# Supplementary Data

# for:

# A novel RNA motif mediates the strict nuclear localization of a long non-coding RNA

Bing Zhang[1], Lalith Gunawardane[1], Farshad Niazi[1,2], Fereshteh Jahanbani[1,3] , Xin Chen[1] and Saba Valadkhan[1]*

[1] Department of Biochemistry, Case Western Reserve University School of Medicine, Cleveland, OH 44106

[2] Present address: Genomic Medicine Institute, Lerner Research Institute, Cleveland Clinic, Cleveland, Ohio 44195, USA

[3] Present address: Center for Genomics and Personalized Medicine, Department of Genetics, Stanford University, Palo Alto, CA 94304

* Corresponding Author.
Email: saba.valadkhan@case.edu
Phone: +1 (216) 368 1068

# Supplementary Figure Legends

**Fig. S1. Sequence of the mouse BORG RNA.** The numbers refer to the mature, spliced BORG transcript. The darker gray highlights mark intron-derived sequences, which are separated from the exons by "/" marks. The size of each intron is shown within the highlighted region, and the sequence of the 5' and 3' splice site junctions, putative branch site and polypyrimidine tract are included. The polyadenylation signal hexamer and the cleavage site at the 3' end of BORG are shown in boldface letters. A forward slash marks the site of cleavage. The region downstream of the cleavage site highlighted in lighter gray contains the putative downstream poly(A) processing elements (the GU-rich region)(1).

**Fig. S2. Characterization of the mouse BORG RNA.** a. BORG is a strictly nuclear transcript. RT-qPCR reactions on cytoplasmic (Cyt) and nuclear (Nuc) cellular fractions obtained from C2C12 cells. Both the RT and PCR reactions have been performed using BORG-specific primers shown in Fig. 1a that flank an exon-exon junction. The lanes marked "- RT" are control reactions in which the reverse transcriptase has been omitted. Error bars represent the standard deviation calculated based on at least three independent experiments. b. BORG is a spliced transcript originating from the reference minus (antisense) genomic strand. To verify the existence of BORG RNA and determine its direction of transcription, we performed strand-specific RT-PCR. Reverse transcription reactions contained primers that bound either the sense or the antisense genomic strands in the BORG locus. Next, we used these reverse transcription reactions as templates in qPCR reactions. Reverse transcription reactions with the primers that bound the sense strand (labeled 5') do not yield a product when used as PCR templates, while those binding the antisense strand (labeled 3') did yield a product. These results indicate that BORG is transcribed in the antisense (minus) genomic reference orientation. The PCR primers flank either the first or second exon-exon junction in BORG, ensuring that the observed RT-PCR product does not originate from genomic DNA contamination. 5' and 3' indicate the forward and reverse primers compared to BORG (binding the sense and antisense genomic strands, respectively), with the numbers indicating the positions of the start and end of the resulting amplicon on BORG RNA. The results of RT-qPCR reactions with random hexamers used in the RT reaction are shown, which also yield a PCR product albeit with reduced efficiency compared to

BORG-specific primers. c. Validation of the sequence of the first and second splice junctions in mature BORG RNA. The sequences surrounding the first (top panel) and second (bottom panel) exon-exon junctions in the mature BORG RNA were amplified by RT-PCR using primers flanking the exon-exon junctions in total cellular RNA obtained from C2C12 cells. The amplicon was purified and sequenced using a reverse primer to verify the sequence of the splice junctions as shown in the sequencing chromograms. Sequencing of RT-PCR products obtained from mouse brain tissue RNA yielded similar results. d. BORG RNA is polyadenylated at its 3' end. The reaction ingredients included or omitted in each reaction are shown on top. e. To confirm the polyadenylation of BORG, we performed PCR reactions using radiolabeled BORG-specific primers on two independently-generated cDNA libraries made with oligo(dT) primers. In these experiments, BORG-specific primers used in the PCR reaction are located very close to the 3' end of BORG to ensure that the observed oligo(dT)-mediated priming does not result from the interaction of oligo(dT) primers with an A-rich stretch of BORG RNA. We verified that there are no A-rich stretches between the amplicon and the 3' cleavage site determined by 3'RACE analysis in BORG. As the BORG-specific PCR primers used in this study do not flank an exon-exon junction, we included a mock reaction that lacked reverse transcriptase (lane marked No RT-BORG PCR) to check for the presence of genomic DNA contamination, which proved to be negligible. f. To further prove the specificity of these amplifications, we performed radioactive PCR reactions using BORG-specific primers on cDNAs made with primers specific to Stat3, IKK and c-Met genes, which did not lead to a detectable product. The specificity of the primer used for the RT and those used in the PCR reaction are shown above each lane. A positive control reaction with c-Met-specific PCR primers is also shown. g. and h. Analysis of the level of U6 and U99 (nuclear markers) and MT-RNR1 and 18S ribosomal RNA (cytoplasmic markers) in the nuclear and cytoplasmic fractions (Nuc and Cyt, respectively) using both RT-PCR with radiolabeled primers (g) and quantitative RT-PCR (h). The two sets of nuclear and cytoplasmic markers yield virtually identical results.

**Fig. S3. Characterization of subcellular localization of BORG.** a. RT-PCR reactions with radiolabeled primers indicate the nuclear localization of BORG in 4T1 mouse breast cancer cells. b. Validation of fractionation purity by RT-qPCR assays on nuclear and cytoplasmic fractions from 4T1 and

BORG-overexpressing NMuMG cells using U6 and MT-RNR1 as nuclear and cytoplasmic markers, respectively. Cyt and Nuc refer to cytoplasmic and nuclear fractions. Identity of the fractionated cell line is shown at the top. c. RT-qPCR reactions on nuclear and cytoplasmic fractions from HEK293 cells indicate the nuclear localization of BORG in these cells. d. RT-qPCR assays on nuclear and cytoplasmic fractions (Nuc and Cyt, respectively) obtained from HEK293 cells to determine the purity of fractions, using U6 and MT-RNR1 as nuclear and cytoplasmic markers. e. RT-qPCR reactions on nuclear and cytoplasmic fractions from HeLa cells indicate the nuclear localization of BORG in these cells. U6 and MT-RNR1 are used as nuclear and cytoplasmic markers to indicate the purity of the fractions. f. RT-qPCR assays to monitor the purity of the nuclear and cytoplasmic fractions from BMP2-treated cells using U99 and MT-RNR1 as nuclear and cytoplasmic markers, respectively. g. RT-qPCR reactions on nuclear and cytoplasmic fractions (Nuc and Cyt, respectively) obtained from BORG-overexpressing clones 1 and 2 to show the purity of the fractions using U99 and MT-RNR1. h and i. Analysis of the level of BORG in vector-transfected (h) and BORG overexpressing (i) C2C12 cells after the addition of DRB to block Pol II transcription. Dotted red lines mark the value on the Y axis corresponding to 50% of the starting level of BORG.

**Fig. S4. The localization of BORG is not altered as a result of blocking nonsense-mediated decay pathways.** a. RT-qPCR assays using BORG-specific primers on cellular fractions from C2C12 untransfected cells and those overexpressing BORG from a low expression level transgene (+BORG) after an hour of heat shock and the indicated time point of post-heat-shock recovery. b. Similar assays as in panel a. on fractions from cells treated with cycloheximide for the indicated time points. c. RT-qPCR probing of fractionated RNAs from the early time points of cycloheximide treatment indicate a rise in the level of both total and cytoplasmic lncRNA GAS5 transcript, which is known to be an NMD target (2). The tested time points match those in Fig. 2e. d and e. Validation of fractionation purity for the heat shock- and cycloheximide-treated samples using U99 and MT-RNR1 as nuclear and cytoplasmic markers, respectively. f. RT-qPCR reactions on nuclear and cytoplasmic fractions from cells in which BORG has been induced by BMP2 treatment, followed by cycloheximide addition. Nuc and Cyt refer to nuclear and cytoplasmic fractions, respectively. For details, see Methods.

