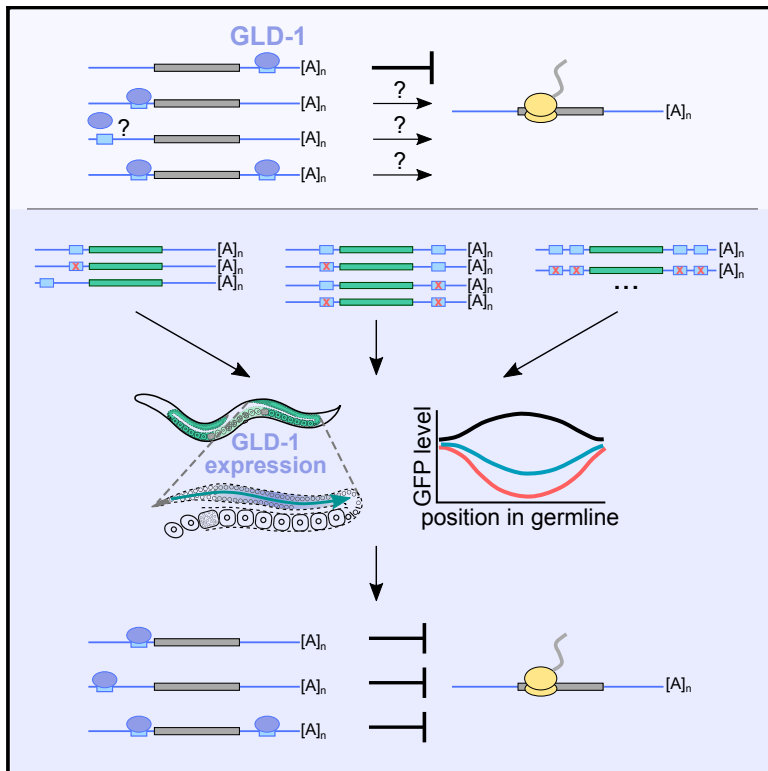


## Post-transcriptional Regulation by 3' UTRs Can Be Masked by Regulatory Elements in 5' UTRs

### Graphical Abstract



### Authors

Kathrin Theil, Margareta Herzog,  
Nikolaus Rajewsky

### Correspondence

rajewsky@mdc-berlin.de

### In Brief

The main hubs in post-transcriptional regulation are the UTRs of mRNAs. While 3' UTRs are well studied, studies on specific regulation by 5' UTRs are scarce. Theil et al. show that 5' UTR sites can be highly functional and that they can mask critical functions of 3' UTR sites.

### Highlights

- GLD-1 binding sites in 5' UTRs can be strong repressive elements
- Regulation via 5' UTRs does not stringently depend on binding site position
- 5' and 3' UTR sites within the same transcript repress translation independently
- Sites in one UTR can mask functions of sites in the other UTR



# Post-transcriptional Regulation by 3' UTRs Can Be Masked by Regulatory Elements in 5' UTRs

Kathrin Theil,<sup>1</sup> Margareta Herzog,<sup>1</sup> and Nikolaus Rajewsky<sup>1,2,\*</sup><sup>1</sup>Systems Biology of Gene Regulatory Elements, Berlin Institute for Medical Systems Biology (BIMSB), Max Delbrück Center for Molecular Medicine in the Helmholtz Association (MDC), 13125 Berlin, Germany<sup>2</sup>Lead Contact\*Correspondence: [rajewsky@mdc-berlin.de](mailto:rajewsky@mdc-berlin.de)  
<https://doi.org/10.1016/j.celrep.2018.02.094>

## SUMMARY

In mRNA sequences, 3' UTRs are thought to contain most elements that specifically regulate localization, turnover, and translation. Although high-throughput experiments indicate that many RNA-binding proteins (RBPs) also bind 5' UTRs, much less is known about specific post-transcriptional control exerted by 5' UTRs. GLD-1 is a conserved RBP and a translational repressor with essential roles in *Caenorhabditis elegans* germ cell development. Previously, we showed that GLD-1 binds highly conserved sites in both 3' and 5' UTRs. Here, by targeted single-copy insertion of transgenes, we systematically tested *in vivo* functionality of 5' and 3' UTR binding sites individually and in combination. Our data show that sites in 5' UTRs mediate specific and strong translational repression, independent of exact position. Intriguingly, we found that the functionality of 3' UTR sites can be masked by 5' UTR sites and vice versa. We conclude that it is important to study both UTRs simultaneously.

