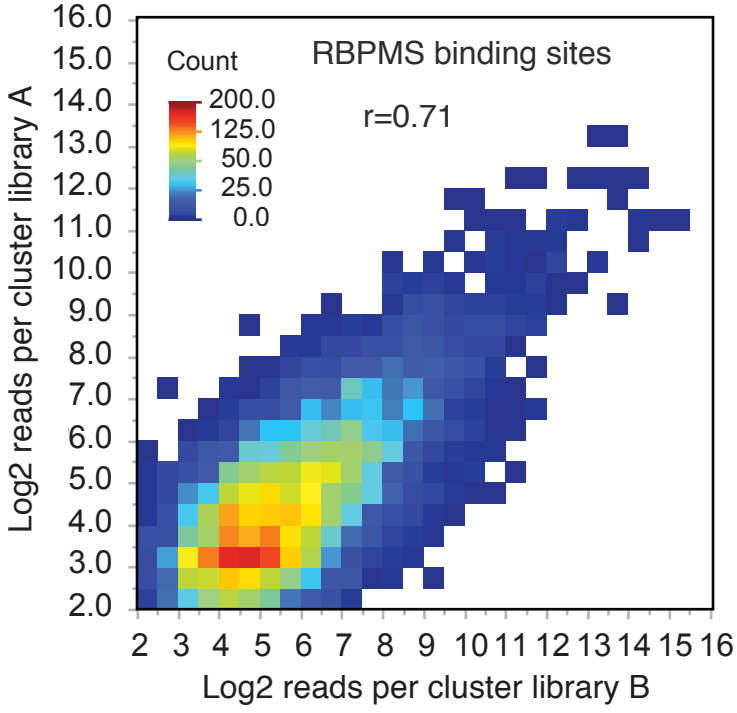
A


Log2 reads library A
r=0.5871
r=0.6517
r=0.5326
r=0.5871
Total clusters per gene library A
r=0.3795
r=0.6359
r=0.6517
r=0.3795
Log2 reads library B
r=0.6951
r=0.5326
r=0.6359
r=0.6951
Total clusters per gene library B
0 2 5 7 9 12
0 3 6 9 13 18
0 3 6 9 13
0 3 6 9 12

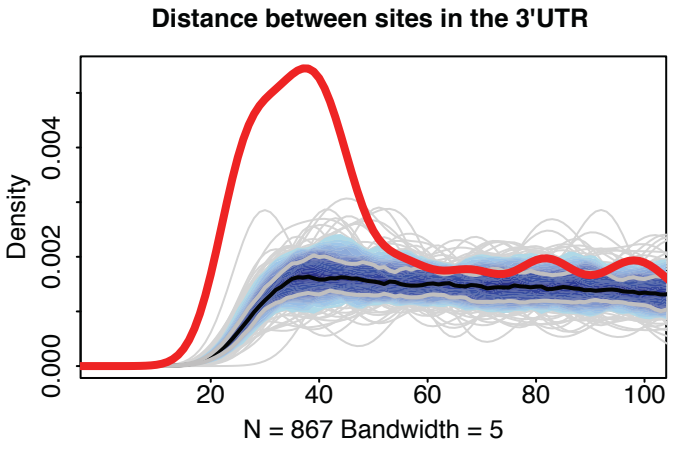
B

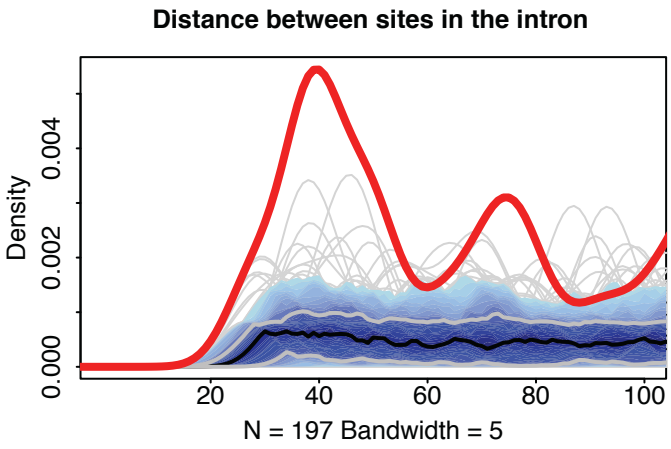

Clusters present in both libraries
Clusters present only in one library
Cumulative fraction
Conversion specificity
1.00
0.75
0.50
0.25
0.00
-0.5
0
0.5
1
1.5
2