**Fig. S5. Analysis of the subcellular localization of Mut and Exon transcripts.** a and b. RT-qPCR assays indicate that all BORG truncation mutant transcripts are strictly localized to the nuclei. Nuc and Cyt refer to nuclear and cytoplasmic fractions, respectively. The PCR reactions were performed with primers that flanked the region deleted in each construct (Table S3) to enable us to distinguish the transgene-expressed transcript from the endogenous BORG RNA (shown) based on size (see panel d below). c. Validation of fractionation purity for the Mut1-6 constructs using RT-qPCR assays on fractionated RNAs against U6 and MT-RNR1 as nuclear and cytoplasmic markers, respectively. d. The expression of the mutant constructs do not alter the localization of the endogenous RNA. The RT-PCR assays shown are performed with radiolabeled primers that flank the region deleted in mutants 3 and 4. Due to the small size of the deleted region, an amplicon corresponding to the endogenous transcript could also be detected in the same PCR assay. Cyt and Nuc refer to the cytoplasmic and nuclear fractions, respectively.

**Fig. S6. Validation of the expression level and purity of fractionation for the Frag constructs and Ex2- and Ex3-mutant constructs.** a. Quantitative RT-PCR to define the level of nuclear and cytoplasmic-localized Frag3 and Frag4 RNAs. Nuc and Cyt refer to nuclear and cytoplasmic fractions, respectively. b. RT-qPCR assays for defining the purity of the nuclear and cytoplasmic fractions using U99 and 18S rRNAs as nuclear and cytoplasmic markers, respectively. c. A transcript containing sequences that are reverse and complement of those in Frag5 construct is dominantly cytoplasmic as determined by RT-qPCR. d. Quantitative RT-PCR assays indicate the purity of the nuclear and cytoplasmic fractions used in panel c. U99 and 18S rRNA are used an nuclear and cytoplasmic markers, respectively. e. Quantitative RT-PCR to define the level of nuclear and cytoplasmic-localized Ex2-Mut3 and Ex3-Mut7 RNAs. Nuc and Cyt refer to nuclear and cytoplasmic fractions, respectively. f. and g. Validation of the fractionation purity of RNAs used in Fig. 5. Nuc and Cyt refer to nuclear and cytoplasmic fractions, respectively. Either U6 (f) or U99 (g) and MT-RNR1 are used as nuclear and cytoplasmic markers, respectively.

**Fig. S7. Validation of purity of nuclear and cytoplasmic fractions from AGCCC insertion and mutation experiments**. a. U99 and MT-RNR1 are used as nuclear and cytoplasmic markers for defining the purity of the fractions obtained from cells transfected with Frag2 mutants. The identity of each mutant (corresponding to those in Fig. 6b) is shown at the bottom. b. A similar analysis as in panel a. for fractions

obtained from AGCCC insertion mutants analyzed in Fig. 6c.

**Fig. S8. Analysis of the specificity of the correlation between AGCCC copy number and nuclear localization of RNAs.** a. Distribution of the number of -8(A/T),-3(G/C),AGCCC motifs per transcript in lncRNAs, antisense RNAs and mRNAs annotated in the Ensembl database. In the box graph in this and the following two panels, the dotted line marks the median, and the box marks the 25-75$^{th}$ percentile distribution. The whiskers marks the 0 (when applicable) and 100 percentile assuming a normal distribution. b. and c. Distribution of length of mature RNAs (b) and number of introns (c) in the three groups of transcripts. d. and e. Distribution of the number of -8(A/T),-3(G/C),AGCCC motifs per transcript divided by the length of the transcript (in d) and number of exons per transcript (in e). f. and g. Correlation between the exon number (f) and transcript length (g) with nuclear/cytoplasmic ratio of RT-qPCR signals for protein-coding RNAs. In these and the following graphs, the size of the circles representing each RNA is correlating with the size of the RNA (see Table S2). Error bars represent two standard deviations. The Spearman's correlation coefficient (r) and two-tailed p-values (p) are shown. h. and i. Correlation of the nuclear/cytoplasmic ratio of the RT-qPCR signal for protein-coding RNAs with the number of AGCCC motifs per exon (h) and the number of AGCCC motifs divided by the length of the RNA (i). Neither improve the correlation with nuclear/cytoplasmic ratio compared to the absolute number of motifs per RNA (Fig. 6d). -8,-3,AGCCC refers to the motif containing both the AGCCC pentamer and the sequence restrictions at positions -3 and -8. j. In the absence of the sequence restrictions at positions -3 and -8, the AGCCC motif shows a weaker correlation with the nuclear localization of RNAs as evidenced by the values of Spearman's correlation coefficient (r) and two-tailed p-values (p).

## Supplementary References


1. **Darnell JE Jr**. 2013. Reflections on the history of pre-mRNA processing and highlights of current knowledge: a unified picture. RNA **19**:443–460.