## INTRODUCTION

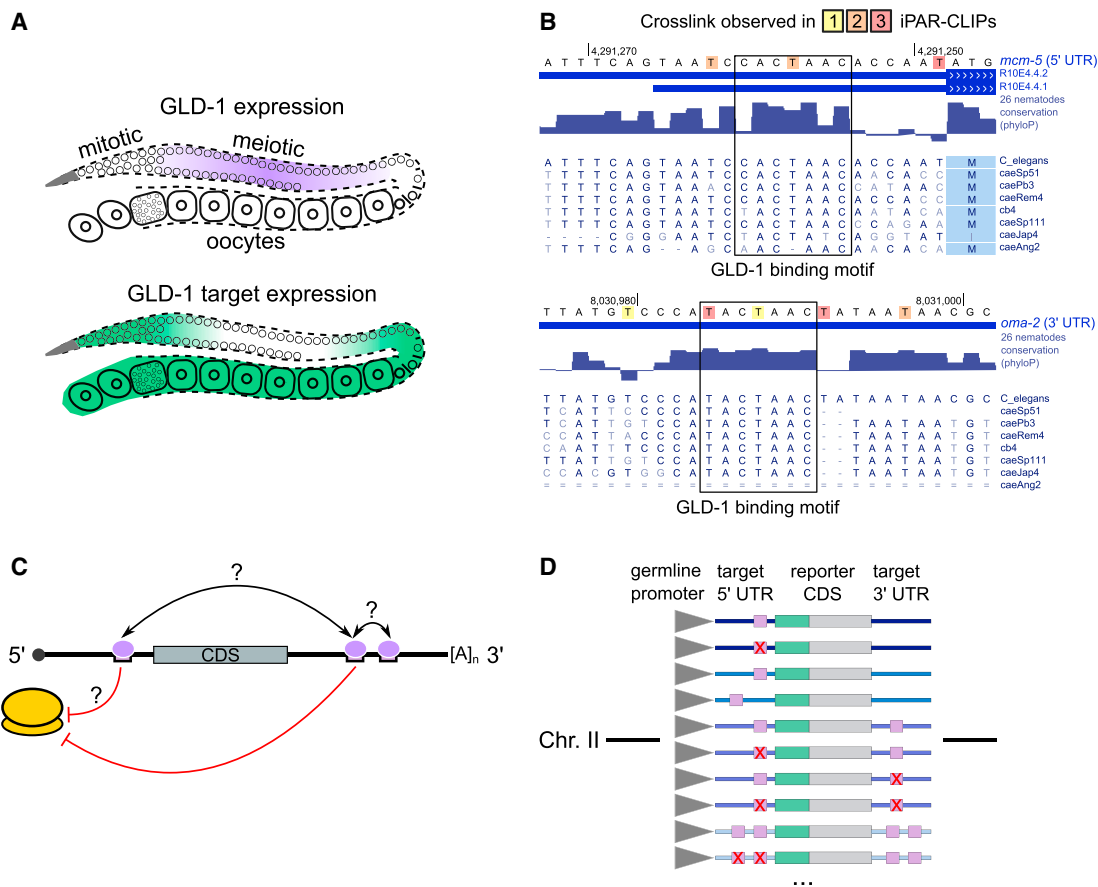
In the life of animal mRNAs, translation, transport, storage, and turnover are predominantly regulated by *cis*-regulatory sites that are bound by *trans*-acting RNA-binding proteins (RBPs) or ribonucleoprotein complexes. These sites usually reside in the UTRs of mRNAs. UTR length can differ greatly between genes and depending on expression context, e.g., tissue type. Interestingly, while average 5' UTR length stayed relatively constant during animal evolution (100–200 nt), 3' UTR length expanded with the number and complexity of different cell types and tissues (Pesole et al., 2001; Sood et al., 2006). For instance, 3' UTRs in the roundworm *Caenorhabditis elegans* are approximately 200 nt in length, while human 3' UTRs are, on average, 1,000 nt long (Mangone et al., 2010; Pesole et al., 2001). Importantly, 5' UTRs act as platforms on which ribosomes bind, scan, and initiate translation. This, together with the limited length, is thought to restrict the regulatory potential of 5' UTRs. Consequently, the vast majority of studies elucidating mechanisms of specific spatial or temporal post-transcriptional gene regulation have focused on dissecting

3' UTRs, while there are only few studies assaying RBPs that bind 5' UTRs to induce specific post-transcriptional control (e.g., Graindorge et al., 2013; Gray and Hentze, 1994; Medenbach et al., 2011). In general, 3' UTRs are commonly treated as modular units that mediate regulation independently of 5' UTRs.

Recent data motivated us to study whether 3' and 5' UTRs can act together to mediate regulation. Transcriptome-wide RNA-RBP interaction studies have, across animals and systems, reproducibly identified hundreds of RBP binding sites located in 5' UTRs (Baltz et al., 2012; Bliin et al., 2015; Rybak-Wolf et al., 2014; Van Nostrand et al., 2017). GLD-1 (germline development defective 1) is an RBP belonging to the highly conserved STAR (signal transduction and activation of RNA) protein family (Jones and Schedl, 1995; Vernet and Artzt, 1997) and promotes meiosis and oogenesis in the *C. elegans* germline by repressing translation of its target transcripts (Figure 1A; reviewed in Lee and Schedl, 2010). We established iPAR-CLIP (*in vivo* photoactivatable-ribonucleoside-enhanced crosslinking and immunoprecipitation) to identify RBP binding sites in *C. elegans* transcriptome-wide (Jungkamp et al., 2011). This method revealed that GLD-1 binds conserved sites in both UTRs of its target mRNAs (examples are given in Figure 1B). Furthermore, we found that GLD-1 5' UTR sites tend to reside close to the translation start codon. While many GLD-1 binding sites in 3' UTRs have been shown to be functionally important, so far, only one case of 5' UTR-mediated regulation by GLD-1 has been studied mechanistically (*gna-2* transcript; Lee and Schedl, 2004). In this case, it was proposed that GLD-1 binding to the 5' UTR prevents nonsense-mediated mRNA decay of *gna-2* by repressing translation of upstream open reading frames (uORFs). Interaction with a uORF was also proposed as a mechanism for one of the few other RBPs for which sites in 5' UTRs were studied mechanistically (Medenbach et al., 2011). We observed that the majority of 5' UTRs in which GLD-1 binding sites reside do not contain uORFs, suggesting functions of 5' UTRs that have escaped detection. Moreover, although we found many transcripts to be bound in both 5' and 3' UTR (Figure S1A), joint regulation by GLD-1 sites in both UTRs has not been studied systematically to date.