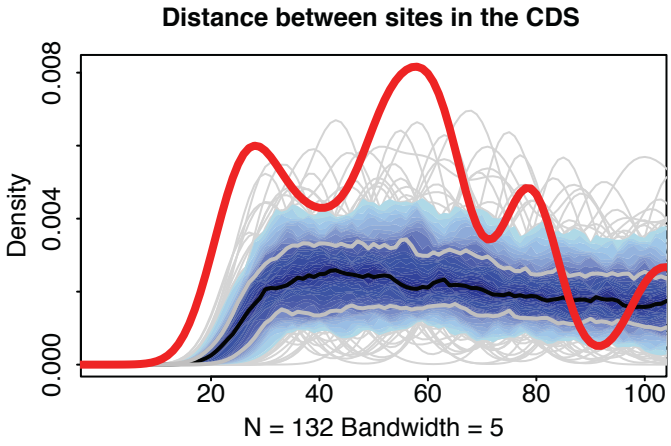
C

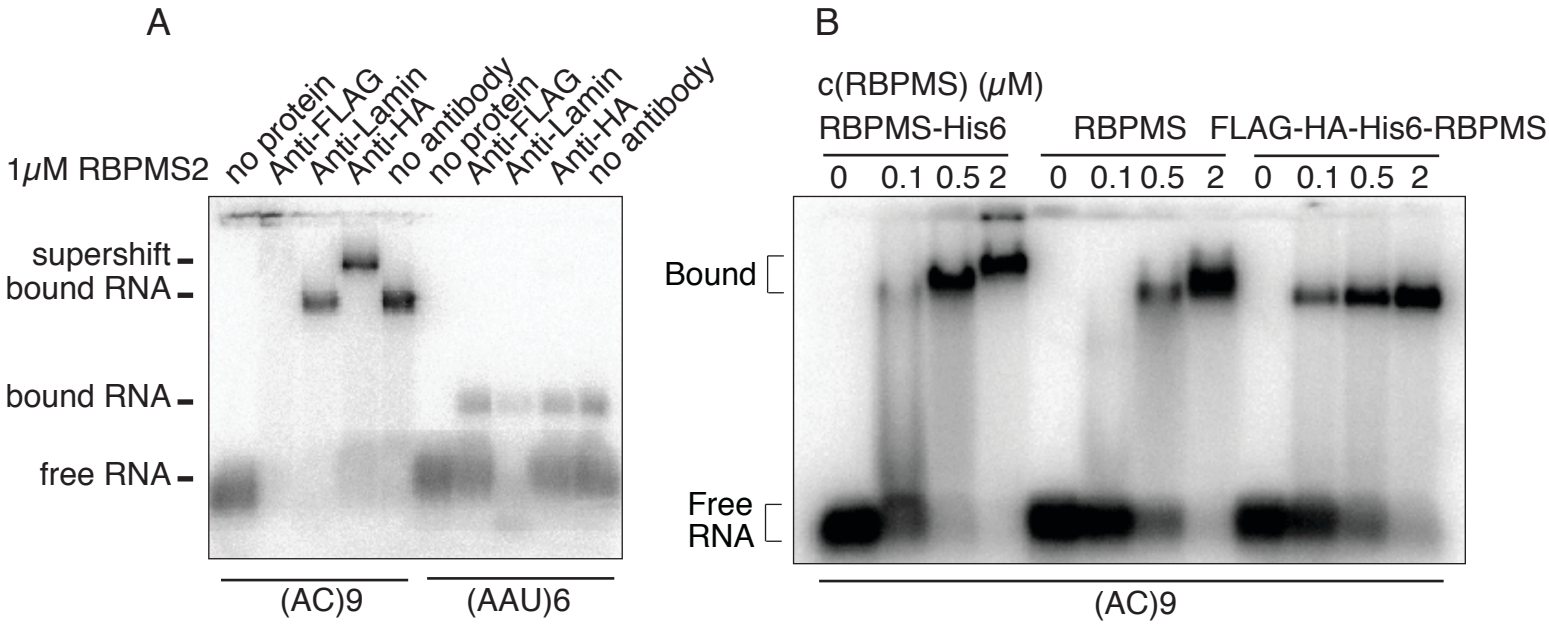

Count
200.0
125.0
50.0
25.0
0.0
RBPMS binding sites
r=0.71
Log2 reads per cluster library A
Log2 reads per cluster library B
16.0
15.0
14.0
13.0
12.0
11.0
10.0
9.0
8.0
7.0
6.0
5.0
4.0
3.0
2.0
2 3 4 5 6 7 8 9 10 11 12 13 14 15 16