2. **Tani H, Torimura M, Akimitsu N**. 2013. The RNA degradation pathway regulates the function of GAS5 a non-coding RNA in mammalian cells. PLoS ONE **8**:e55684.

```
   1 cacccagaag tgtgctattc taccagctat ttggtatctc tagggcaaca aacttgacaa
  61 tcaagatcag ttaccacatg cctctctcct cgataaagag caagctgcga agggcgatta
 121 cagctggaca caggtgagtg agccagaagg aactgaccag ccttaggaag gaaccgtcac
 181 agctaggcca tgctccagat attaatttct gagtcactgg ccaggagcga gatcccagtg
 241 taagagtccc ctgtgatgta acgtctggaa gaattccaca gagaaagaag tcagaaactt
 301 ctacacagct gacagatcaa gatcaccagg gctgggaagt tttctcagga gtaaagtgtg
 361 aggagcggag tttaaatccc agcagccgtg taaaagctgc tgcaggagca tgtctgcaat
 421 ctcaggcctg ggggaagtga gcaaagacag cagctgctga gaacttactg ggcagccgtg
 481 gagacaaagt gccaagctcc agattccgtg ggagccccag gaaataaggt gaaag/gtaa
     gggtggagca -9318 nts- gccttgaatc acatcttctc ctatttgcct tcag/aacaa
 541 gatgtagaaa gaactctcag ctccctcagc atcctgtctg cctgccttca agctttcgtg
 601 cttcccacca tgatgacaac ggactgaacc tcagaatcta caagtcagcc ccaattaaat
 661 gttgttttag tcatggtgtc ttttcacagc aatgaaaatt ctaagataga agttactaac
 721 aaggactagg gtattgctgt gataggcctg accatgcttt tgtggatttt gtaactttgg
 781 atttggaaag cagtggaatg ctttaagtgg gacttaatgg gacagcctag taggaatagg
 841 gaggacattg gtgctaaggg tgaactctgc gggtctggct caagagattt cagaggagaa
 901 taacttgagt atgtggcctg gagactgctt ttgtgatact ttggtgagga atgtaggtac
 961 tttttgccct cgtttgaagt gtctgcctga ggctcaagtg aagagattca aattaattgc
1021 aatgacagaa gtctcagaaa ag/gtaagtg tactttgtcc tctggtggat tcttatgagg
     aacattttga tcaactgcag -2794 nts- catagatctc ttttgactac ttcctgctca
     ctaagggcct tgaattccta g/gggcaagt ccaatcttca tcatcaggat atcttcaatc
1081 agcaaccagc aaaaaacagc aaaacagcaa atgcaacagc aggaaccagc aacagcagga
1141 agagcagcag ccactacctc ctctggtgtt tatatactgc tctggtattt atattctctc
1201 tcaagaattc ctagaaatct actgtcttca gctggcaaat cacatccctg ccagagcacg
1261 agacaaatca tagtcagctg ctgtggacaa tctgaagcag acccacatcc cacccctgga
1321 attaaaacaa aaacacattc atagaacata actaggggaa aaaaaaccaa tacagtgagc
1381 tttgggatgg ccagctttct tacacggagg ctcagagctc cacagtaaga gagggtggtg
1441 ggagtggtga caggcagaaa gactgtcatc tttcatgact gcctgcataa gaagctaagc
1501 cccagggata gaggctcagc ctctttatga aagaatagca tgtctctaaa agaacatgtt
1561 tagagggaaa gactattaga gccatatttt ggaaaatata atgagacaca atatttttt
1621 gtggatacag ttttataata agggtggtag caggaaatgt ctcctgacca ccacctgcta
1681 tcctgatgag ggaaccagct gctatttgga tgcctgtggt aaataccacc cactttgatg
1741 ccatcactct gatagcccag actcacttac aacccccttca ctggggaaag agggattttt
1801 ggtattacag ctgccatcat tgcagccatt attgcaggca ttacaggggc aacaacagct
1861 actgttgccc agaccgccac tatgagaggc tgtcaacgct gtagtctcta agtcagctga
1921 ggtgctacag gcataggagc tactaaatca acatctttat gaggccatcc acattttaca
1981 acaacagatt gactggcaga agagctggct cttgttaggg acatgtgtct caatagctta
2041 aagataccta gtcagacaaa ttcctattcc tgaaggcaag aaaaccactc acttgcccaa
2101 ttggtggtag atataaccac gtccataacc aactttggaa gcgggaacca gggctcctat
2161 gacagaaaag aagcaggctt ctctctgcaa ttcataaaca ccgacaggaa tatgttgctt
2221 ttcctaagcc cttcgctgtc tttccaagaa gaatttttt gtccttaacc aaatgtgaaa
2281 tgcacatttt gagcacaaag aatcaaagct ccaacagcaa actgacatga catcctttac
2341 aggaagtaaa atatttaaag acagagtcag agtcttctat tcccctcccc ccaggaagaa
2401 acccctttaat attcccatta accttgatgt gaatcacagc gctccttctt ccctgtaggc
2461 agataatcat gactcaccac agacccctga ctgtgggctt catgtaattc tggcagcaaa
2521 ggccttcccc gggggagctg tgtgaggcca ttggctgtga aatctttaca aagtccaggg
2581 atggactagg agtaaccctg gagcaagctt tgcactgtgt atcaaaacaa gactccacat
2641 tggggcatcc acactgaggc caccttgtga actaggcctc accttggata agaaagaaat
2701 tgcattattt ttactttcaa atgtgttaaa g**aataaa**ttt tacttaaata tattctatat
2761 tcaat**c/a**ga tatgcccact gctggagact gagtatttgt gttttacacc aaaatttctt
     ttgaaaccta agccctaagg agctagcatc aggaagtggg gttgttggaa gtgattaggt
```

**Fig. S1**

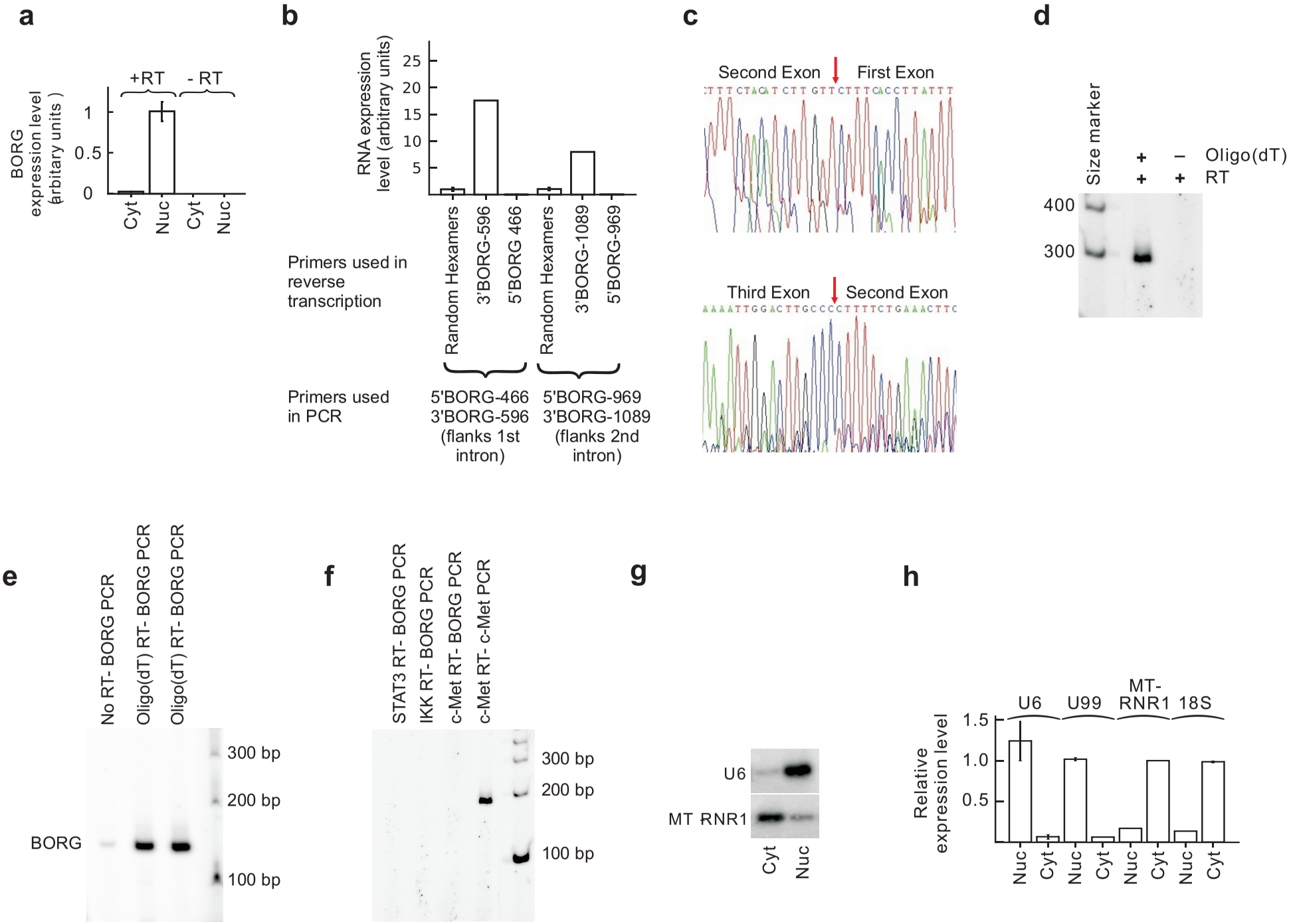

a
BORG expression level (arbitary units )
+RT
- RT
1
0.5
0
Cyt
Nuc
Cyt
Nuc
b
RNA expression level (arbitrary units)
25
20
15
10
5
Random Hexamers
3'BORG-596
5'BORG 466
Random Hexamers
3'BORG-1089
5'BORG-969
Primers used in reverse transcription
Primers used in PCR
5'BORG-466
3'BORG-596
(flanks 1st intron)
5'BORG-969
3'BORG-1089
(flanks 2nd intron)
c
Second Exon
First Exon
Third Exon
Second Exon
d
Size marker
+
+
–
+
Oligo(dT)
RT
400
300
e
No RT- BORG PCR
Oligo(dT) RT- BORG PCR
Oligo(dT) RT- BORG PCR
BORG
300 bp
200 bp
100 bp
f
STAT3 RT- BORG PCR
IKK RT- BORG PCR
c-Met RT- BORG PCR
c-Met RT- c-Met PCR
300 bp
200 bp
100 bp
g
U6
MT RNR1
Cyt
Nuc
h
U6
U99
MT-RNR1
18S
Relative expression level
1.5
1.0
0.5
Nuc
Cyt
Nuc
Cyt
Nuc
Cyt
Nuc
Cyt