The *C. elegans* germline is a well-established model to study post-transcriptional regulation, and germline expression patterns are thought to be predominantly determined by signals within mRNA 3' UTRs (Merritt et al., 2008). We reasoned that assessing regulation by the RBP GLD-1 within this *in vivo* system offers an intriguing opportunity to test functions of sites in 5' UTRs and possible crosstalk between active sites within 5' and 3' UTRs.





**Figure 1. Addressing the Function of Individual GLD-1 Binding Sites *In Vivo***

(A) GLD-1 is an RNA-binding protein expressed in the medial gonad of *C. elegans*. Binding by GLD-1 leads to translational repression of its target RNAs. (B) Examples of GLD-1 binding sites investigated in this study (see also Figure S1 and Table 1). GLD-1 binds conserved sites in target 5' UTRs, in addition to 3' UTR sites (Jungkamp et al., 2011). Crosslinking sites from iPAR-CLIP experiments, indicating direct protein-RNA contacts, were found in the immediate vicinity of GLD-1 binding motifs. While the motifs themselves are highly conserved, conservation drops in flanking regions. GLD-1 5' UTR sites tend to reside close to the translation start codon (Jungkamp et al., 2011). Genome browser screen shots adapted from <http://genome.ucsc.edu> (Kent et al., 2002). (C) Individual GLD-1 sites in 5' and 3' UTRs may act independently or cooperate in translational repression. (D) To investigate the regulatory impact of individual GLD-1 binding sites *in vivo*, transgenic worms were generated for four different targets (Table 1) by single-copy insertion of reporter sequences into the same genetic locus (Frøkjær-Jensen et al., 2008). Examples of constructs with wild-type, shifted, or mutated binding sites are shown (see also Experimental Procedures). Violet boxes indicate GLD-1 binding sites. CDS, coding sequence. See also Figure S1.

We set out to investigate the role of GLD-1 5' UTR sites and their interplay with 3' UTR sites within the same transcript; specifically, we sought to address (1) whether GLD-1 5' UTR sites are functional, (2) whether the position of the GLD-1 binding site within the 5' UTR architecture is critical for function, and (3) whether and how sites in target 5' and 3' UTRs jointly impact translation (Figure 1C). To this end, we inserted single copies of transgenes into a specific locus of the *C. elegans* genome, consisting of a GFP::H2B (histone H2B) coding sequence (CDS) flanked by selected GLD-1 target 5' and 3' UTRs (Figure 1D). Since the transgenic worms we generated are genetically identical, we were able to compare reporter expression pattern changes caused by engineered sequence alterations. We systematically recorded and quantified GFP expression across the germline. Our data show that all tested sites in 5'

UTRs were functional, while exact position did not seem to be critical. Most importantly, our data show that 5' UTR sites can mask regulatory effects from sites in 3' UTRs, and vice versa.

## RESULTS

To understand endogenous regulation by GLD-1 binding sites in target 5' and 3' UTRs, it is indispensable to perform *in vivo* studies. We generated transgenic *C. elegans* strains for 19 different reporter constructs (Figure 1D). We used the Mos1 single-copy insertion (MosSCI) technique, which allows single-copy insertions at a defined locus in the *C. elegans* genome and, thus, reproducible expression at close to endogenous levels (Frøkjær-Jensen et al., 2008). Reporter transgenes were constructed as follows (from 5' to 3'): germline-permissive

**Table 1. Selected Candidates for *In Vivo* Reporter Analysis**

Target	Length (in Nucleotides)		No. of Binding Sites		Motif	
	5' UTR	3' UTR	5' UTR	3' UTR	5' UTR	3' UTR
R09E10.6	45	25	1	0	AACTAAC/ TACTAAC <sup>a</sup>	–
<i>gipc-1</i>	62	35	1	0	TACTAAC	–
<i>mcm-5</i>	35	120	1	1	CACTAAC	TCTTAAC
<i>oma-2</i>	71	291	2	2	CACTAAC	TACTAAC
					TACTAAC	CACTAAC

<sup>a</sup>Extended GLD-1 binding site.

promoter (*Pgld-1*), GLD-1 target 5' UTR, GFP::H2B CDS, and GLD-1 target 3' UTR. The H2B CDS was included to concentrate the GFP signal to the nucleus for better visualization. To investigate the impact of individual binding sites, we identified GLD-1 motifs in iPAR-CLIP clusters (Jungkamp et al., 2011) and mutated them by replacing the two most conserved bases (positions 4 and 6) within the GLD-1 heptamer consensus sequence (Wright et al., 2011). Since the genetic background is identical in all generated strains (same insertion site), functionality of single-base substitutions can be dissected. Additionally, since GLD-1 is only expressed in the medial gonad, other regions of the germline can serve as internal controls to discriminate GLD-1-dependent from -independent effects on reporter expression. We selected candidates for in-depth reporter analysis by the following criteria: (1) presence of at least one 5' UTR binding site close to the start codon, (2) no additional binding site in the target mRNA CDS, (3) enriched expression in the germline, (4) reproducible clusters in iPAR-CLIP experiments (Jungkamp et al., 2011), and (5) amenability to site-directed mutagenesis. Selected candidates are listed in Table 1, and corresponding genome browser snapshots with iPAR-CLIP crosslink sites and conservation tracks are shown in Figures 1B and S1B.