Farazi_Supplementary Figure 4

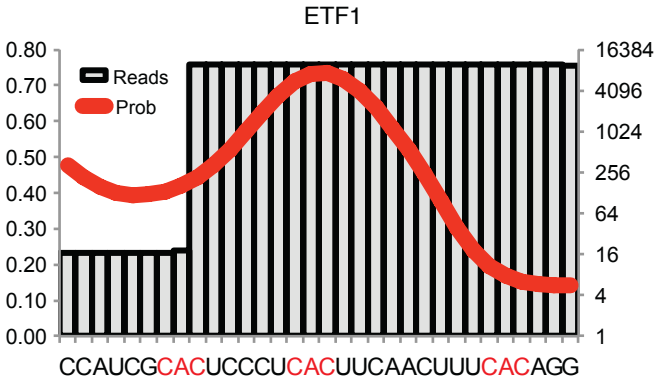


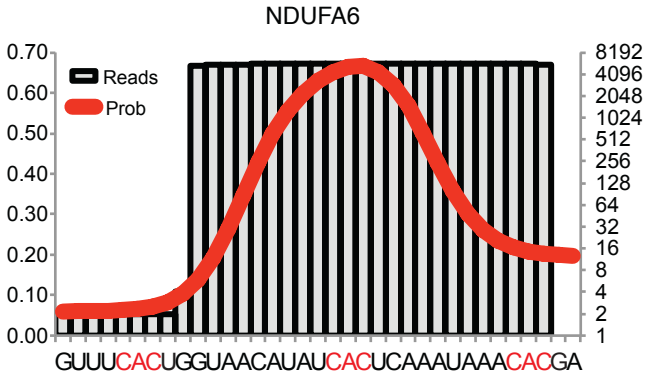


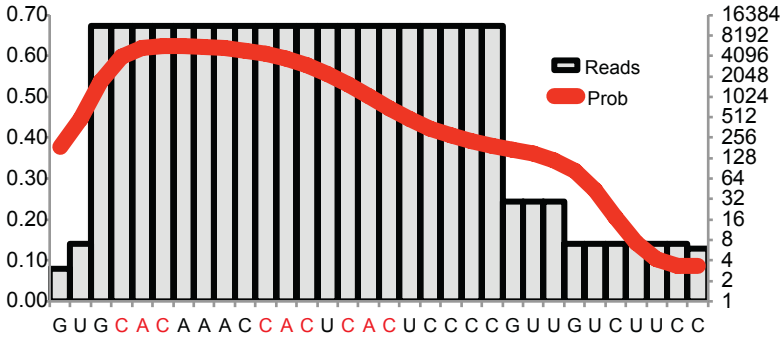


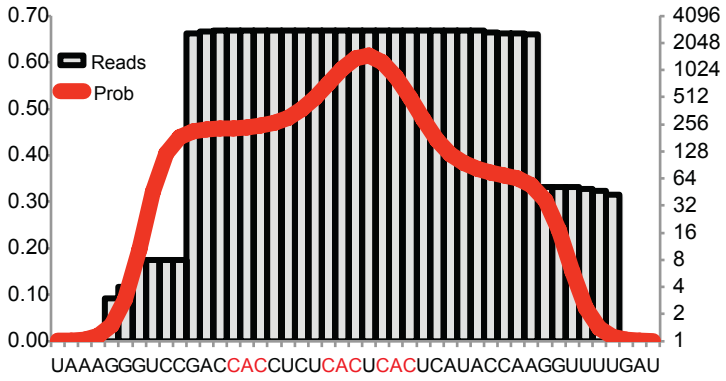

A
1μM RBPMS2
no protein
Anti-FLAG
Anti-Lamin
Anti-HA
no antibody
no protein
Anti-FLAG
Anti-Lamin
Anti-HA
no antibody
supershift
bound RNA
bound RNA
free RNA
(AC)9
(AAU)6
B
c(RBPMS) (μM)
RBPMS-His6
RBPMS
FLAG-HA-His6-RBPMS
0 0.1 0.5 2
0 0.1 0.5 2
0 0.1 0.5 2
Bound
Free RNA
(AC)9