**Figure S2**

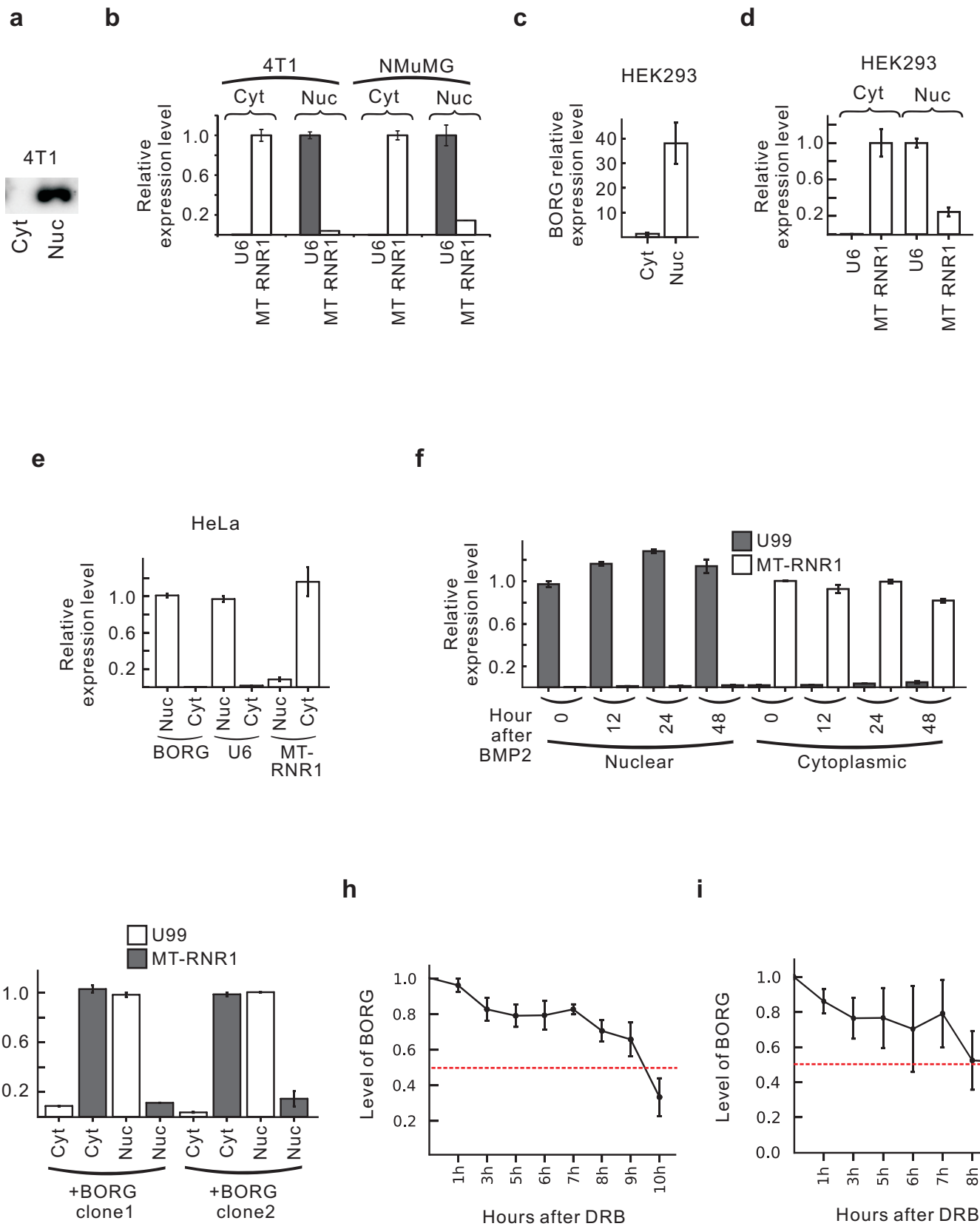

a
4T1
Cyt
Nuc
b
4T1
NMuMG
Cyt
Nuc
Cyt
Nuc
Relative expression level
1.0
0.6
0.2
U6
MT-RNR1
c
HEK293
BORG relative expression level
40
30
20
10
Cyt
Nuc
d
HEK293
Cyt
Nuc
Relative expression level
1.0
0.6
0.2
U6
MT-RNR1
e
HeLa
Relative expression level
1.0
0.6
0.2
Nuc
Cyt
BORG
U6
MT-RNR1
f
U99
MT-RNR1
Relative expression level
1.0
0.6
0.2
Hour after BMP2
0
12
24
48
Nuclear
Cytoplasmic
g
U99
MT-RNR1
Relative expression level
1.0
0.6
0.2
Cyt
Nuc
+BORG clone1
+BORG clone2
h
Level of BORG
1.0
0.8
0.6
0.4
0.2
1h
3h
5h
6h
7h
8h
9h
10h
Hours after DRB
i
Level of BORG
1.0
0.8
0.6
0.4
0.2
0.0
1h
3h
5h
6h
7h
8h
9h
10h
Hours after DRB