### Single 5' UTR Binding Sites Are Functional *cis*-Regulatory Elements

Many GLD-1 targets with a 5' UTR binding site harbor additional iPAR-CLIP clusters in other transcript regions (Figure S1A; Jungkamp et al., 2011). Hence, we first asked whether targets solely bound in their 5' UTRs are efficiently repressed. We generated transgenic reporters for two targets, exclusively bound in their 5' UTRs. The first candidate, R09E10.6, harbors an extended GLD-1 binding site consisting of two GLD-1 consensus motifs (Figure 2A), while the second candidate, *gipc-1*, contains only one strong GLD-1 motif in close proximity to the start codon (Figure 2C). For both targets, reporters with intact GLD-1 binding sites were expressed in a typical GLD-1-dependent pattern with repression in the medial gonad, where GLD-1 levels are high (Figures 2B and 2D). Mutation of sites led to increased GFP expression in this region, indicating that repression, indeed, depends on GLD-1 binding (Figures 2B and 2D; RNA levels in Figures S2A and S2B). We conclude that binding sites in transcript 5' UTRs are functional and that one 5' UTR site is sufficient to confer repression.

### 5' UTR Binding Sites Can Act Independently of Relative Position

The majority of 5' UTR binding sites detected by iPAR-CLIP reside in close proximity to the translation start codon (Jungkamp et al., 2011). To investigate whether this architecture and the exact context of the binding motif are required to mediate repression, we generated variants of the *gipc-1* reporter construct (5' UTR length: 62 nt) with upstream-shifted GLD-1 binding motifs (Figure 2C). We devised two different strategies. First, we deleted the original GLD-1 binding site and generated a new binding motif by substituting bases 25 nt upstream of the original motif. Second, we shifted the original binding site by inserting a spacer sequence of 12 nt upstream of the start AUG (Figure 2C; RNA levels in Figure S2B). Both reporters exhibited the characteristic repression in the GLD-1 expression domain indicating GLD-1 dependency (Figure 2D).

As visual evaluation of the repressive state might be prone to biases, we aimed at quantitatively assessing reporter expression along the germline (cf. Farley and Ryder, 2012; Wright et al., 2011). To this end, we measured GFP pixel intensities in worm gonads of each two independent transgenic lines per reporter construct. We recorded GFP intensities from the distal tip of the gonad, where reporter expression is not subject to regulation by GLD-1, over the medial gonad, where GLD-1 levels are high, to the bend, where GLD-1 levels decrease again (Figure 2E, scheme). Independent lines for the same construct showed highly reproducible expression patterns (Figure S2C). While deleting the GLD-1 binding site led to a clearly distinct reporter expression profile, both constructs with shifted binding sites exhibited an expression pattern indistinguishable from that of the wild-type reporter (Figure 2E).

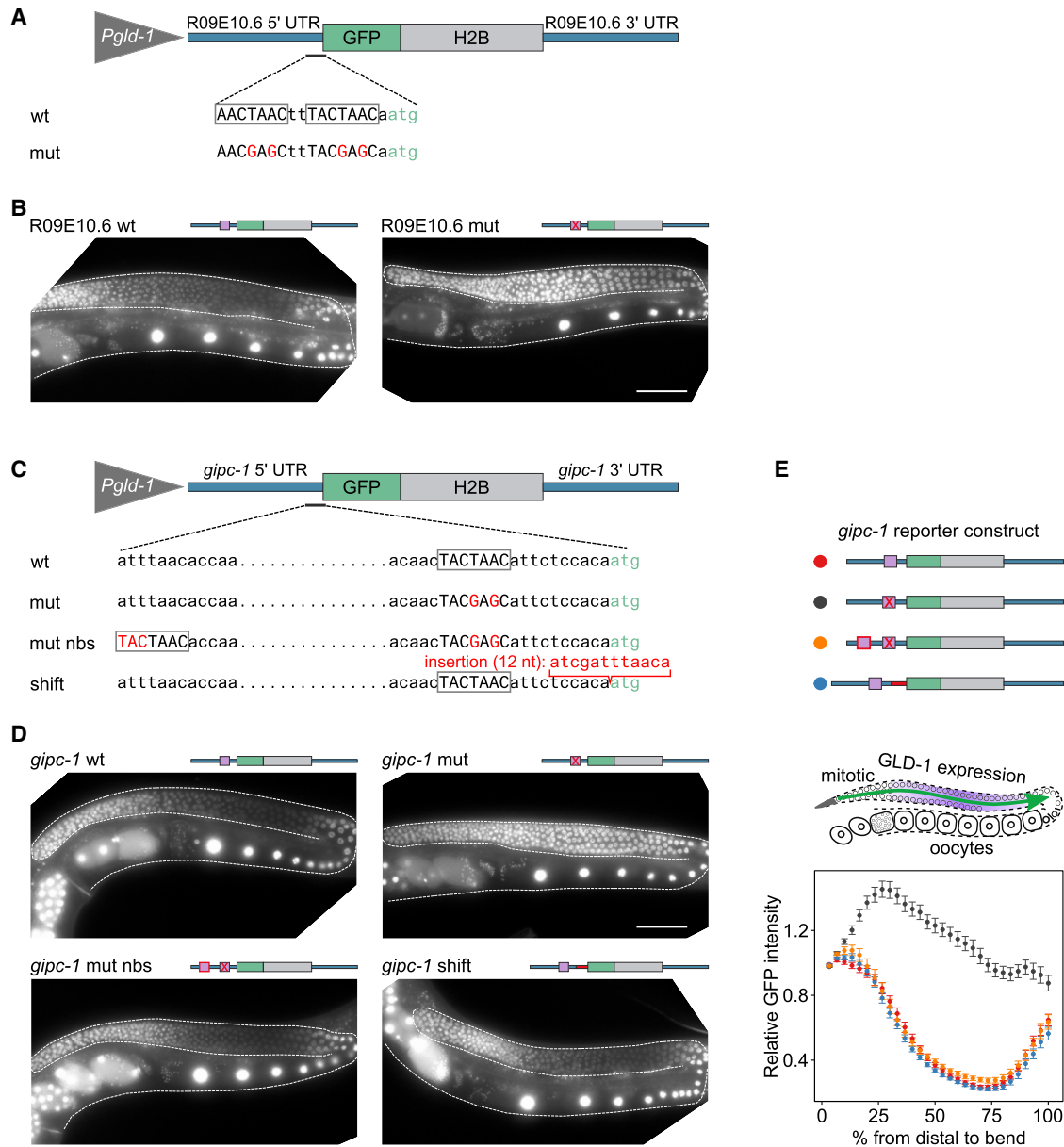
In conclusion, close proximity to the translation start codon and immediate context of the binding site are not stringent requirements for translational repression.