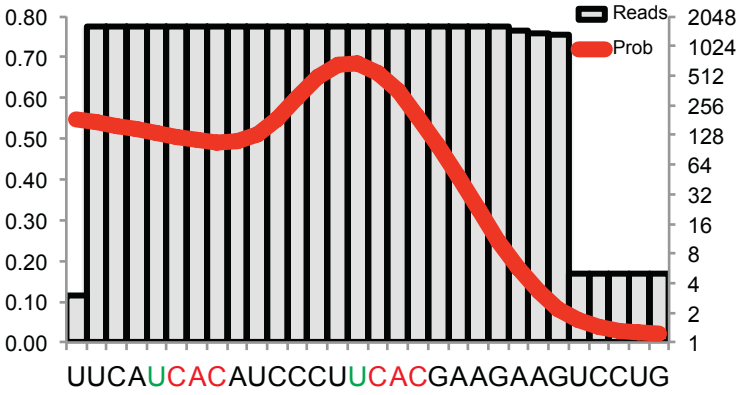
Farazi_Supplementary Figure 6

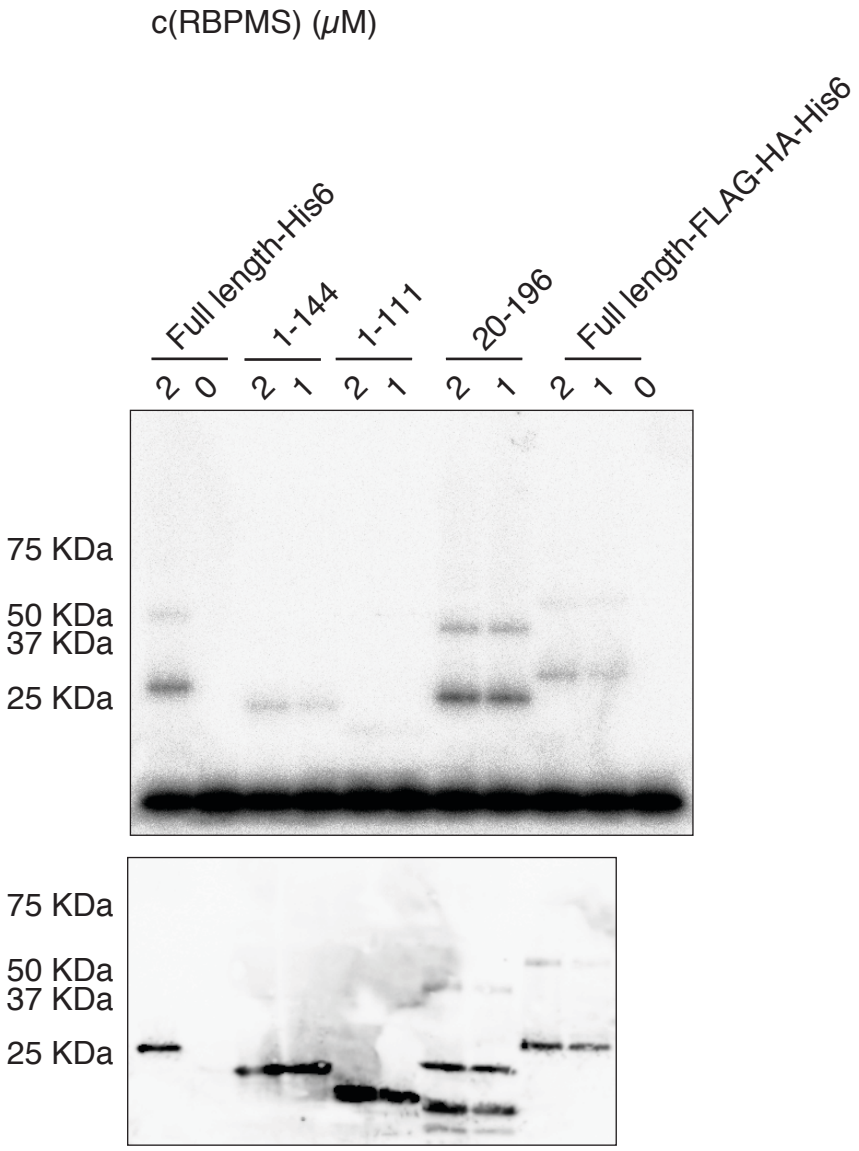






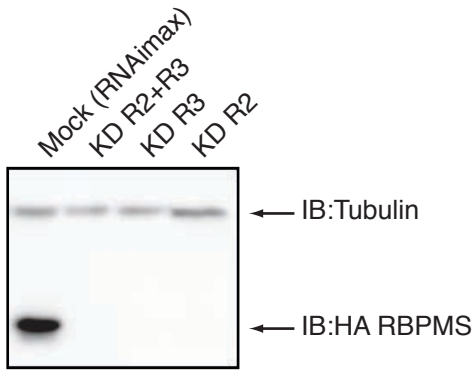


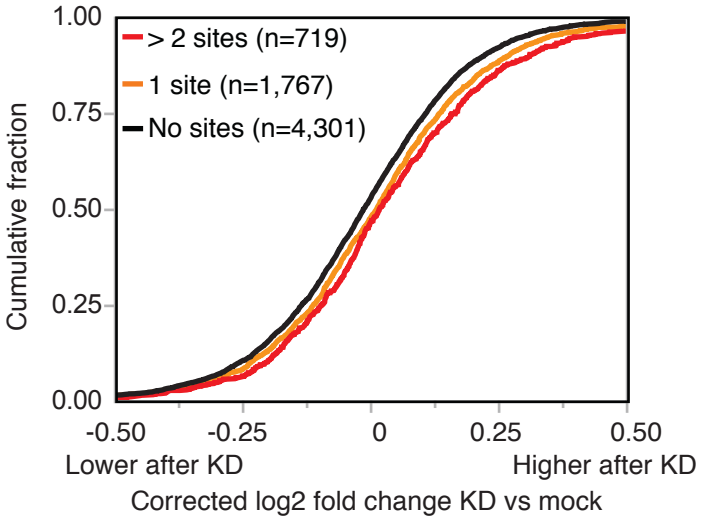



c(RBPMS) (μM)
Full length-His6
1-144
1-111
20-196
Full length-FLAG-HA-His6
2 0 2 1 2 1 2 1 2 1 0
75 KDa
50 KDa
37 KDa
25 KDa
75 KDa
50 KDa
37 KDa
25 KDa