**Figure S3**

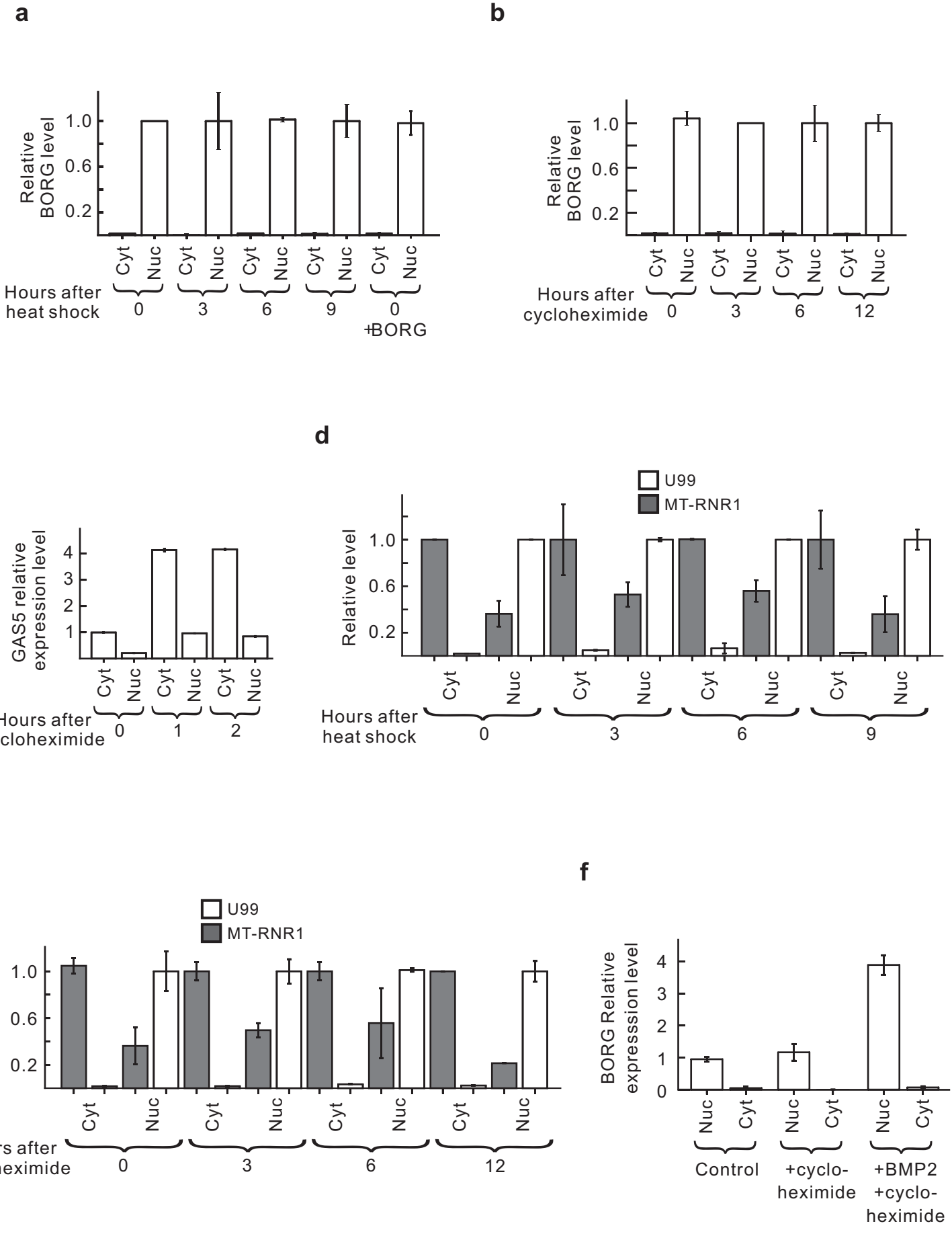
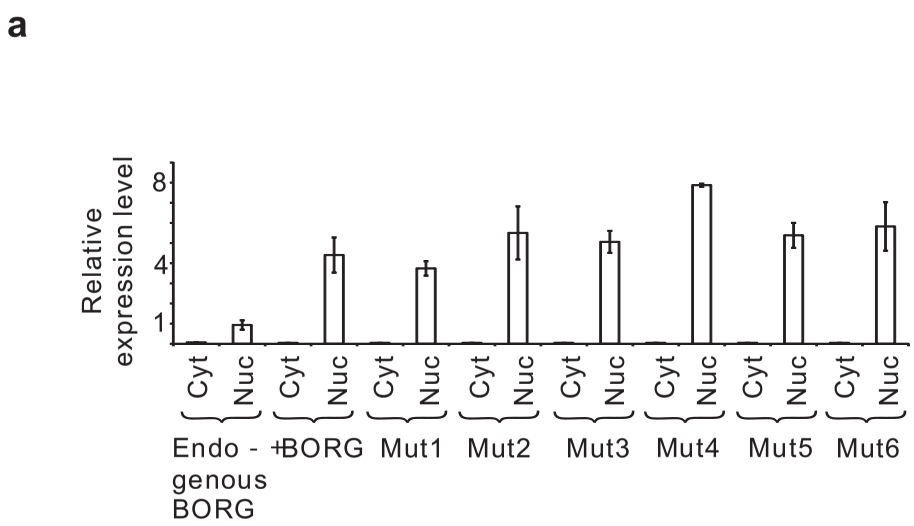
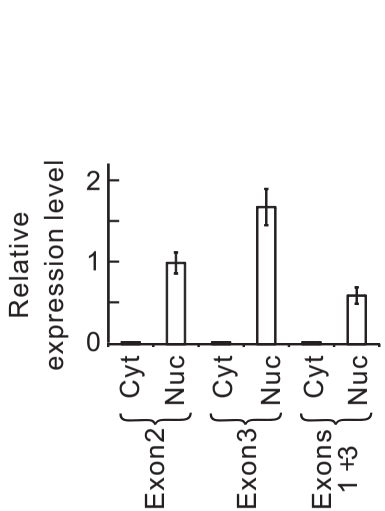
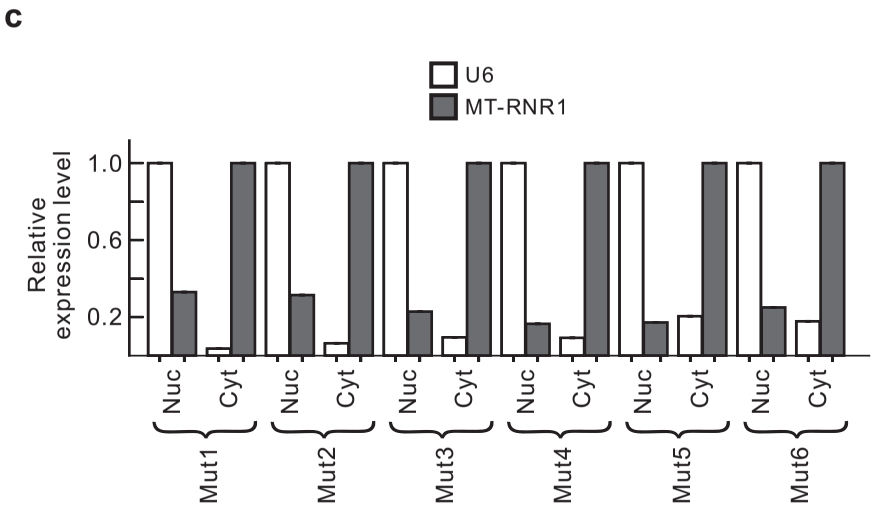

a
Relative BORG level
1.0
0.6
0.2
Cyt
Nuc
Hours after heat shock
0
3
6
9
0
+BORG
b
Relative BORG level
Hours after cycloheximide
0
3
6
12
c
GAS5 relative expression level
4
3
2
1
Hours after cycloheximide
0
1
2
d
U99
MT-RNR1
Relative level
Hours after heat shock
0
3
6
9
e
U99
MT-RNR1
Relative level
Hours after cycloheximide
0
3
6
12
f
BORG Relative expression level
Control
+cyclo-heximide
+BMP2 +cyclo-heximide

**Figure S4**


a
Relative expression level
8
4
1
Cyt
Nuc
Endo - genous BORG
+BORG
Mut1
Mut2
Mut3
Mut4
Mut5
Mut6


b
Relative expression level
2
1
0
Cyt
Nuc
Exon2
Exon3
Exons 1 +3


c
U6
MT-RNR1
Relative expression level
1.0
0.6
0.2
Nuc
Cyt
Mut1
Mut2
Mut3
Mut4
Mut5
Mut6