### 5' and 3' UTR Binding Sites within the Same Target Can Both Act as Potent Functional Elements

Many GLD-1 targets with 5' UTR binding sites are also bound in other transcript regions, mainly in 3' UTRs (Figure S1A; Jungkamp et al., 2011). Hence, we asked how individual sites within 5' and 3' UTRs of the same target contribute to repression. Conceivable scenarios are: (1) only one of the sites is functional, (2) both sites are functional and contribute to repression independently, and (3) both sites cooperate. To systematically address the mode of repression in “multi-site” targets, we generated reporters for two transcripts with binding sites in both 5' and 3' UTRs.

### The *mcm-5* 5' UTR Binding Site Confers Stronger Repression than the 3' UTR Site

The first candidate, *mcm-5*, harbors two reproducibly detected GLD-1 binding sites, one in each UTR. We individually and jointly mutated these sites to evaluate their functional impact (Figure 3A). Additionally, to exclude potential other regulatory elements within the target 3' UTR, we generated a reporter strain where we replaced the *mcm-5* 3' UTR with an unrelated 3' UTR without GLD-1 binding sites (*unc-54* 3' UTR; Figure S3A). While the 5' UTR site alone exhibited a repressive effect similar to that of two intact sites, the sole presence of one functional 3'



**Figure 2. GLD-1 5' UTR Binding Sites Mediate Translational Repression In Vivo, Independent of Exact Surrounding Sequence Context**

Reporter constructs with wild-type, mutated, or shifted GLD-1 binding sites were introduced into *C. elegans*.

(A and C) Transgene design of R09E10.6 (A) and *gipc-1* (C) reporter constructs with blow-up of GLD-1 binding site close to the translation start codon. *Pgld-1*, *gld-1* promoter; capital letters indicate GLD-1 binding motif; boxes indicate intact sites; red indicates mutated bases in the binding motif; dots indicate interjacent nucleotides.

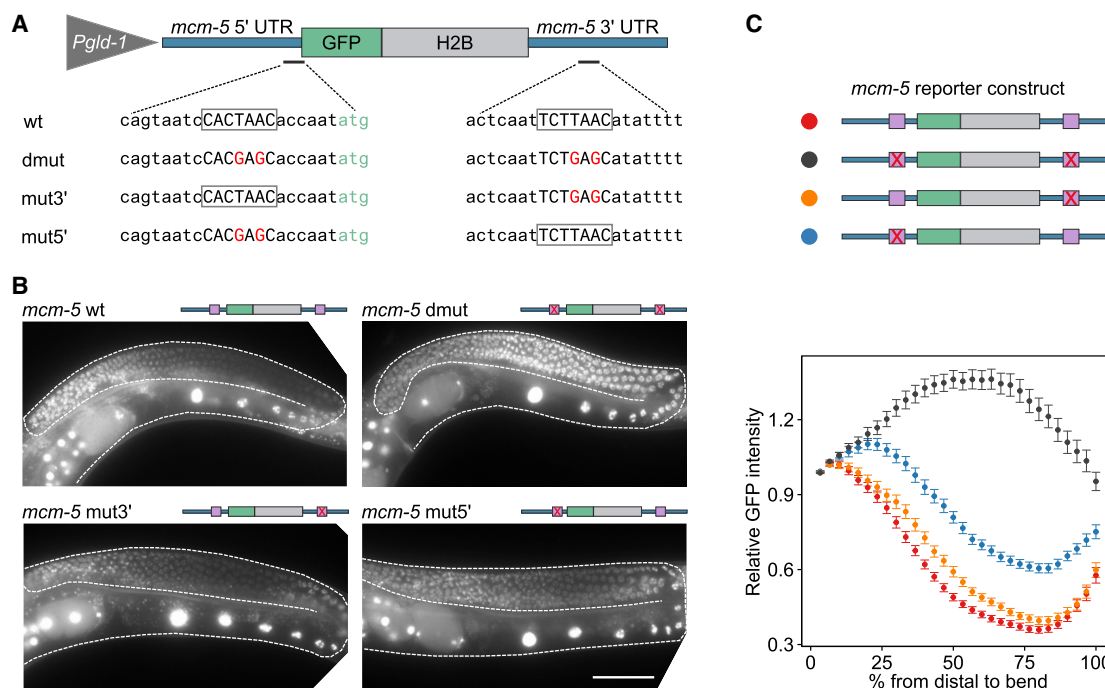
(B and D) For both R09E10.6 (B) and *gipc-1* (D) reporters, mutation of the GLD-1 binding site leads to reporter derepression in the medial gonad, where GLD-1 is expressed. Shifting the GLD-1 binding site away from the start codon retains reporter repression (D). Dashed lines encircle gonads. Scale bars, 50  $\mu$ m.