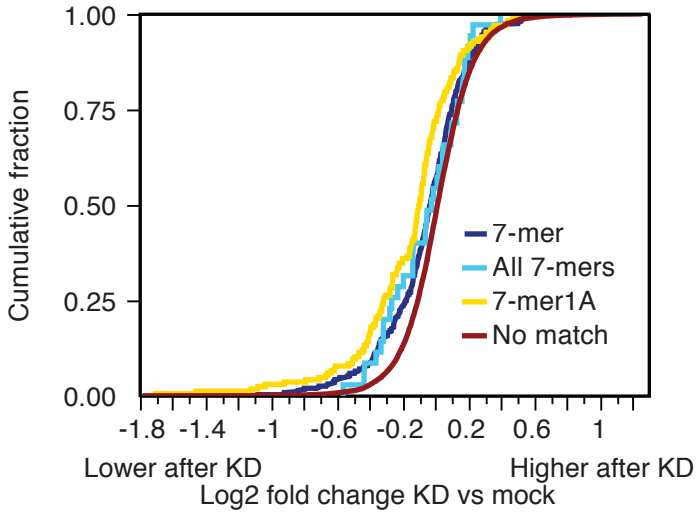
Farazi_Supplementary Figure 8

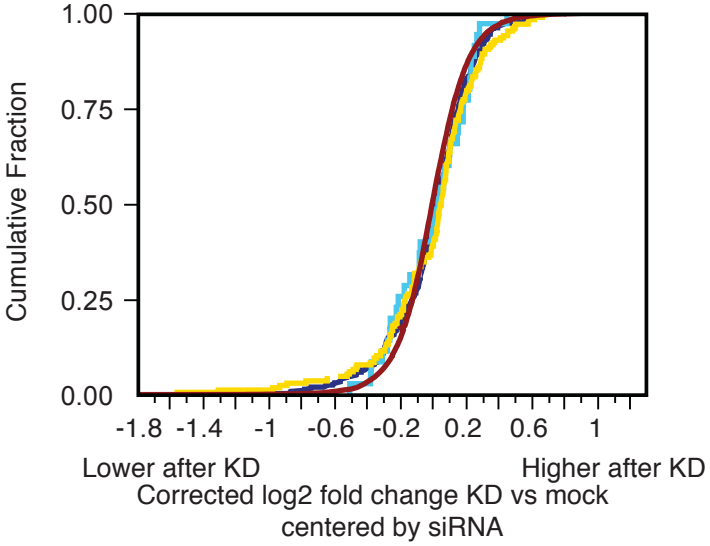
A



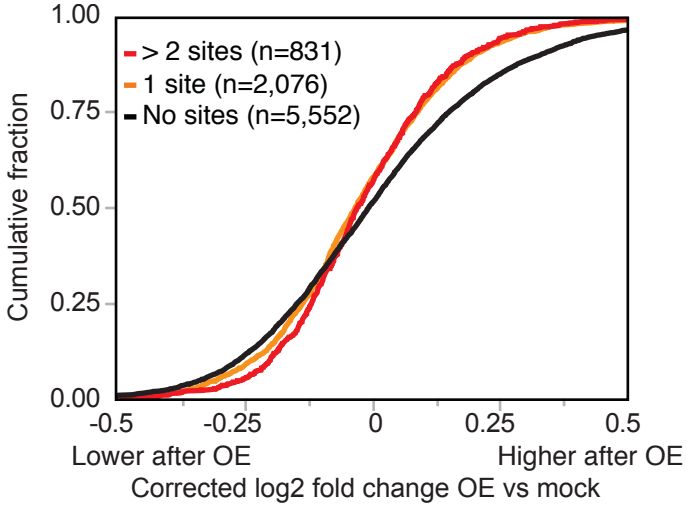


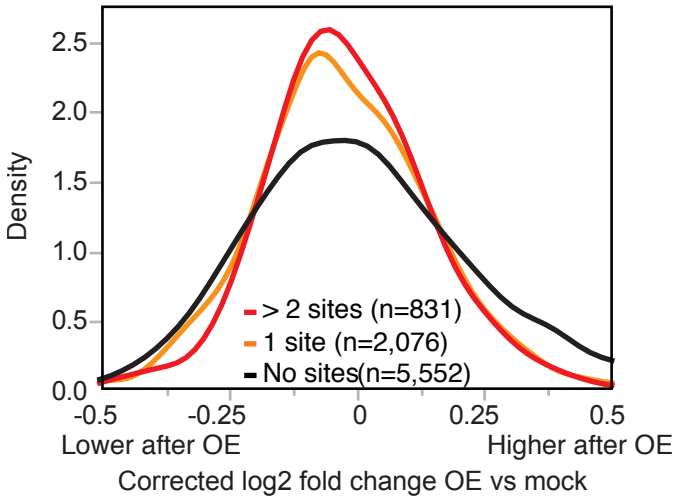