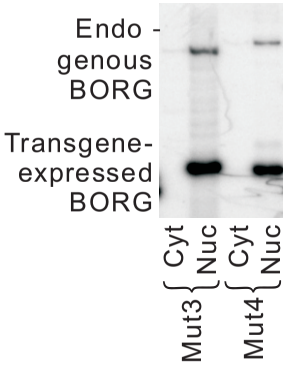

d
Endo - genous BORG
Transgene-expressed BORG
Cyt
Nuc
Mut3
Mut4

**Figure S5**

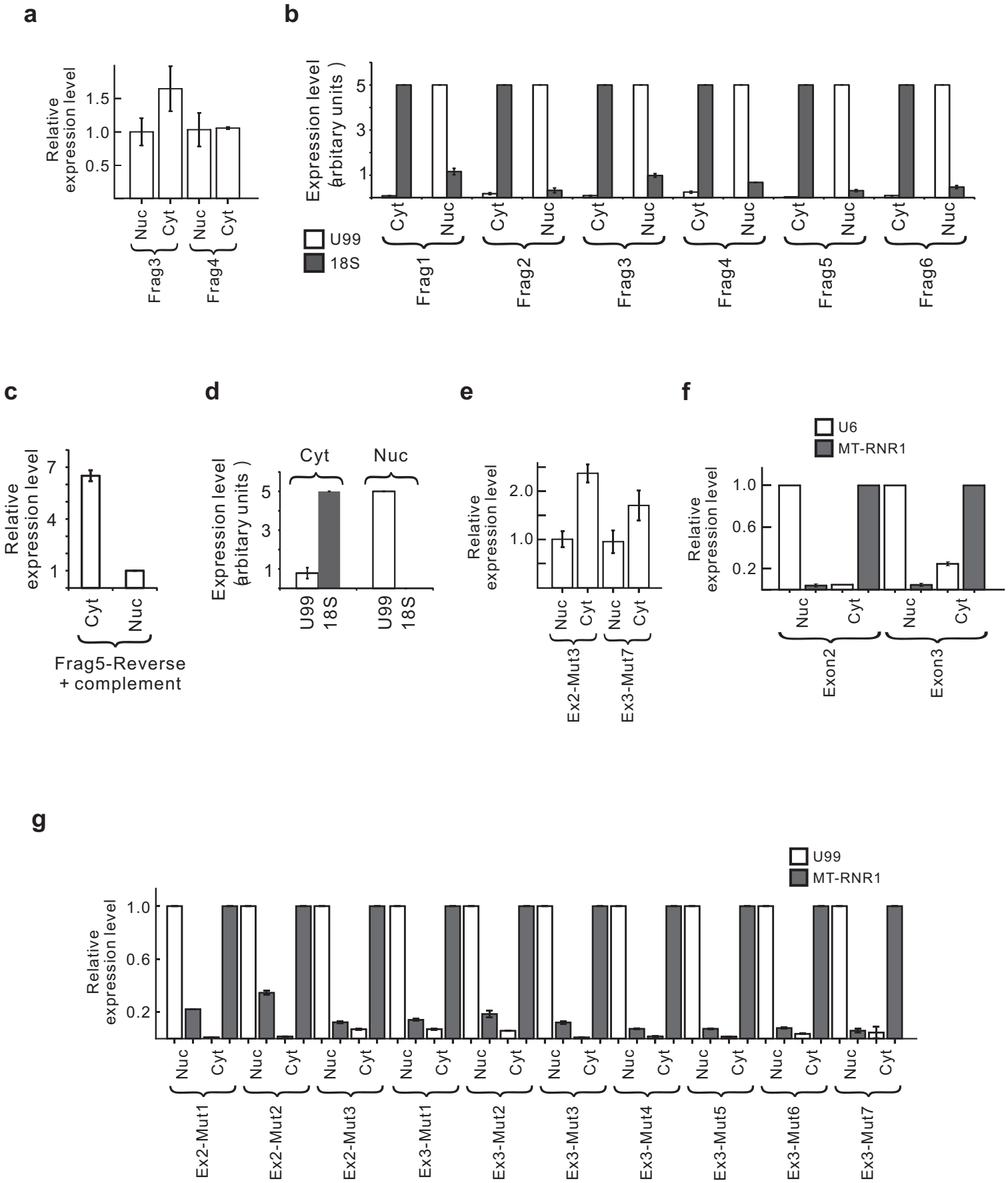
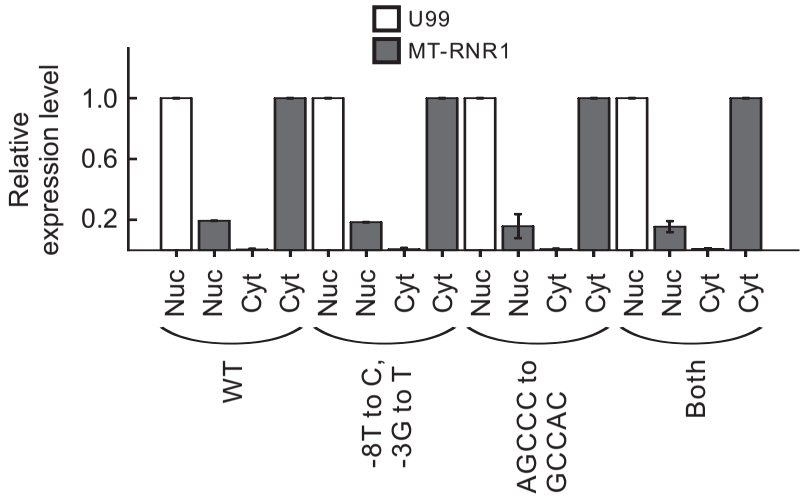

a
Relative expression level
1.5
1.0
0.5
Nuc
Cyt
Nuc
Cyt
Frag3
Frag4
b
Expression level (arbitary units )
5
3
1
U99
18S
Cyt
Nuc
Frag1
Frag2
Frag3
Frag4
Frag5
Frag6
c
Relative expression level
7
4
1
Cyt
Nuc
Frag5-Reverse + complement
d
Expression level (arbitary units )
Cyt
Nuc
5
3
1
U99
18S
U99
18S
e
Relative expression level
2.0
1.0
Nuc
Cyt
Nuc
Cyt
Ex2-Mut3
Ex3-Mut7
f
U6
MT-RNR1
Relative expression level
1.0
0.6
0.2
Nuc
Cyt
Nuc
Cyt
Exon2
Exon3
g
U99
MT-RNR1
Relative expression level
1.0
0.6
0.2
Nuc
Cyt
Ex2-Mut1
Ex2-Mut2
Ex2-Mut3
Ex3-Mut1
Ex3-Mut2
Ex3-Mut3
Ex3-Mut4
Ex3-Mut5
Ex3-Mut6
Ex3-Mut7

**Figure S6**

a


U99
MT-RNR1
Relative expression level
1.0
0.6
0.2
Nuc
Nuc
Cyt
Cyt
WT
-8T to C, -3G to T
AGCCC to GCCAC
Both

b

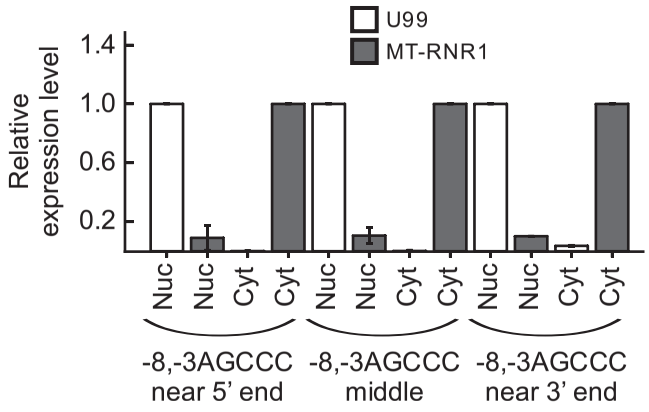

U99
MT-RNR1
Relative expression level
1.4
1.0
0.6
0.2
Nuc
Cyt
-8,-3AGCCC near 5' end
-8,-3AGCCC middle
-8,-3AGCCC near 3' end