(E) Quantification of reporter expression along the germline (scheme) confirms functionality of binding sites in *gipc-1* reporters. Measured GFP pixel intensities were normalized to the mean of the first 5% of each gonad (where GLD-1 is not expressed), and grouped into 30 bins. Displayed are mean  $\pm$  SEM for each bin. Per reporter construct, two independent transgenic lines were analyzed and averaged (number of gonads: wt, n = 46; mut, n = 56; mut nbs, n = 44; shift, n = 50). The method was adapted from Wright et al. (2011).

See also Figure S2.

UTR site resulted in much weaker repression (Figure 3B). This observation is in line with the motif score proposed by Wright et al. (2011), which predicts strong binding for the 5' UTR while

the 3' UTR motif was not predicted to be bound with high affinity. However, the 3' UTR motif agrees with the consensus motif derived from GLD-1 iPAR-CLIP experiments; and T-to-C



**Figure 3. *mcm-5* 5' and 3' UTR GLD-1 Binding Sites Both Individually Mediate Different Degrees of Translational Repression**

Reporter constructs with wild-type or mutated GLD-1 binding sites were introduced into *C. elegans*.

(A) Transgene design with blow-up of GLD-1 binding sites (analogous to Figures 2A and 2C).

(B) Mutation of the 5' UTR GLD-1 binding site leads to partial reporter derepression in the medial gonad, where GLD-1 is expressed. Mutation of the 3' UTR site results in similar repression as in the wild-type reporter. Mutation of both sites leads to complete derepression. Dashed lines encircle gonads. Scale bar, 50  $\mu$ m.

(C) Quantification of reporter expression (analogous to Figure 2E) along the germline confirms functionality of both binding sites. Displayed are mean  $\pm$  SEM for each bin. Per reporter construct, two independent transgenic lines were analyzed and averaged (number of gonads: wt, n = 56; dm, n = 64; mut3', n = 54; mut5', n = 48).

See also Figures S3 and S4.

conversions, indicating protein crosslinking, were reproducibly detected nearby the 3' UTR motif (Figure S1B). Replacement of the *mcm-5* 3' UTR by the *unc-54* 3' UTR preserved the characteristic GLD-1-dependent local repression in the medial gonad, suggesting that the target 5' UTR alone is sufficient to confer regulation (Figure S3A).

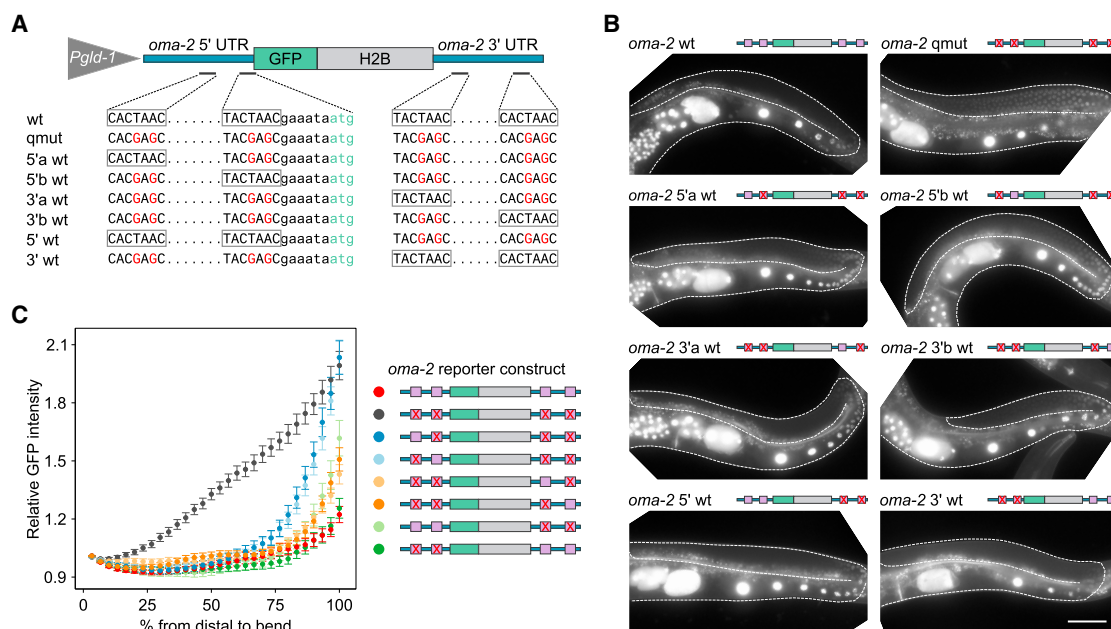
Quantification of GFP reporter expression for *mcm-5* strains confirmed a stronger repressive effect of the 5' UTR GLD-1 binding site compared to the 3' UTR site (Figures 3C and S3B). This was not due to differences in RNA stability, as all reporters were expressed at similar levels (Figure S3C); yet, the quantification unequivocally demonstrated that the 3' UTR site is functional. Only the mutation of both 5' and 3' UTR sites led to full derepression, with a clearly distinct reporter expression pattern along the gonad for the double mutant. The reporter with intact 5' UTR site and replaced 3' UTR exhibited reduced expression in the medial gonad, characteristic for GLD-1-dependent regulation, indicating that the 5' UTR alone is sufficient to confer repression (Figure S3B). We conclude that both 5' and 3' UTR sites are functional.