B

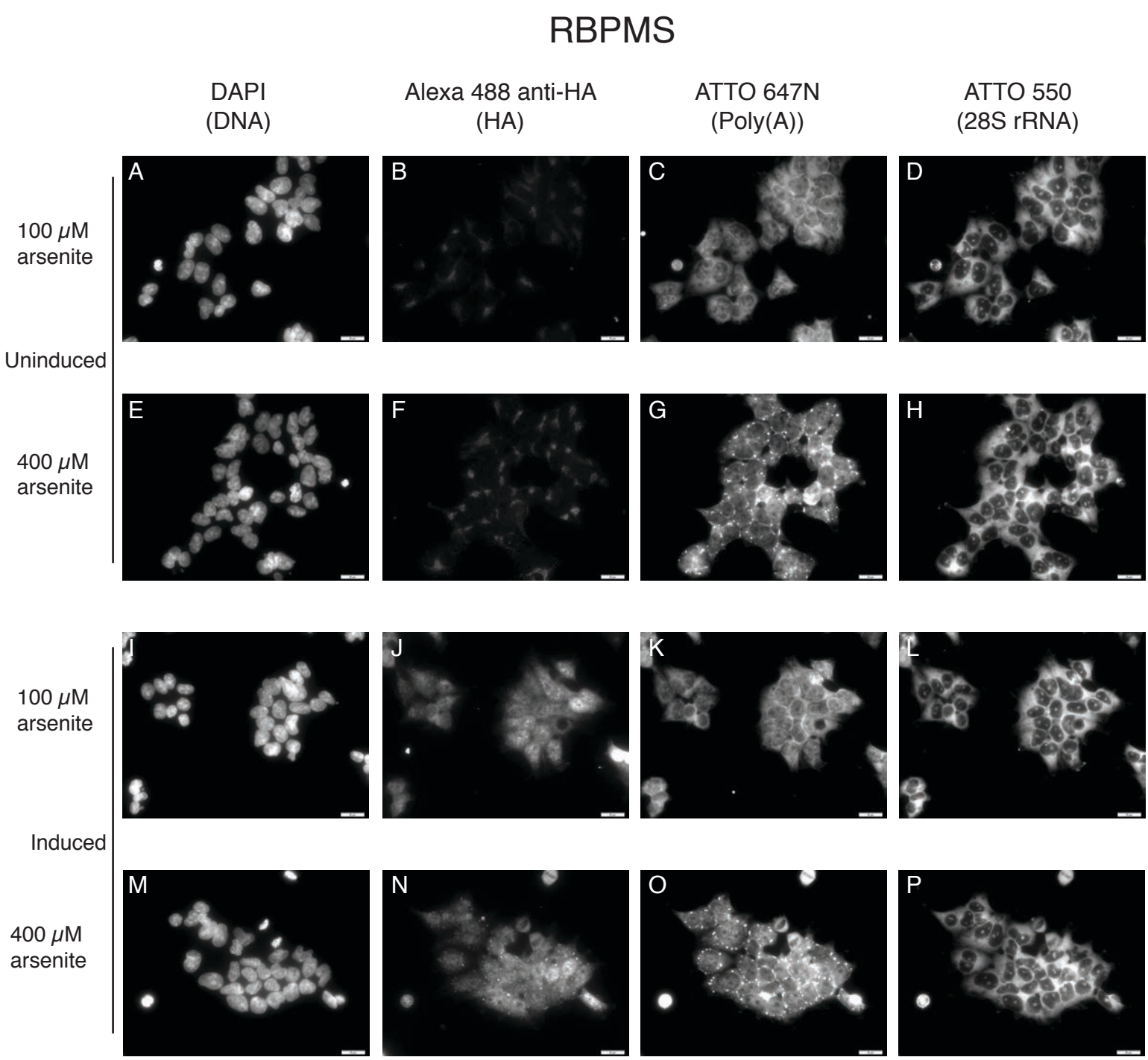




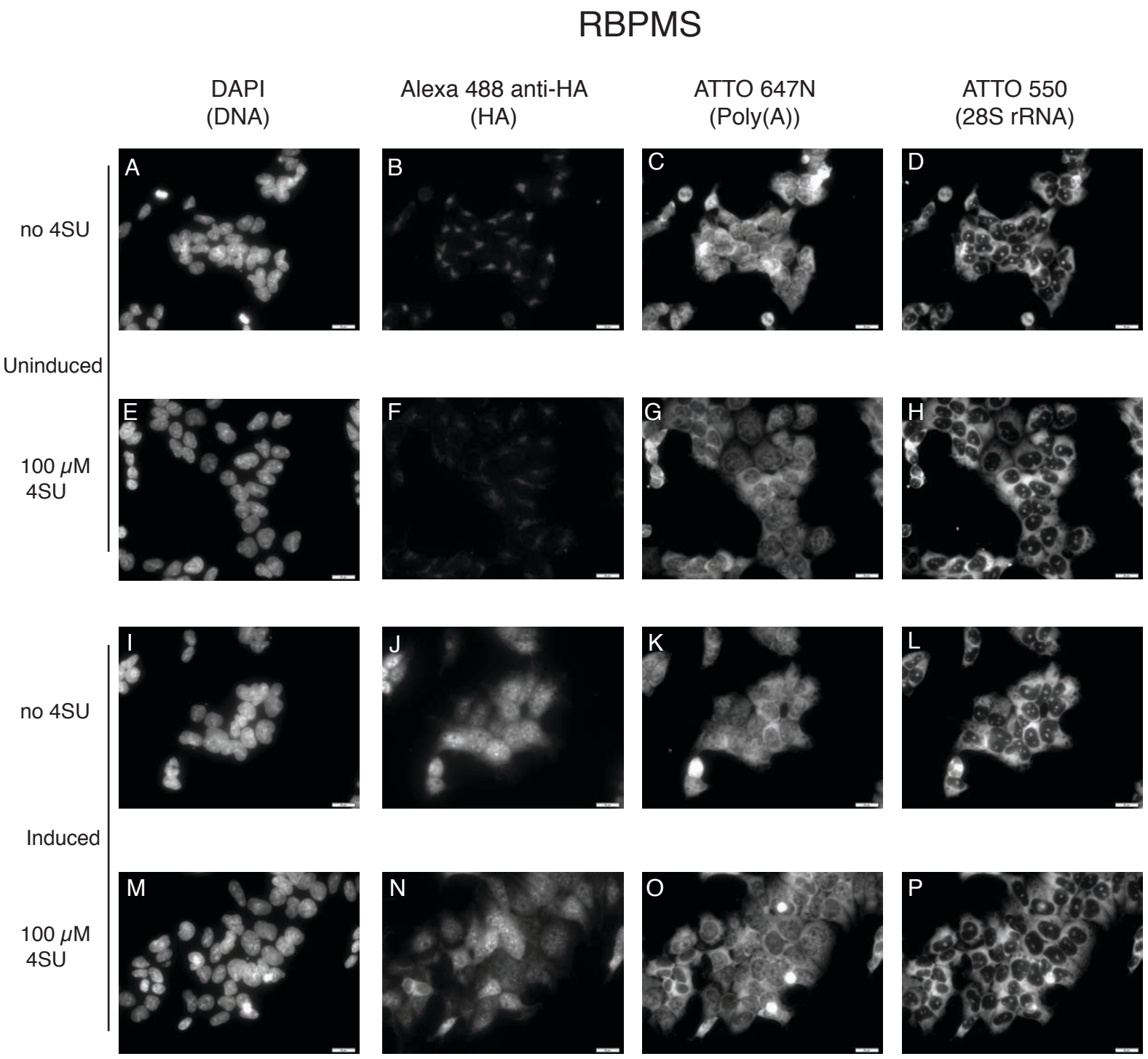
C



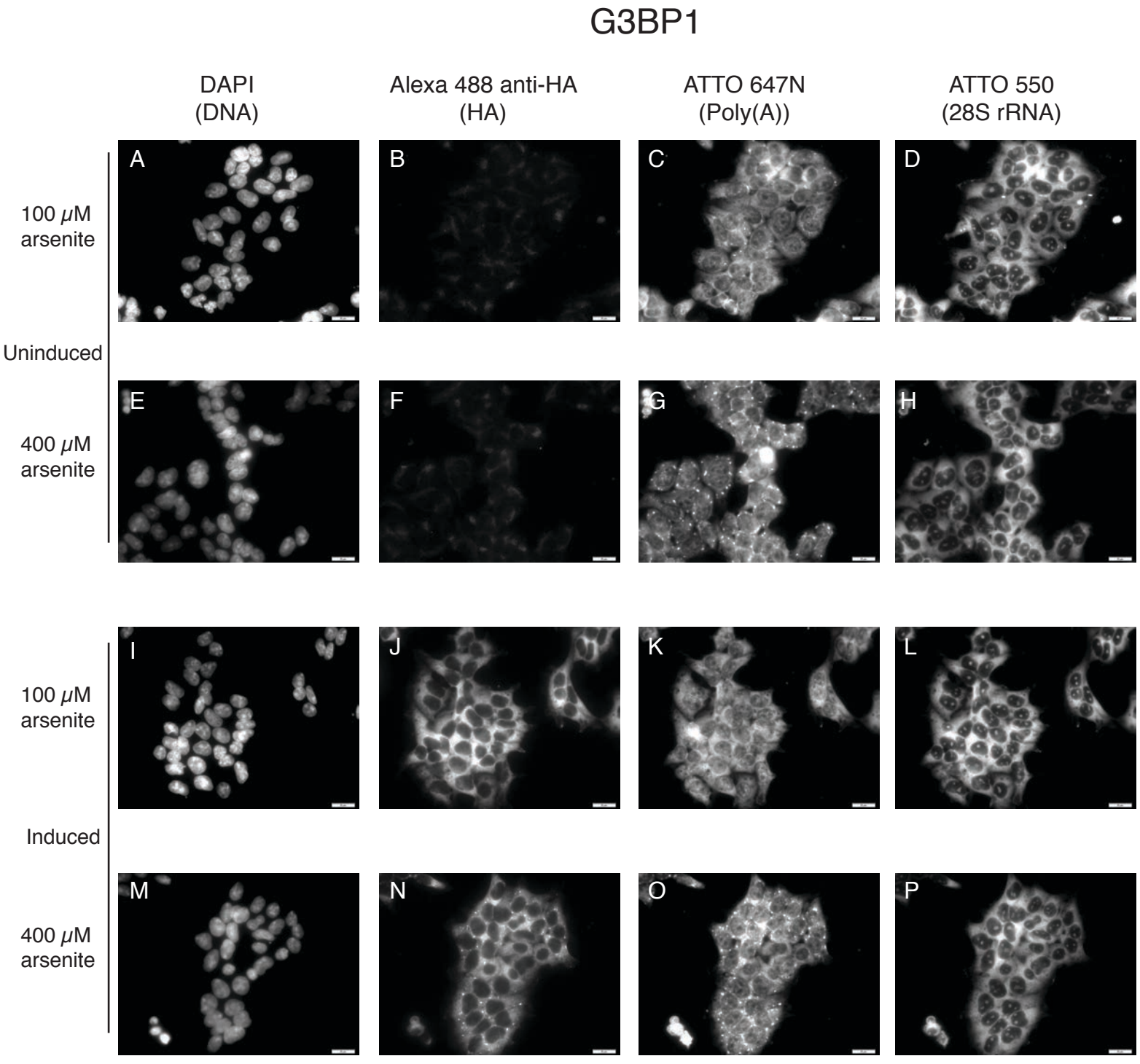


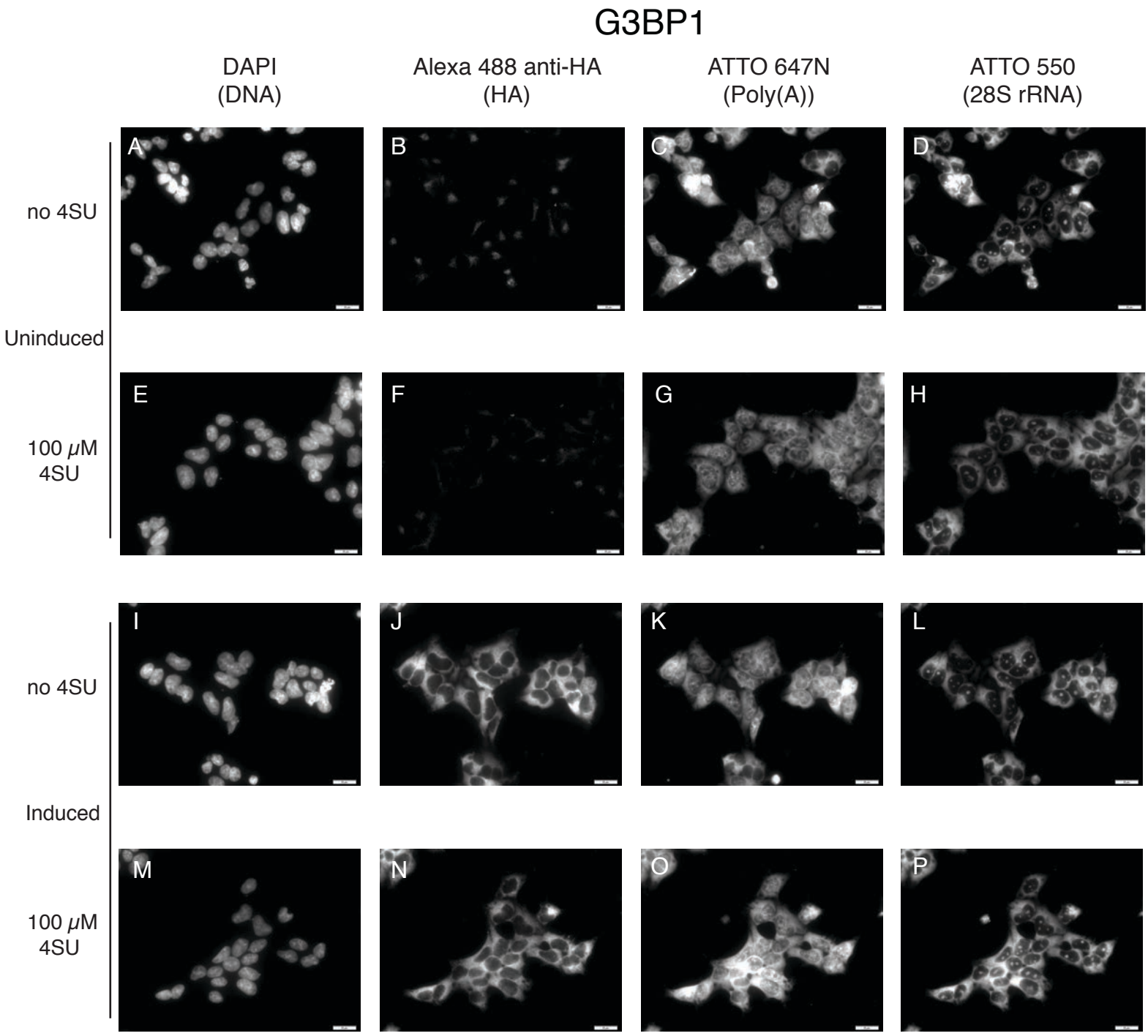

RBPMS
DAPI
(DNA)
Alexa 488 anti-HA
(HA)
ATTO 647N
(Poly(A))
ATTO 550
(28S rRNA)
A
B
C
D
100 μM
arsenite
Uninduced
E
F
G
H
400 μM
arsenite
I
J
K
L
100 μM
arsenite
Induced
M
N
O
P
400 μM
arsenite

Farazi_Supplementary Figure 9B



Farazi_Supplementary Figure 9C




G3BP1
DAPI
(DNA)
Alexa 488 anti-HA
(HA)
ATTO 647N
(Poly(A))
ATTO 550
(28S rRNA)
no 4SU
100 μM
4SU
Uninduced
Induced
A
B
C
D
E
F
G
H
I
J
K
L
M
N
O
P