**Figure S7**

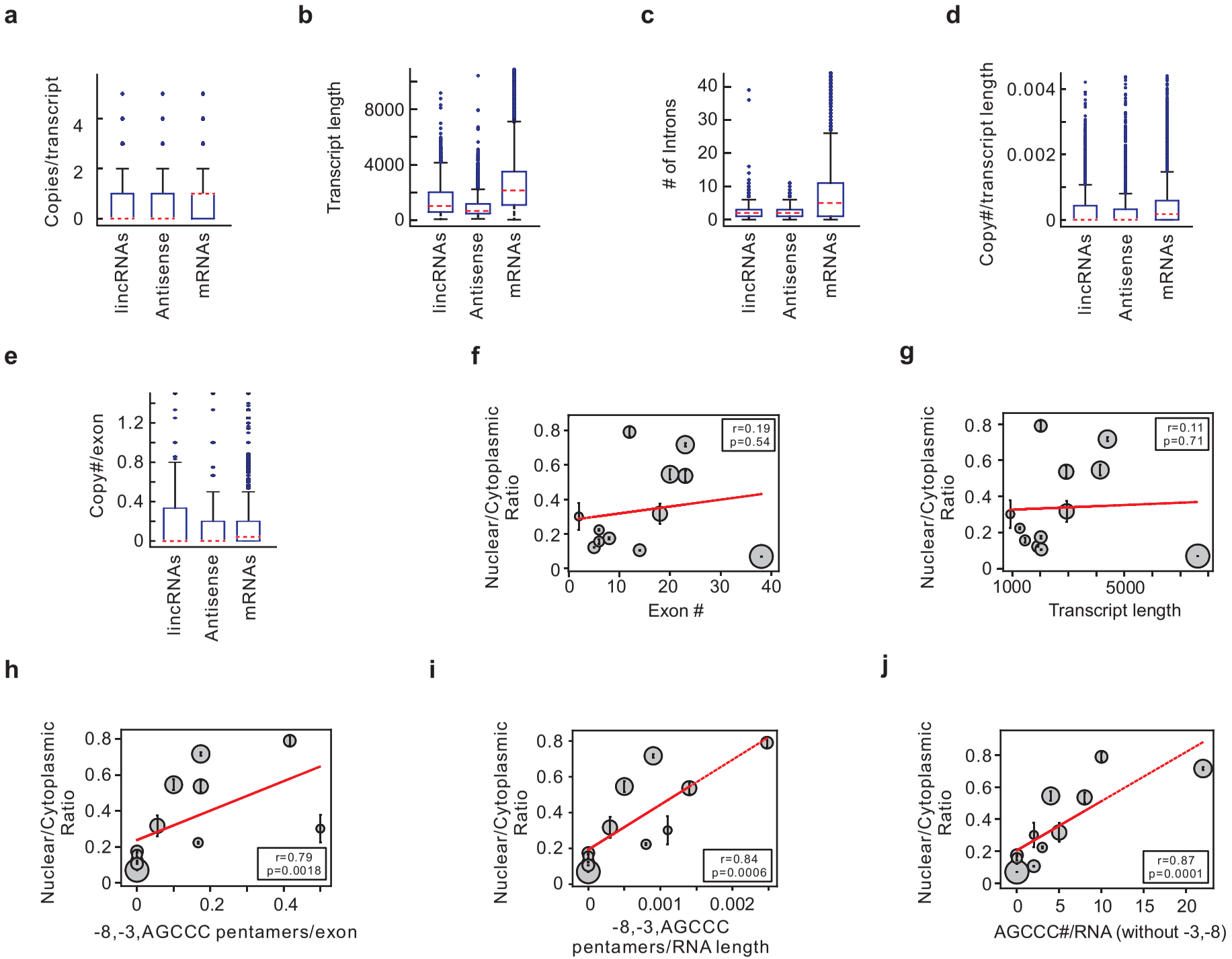

a
Copies/transcript
0
2
4
lincRNAs
Antisense
mRNAs
b
Transcript length
0
4000
8000
lincRNAs
Antisense
mRNAs
c
# of Introns
0
10
20
30
40
lincRNAs
Antisense
mRNAs
d
Copy#/transcript length
0
0.002
0.004
lincRNAs
Antisense
mRNAs
e
Copy#/exon
0
0.4
0.8
1.2
lincRNAs
Antisense
mRNAs
f
Nuclear/Cytoplasmic Ratio
r=0.19
p=0.54
Exon #
g
Nuclear/Cytoplasmic Ratio
r=0.11
p=0.71
Transcript length
h
Nuclear/Cytoplasmic Ratio
r=0.79
p=0.0018
-8,-3,AGCCC pentamers/exon
i
Nuclear/Cytoplasmic Ratio
r=0.84
p=0.0006
-8,-3,AGCCC pentamers/RNA length
j
Nuclear/Cytoplasmic Ratio
r=0.87
p=0.0001
AGCCC#/RNA (without -3,-8)

**Figure S8**

**Table S1: BORG constructs used in this study**

| Construct name | Corresponding sequence on BORG | Length | # of AGCCC sequence motifs including -3 and -8 restrictions |
|---|---|---|---|
| BORG, full length | 1-2766 | 2766 | 5 |
| Mutant 1 | 1-353+536-2766 | 2584 | 4 |
| Mutant 2 | 1-535+705-2766 | 2597 | 4 |
| Mutant 3 | 1-704+1043-2766 | 2428 | 5 |
| Mutant 4 | 1-1042+1350-2766 | 2459 | 5 |
| Mutant 5 | 1-1349+2040-2766 | 2076 | 3 |
| Mutant 6 | 1-2039+2512-2766 | 2294 | 4 |
| Exon2 | 536-1042+2604-2766 | 670 | 1 |
| Exon3 | 1043-2766 | 1724 | 3 |
| Exons1+3 | 1-535+1043-2766 | 2259 | 4 |
| Frag1 | 1-120+354-535+2604-2766 | 465 | 1 |
| Frag2 | 1-120+536-704+2604-2766 | 452 | 1 |
| Frag3 | 1-120+705-1042+2604-2766 | 621 | 0 |
| Frag4 | 1-120+1043-1349+2604-2766 | 590 | 0 |
| Frag5 | 1-120+1350-2039+2604-2766 | 973 | 2 |
| Frag6 | 1-120+2040-2511+2604-2766 | 755 | 1 |
| Frag5-Reverse+complement | 1-120+2039-1350+2604-2766 | 973 | 0 |
| Ex2-mut1 | 581-1042+2604-2766 | 625 | 1 |
| Ex2-mut2 | 631-1042+2604-2766 | 575 | 1 |
| Ex2-mut3 | 671-1042+2604-2766 | 535 | 1 |
| Ex3-mut1 | 1043-2139+2604-2766 | 1260 | 2 |
| Ex3-mut2 | 1043-1990+2604-2766 | 1111 | 2 |
| Ex3-mut3 | 1043-1880+2604-2766 | 1001 | 2 |
| Ex3-mut4 | 1043-1730+2604-2766 | 851 | 1 |
| Ex3-mut5 | 1043-1580+2604-2766 | 701 | 1 |
| Ex3-mut6 | 1043-1540+2604-2766 | 661 | 1 |
| Ex3-mut7 | 1043-1430+2604-2766 | 551 | 0 |

**Table S2: Sequence features of protein-coding RNAs analyzed in Figs. 6d and e.**

| Name | Length | N/C ratio | -8,-3,AGCCC copy # |
|---|---|---|---|
| hsf1 | 2018 | 0.7905158315 | 5 |
| actb | 1920 | 0.1220105646 | 0 |
| gapdh | 1272 | 0.2230156096 | 1 |
| stat3 | 2927 | 0.5363502412 | 4 |
| hspb1 | 930 | 0.3012853625 | 1 |
| hspa8 | 2019 | 0.1744099735 | 0 |
| met | 4140 | 0.5456501422 | 2 |
| rpa1 | 2942 | 0.3169825455 | 1 |
| thoc2 | 7622 | 0.069348092 | 0 |
| casp6 | 1456 | 0.1560855027 | 0 |
| hdac1 | 2038 | 0.1058533347 | 0 |
| stat2 | 4397 | 0.7170465199 | 4 |