### Single Intact Binding Sites in the *oma-2* Transcript Are Sufficient to Repress Translation

The second “multi-site” target, *oma-2*, harbors two GLD-1 binding sites in both 5' and 3' UTRs, respectively (Figure 4A). All four

sites are predicted to be bound with high affinity (Jungkamp et al., 2011; Wright et al., 2011). We generated eight different reporter constructs to dissect the contributions of individual sites to repression (Figure 4A). To disentangle the impact of each UTR on repression, we first mutated both sites in either the 5' or the 3' UTR. To evaluate contributions by individual sites, additional reporters were generated with all but one binding sites mutated. Only the deletion of all GLD-1 binding sites in both 5' and 3' UTRs led to clearly visible derepression in the medial gonad (Figure 4B).

Quantification along the gonads of *oma-2* transgenic worms confirmed that all sites individually confer strong repression (Figures 4C and S3D; RNA levels in Figure S3E); GFP expression followed very similar profiles for all reporters with at least one intact binding site. Subtle differences only arose in the last 25% before the bend of the gonad—the area where GLD-1 levels decrease again. In this region, reporters with only one intact site in the *oma-2* 5' UTR were more derepressed than reporters with only one 3' UTR site (blue dots versus orange dots; Figure 4C). Consistently, retaining both 3' UTR sites resulted in stronger repression than retaining both 5' UTR sites (green versus light green dots; Figure 4C). The profiles additionally suggest that both sites within one UTR had similar repressive strength. Of note, motifs in the *oma-2* 5' UTR are the same as in the



**Figure 4. All *oma-2* GLD-1 Binding Sites in Both UTRs Are Functional Repressive Elements**

Reporter constructs with wild-type or mutated GLD-1 binding sites were introduced into *C. elegans*.

(A) Transgene design with blow-up of GLD-1 binding sites (analogous to Figures 2A and 2C).

(B) Only mutation of all sites results in derepression of reporter expression in the medial gonad, where GLD-1 levels are high. Dashed lines encircle gonads. Scale bar, 50  $\mu$ m.

(C) Quantification of reporter expression (analogous to Figure 2E) indicates that single sites independently confer strong repression. Displayed are mean  $\pm$  SEM for each bin. Per reporter construct, two independent transgenic lines were analyzed and averaged (number of gonads: wt, n = 54; qmut, n = 64; 5'a wt, n = 56; 5'b wt, n = 58; 3'a wt, n = 48; 3'b wt, n = 56; 5' wt, n = 40; 3' wt, n = 56).

See also Figures S3 and S4.

3' UTR, underlining that a motif alone is not sufficient to accurately predict repressive strength. Importantly, the expression of reporters with single intact sites was very similar to wild-type reporter expression (with four intact sites), and only the deletion of all sites resulted in significant derepression. We conclude that all sites were individually capable of mediating strong repression *in vivo*.

Interestingly, the expression of *oma-2* and *mcm-5* reporters followed different patterns along the germline (Figures 3C and 4C). Since both reporters are transcribed from the same promoter, this is likely a consequence of differential, GLD-1-independent post-transcriptional regulation. The *oma-2* reporter is repressed in the distal germline, where FBF proteins mediate target repression. In fact, *oma-2* was found to be a direct target of FBF-1 (Prasad et al., 2016). This suggests that the distinct expression pattern of the *oma-2* reporter might arise from regulation by at least two RBPs.

### Modeling Repression of “Multi-site” GLD-1 Targets Suggests that Sites Act Independently

To understand how different GLD-1 binding sites act together in *mcm-5* and *oma-2* regulation, we modeled theoretical wild-type reporter repression. We first estimated the degree of repression for reporters with single intact binding sites and for wild-type reporters, with two and four intact sites, respectively. For this, we assumed reporter repression in the medial gonad of wild-

type reporters to be maximal and GLD-1-independent expression throughout the gonad of reporters without GLD-1 sites. Based on estimated repression of “single-site” reporters, we calculated theoretical wild-type reporter repression according to two models: (1) assuming that sites within a target independently confer repression (multiplication of repression fold changes) and (2) assuming independent binding but saturation of repression for a transcript bound at one or multiple sites (addition of repression fold changes). While the calculated models are both consistent with repression of the *mcm-5* wild-type reporter (Figure S4A), for *oma-2*, the models do not reflect the experimentally measured repression (Figure S4B).

We believe that this is likely due to uncertainties in repression estimates. Since the *oma-2* reporter is not expressed in the distal germline, quantifications are much more affected by noise in background fluorescence. With the uncertainties in our estimations, we are not able to discriminate between (1) independent contributions to repression and (2) independent binding with a repression maximum reached by binding of one site. However, for cooperative repression, the experimentally determined wild-type reporter repression is expected to be much stronger than theoretical estimations by any of the two models. Thus, based on our calculations, cooperativity of sites is unlikely.