**TABLE S3: Sequence of PCR primers used in this study.** Primer sequences are shown in a 5' to 3' orientation. The rightmost column indicates whether the primers flank an exon-exon junction and thus, whether they are specific to the spliced message. Y: yes, N: No, NA: not applicable.

| Target construct | Sequence of 5' primer | Sequence of 3' primer | Flanks exon-exon junction? |
|---|---|---|---|
| Full length mouse borg | GACAGATCAAGATCACCAGGGCTG GGAAGTTTTCT | GGAAGCACGAAAGCTTGAAGGCAG GCAGACAGGAT | Y |
| Mutant-1 | AGGAGCGAGATCCCAGTGTA | TGAGGTTCAGTCCGTTGTCA | NA |
| Mutant-2 | GGAGACAAAGTGCCAAGCTC | TTCCACTGCTTTCCAAATCC | NA |
| Mutant-3 | CCATGATGACAACGGACTGA | TTTTTGCTGGTTGCTGATTG | NA |
| Mutant-4 | GTCTGCCTGAGGCTCAAGTG | CCCACCACCCTCTCTTACTG | NA |
| Mutant-5 | TAGTCAGCTG CTGTGGACAA TC | TTGCAGAGAGAAGCCTGCTTCT | NA |
| Mutant-6 | ATATAACCAC GTCCATAACC AA | TCTTGTTTTGATACACAGTGCA | NA |
| U6 | GTGCTCGCTTCGGCAGCACA | GGAACGCTTCACGAATTTGCGTGTC AT | N |
| U99 | CCTCCTTTTCTTGGCGGGGA | CGTCTGAGGATAGAACCCGC | N |
| MT-RNR1 | ATTTCATTGGCCGACAGCTA | AGGTAGAGCGGGGTTTATCG | N |
| 18S | TTGACGGAAGGGCACCACCAG | GCACCACCACCCACGGAATCG | N |
| GAS5 | TCTCACAGCCAGTTCTGTGG | CAGCCTCAAACTCCACCATT | Y |
| Human BORG | CGATGGGTCACATGACAAAG | TTCTCCTGGGGGAAAAATAGA | N |
| Frag1 | GGAGACAAAGTGCCAAGCTC | TCTTGTTTTGATACACAGTGCA | NA |
| Frag2 | CCATGATGACAACGGACTGA | TCTTGTTTTGATACACAGTGCA | NA |
| Frag3 | GTCTGCCTGAGGCTCAAGTG | TCTTGTTTTGATACACAGTGCA | NA |
| Frag4 | TAGTCAGCTG CTGTGGACAA TC | TCTTGTTTTGATACACAGTGCA | NA |

| Target construct | Sequence of 5' primer | Sequence of 3' primer | Flanks exon-exon junction? |
|---|---|---|---|
| Frag5 | ATATAACCAC GTCCATAACC AA | TCTTGTTTTGATACACAGTGCA | N |
| Frag6 | GCTCCAACAGCAAACTGACA | TCTTGTTTTGATACACAGTGCA | NA |
| Frag5-reverse+ complement | CCCACCACCCTCTCTTACTG | TCTTGTTTTGATACACAGTGCA | NA |
| Exon2 | GTCTGCCTGA GGCTCAAGT | GGTGGCCTCAGTGTGGATG | NA |
| Ex2-mut1 | GTCTGCCTGA GGCTCAAGT | GGTGGCCTCAGTGTGGATG | NA |
| Ex2-mut2 | GTCTGCCTGA GGCTCAAGT | GGTGGCCTCAGTGTGGATG | NA |
| Ex2-mut3 | GTCTGCCTGA GGCTCAAGT | GGTGGCCTCAGTGTGGATG | NA |
| Exon3 | GCTCCAACAGCAAACTGACA | AAGAAGGAGCGCTGTGATTC | NA |
| Ex3-Mut1 | CCTGAAGGCAAGAAAACCACTC | GGTGGCCTCAGTGTGGATG | NA |
| Ex3-Mut2 | GTCAGCTGA GGTGCTACAGG | GGTGGCCTCAGTGTGGATG | NA |
| Ex3-Mut3 | GCAGGCATTACAGGGGCA | GGTGGCCTCAGTGTGGATG | NA |
| Ex3-Mut4 | TCCTGATGAGGGAACCAGCT | GGTGGCCTCAGTGTGGATG | NA |
| Ex3-Mut5 | GAAGCTAAGCCCCAGGGATA | GGTGGCCTCAGTGTGGATG | NA |
| Ex3-Mut6 | GAAGCTAAGCCCCAGGGATA | GGTGGCCTCAGTGTGGATG | NA |
| Ex3-Mut7 | TAGTCAGCTG CTGTGGACAA TC | GGTGGCCTCAGTGTGGATG | NA |
| Hsf1 | GGGAAACAGGAGTGTATGGACT | CTTGTTGACAACTTTTTGCTGCT | Y |
| Actb | GAA ATC GTG CGT GAC ATC AAA G | TGT AGT TTC ATG GAT GCC ACA G | Y |
| gapdh | AGGTCGGTGTGAACGGATTTG | TGTAGACCATGTAGTTGAGGTCA | Y |
| Stat3 | CACCTTGGATTGAGAGTCAAGAC | AGGAATCGGCTATATTGCTGGT | Y |
| Hspb1 | CACTGGCAAGCACGAAGAAAG | GCGTGTATTTCCGGGTGAAG | Y |

| Target construct | Sequence of 5' primer | Sequence of 3' primer | Flanks exon-exon junction? |
|---|---|---|---|
| Hspa8 | TCTCGGCACCACCTACTCC | CCCGATCAGACGTTTGGCA | Y |
| Met | CGCAGAAGTTCACCACCAAGTC | AGCATCACTTCGTACAAGGCGT | Y |
| Rpa1 | ACATCCGTCCCATTTCTACAGG | CTCCCTCGACCAGGGTGTT | Y |
| Thoc2 | CACATAACCGTTGAGCCTCTC | ACTTGCTTCGGTGCTCTCTTG | Y |
| Casp6 | GTTTACGCATACGACGCCAAA | CCACGGGTACGTCATGCTG | Y |
| Hdac1 | TTCTGTTACGTCAATGACATCGT | GACCCGGTCTGTAGTATAGAAGG | Y |
| Stat2 | CTGAAGGACGAACAGGATGTC | CAGGGTGGTTAATCGGCCAA | Y |
| Bc1 | GGGTTGGGGATTTAGCTCAG | AGGTTGTGTGTGCCAGTTACC | N |
| Hotair | GCGCTAAGTCCTTCCAGAGA | GACTTCCTTCCTTCCGCTCT | Y |
| Tsix | TCTCGTTGATTCACGCTGAC | TGGCACTTTGATCGTCCATA | N |
| Xist | CTTTGGGCTTAGGTGAGCAG | CACCAAATTACTCGGCCACT | N |
| Tug1 | CTG TTG ATC TTG CCG TGA GA | TCA CCA TGG ACA AAA ATC CA | Y |

**Table S4. The number of -8A/T,-3G/C,AGCCC motifs in lncRNAs analyzed in Fig. 6D in mouse and human.**

* The 5' end of human BORG RNA has not yet been determined. The shown number refers to the number of motifs found in the sequences that we have been able to confirm as part of the human BORG RNA (unpublished data).

| lncRNA | # of motifs in mouse RNA | # of motifs in human RNA |
|---|---|---|
| Xist | 15 | 4 |
| HOTAIR | 1 | 2 |
| TUG1 | 1 | 2 |
| TSIX | 2 | 6 |
| BORG | 5 | 1* |
| GAS5 | 0 | 0 |