In conclusion, motifs within “multi-site” targets can individually confer strong translational repression, independent of whether they reside in a 5' or 3' UTR (Figure S4C). 5' UTR sites

can have a stronger impact (as for *mcm-5* reporter) or be less potent (as for *oma-2* reporter) than 3' UTR sites. Modeling reporter repression of “multi-site” targets, based on “single-site” reporters, suggests that repression does not rely on cooperativity of binding sites. Rather, for *mcm-5* and *oma-2*, the presence of additional sites only confers little or no additional repression. Thus, presence of one functional site can mask the presence of additional functional sites (in the same or another UTR of the same transcript).

## DISCUSSION

### GLD-1 Mediates Repression via 5' UTR Sites

Studies on post-transcriptional regulation predominantly focus on regulation by elements in 3' UTRs. Although RBP binding sites were also found in 5' UTRs (e.g., Van Nostrand et al., 2017), their functions have only been studied mechanistically for few RBPs (Graindorge et al., 2013; Gray and Hentze, 1994; Medenbach et al., 2011).

We showed previously that GLD-1 5' UTR binding sites are well conserved: often, high conservation of a GLD-1 binding motif with a drop of conservation in the immediate sequence vicinity is observed (Jungkamp et al., 2011). Additionally, perturbation experiments with subsequent proteome measurements suggested functionality of 5' UTR sites. However, 5' UTR sites were not tested directly. Here, using single-copy insertion via the MosSCI technique (Frøkjær-Jensen et al., 2008), we generated transgenic *in vivo* reporters to evaluate the importance of 5' UTR binding sites in target regulation. Importantly, our *in vivo* reporters contained both endogenous target 5' and 3' UTR sequences. We investigated the expression patterns of GLD-1 targets with sole 5' UTR binding sites and two targets with different numbers and arrangements of sites in both 5' and 3' UTRs. For all reporters, GLD-1 5' UTR sites represented potent repressive elements, individually capable of inhibiting reporter expression.

Most 5' UTR GLD-1 binding sites reside close to the translation start codon, and we proposed previously that this may have functional implications (Jungkamp et al., 2011). We put forward that this arrangement could facilitate interactions between GLD-1 and translation initiation factors. Interestingly, global protein occupancy profiles from human cells show a similar trend toward higher density of protein-bound sites close to the start codon (Baltz et al., 2012).

In our experiments, shifting the GLD-1 binding motif within the *gipc-1* reporter away from the start codon did not abrogate repression. On the contrary, shifted sites were similarly potent in mediating repression as the wild-type GLD-1 site (Figure 2E), indicating that immediate proximity to the start codon is not generally required for 5' UTR-mediated translational repression. Intriguingly, regulation by GLD-1, thus, seems to differ from other mechanistically dissected examples of 5' UTR-mediated regulation (Goossen et al., 1990; Medenbach et al., 2011) in that it does not stringently depend on position. However, the temporal resolution of our assay is limited by the half-life time of the GFP reporter protein and only allows investigation of steady-state protein levels. Thus, we cannot rule out the possibility that shifting the binding site could impact the kinetics of translation initiation.

The only GLD-1 5' UTR target studied in detail so far, *gna-2*, harbors two uORFs and was shown to be both repressed and stabilized by GLD-1 (Lee and Schedl, 2004). Stabilization relies on GLD-1 protecting the transcript from nonsense-mediated decay, likely by inhibiting uORF translation (Lee and Schedl, 2004). The fact that, in this case, a binding site overlaps one uORF stop codon (Jungkamp et al., 2011) suggests that GLD-1 regulation via 5' UTRs might, in general, involve interactions with uORFs. In our study, RNA stability was not majorly affected by the mutation of GLD-1 binding sites (Figures S2A, S2B, S3C, and S3E). Furthermore, the majority of identified targets with 5' UTR sites do not contain uORFs (Jungkamp et al., 2011). Of the GLD-1 targets investigated in the present study, only one mRNA (*mcm-5*) harbored a short uORF (aug guu uaa) that did not overlap with a GLD-1 binding site. We also checked that *trans*-splicing, a frequent processing event in *C. elegans* that leaves mRNAs with a shortened 5' UTR fused to a common spliced leader sequence (Allen et al., 2011), does not generate uORFs for any of these targets. In conclusion, 5' UTR-mediated regulation by GLD-1 seems not to be coupled, generally, to regulation via uORFs.

### 5' UTR Sites Can Mask Functional Sites in 3' UTRs

Although we and others showed that single GLD-1 binding sites are sufficient to mediate repression (Jungkamp et al., 2011; Wright et al., 2011), it is conceivable that multiple sites within a transcript may cooperate in repression and/or use different mechanisms, depending on binding site characteristics. For instance, it has been shown that full translational repression of the *Drosophila msl-2* transcript by the RBP SXL (Sex-lethal) requires both sites in 3' and 5' UTRs (Bashaw and Baker, 1997; Kelley et al., 1997). For many RBPs, functions depend on sequence context and position of binding sites, and interestingly, for a few cases including the SXL-*msl-2* interaction, the mode of regulation has been described to differ depending on whether sites occur in 5' or 3' UTRs (Aeschmann et al., 2017; Beckmann et al., 2005; Kühn, 2015). GLD-1 acts as a translational repressor and additionally stabilizes a subset of targets (Lee and Schedl, 2004; Scheckel et al., 2012). While, so far, there is no evidence for general differences between GLD-1 regulation via 5' and 3' UTR sites, it has been suggested that CDS sites might favor translational repression over transcript stabilization (Brümmer et al., 2013). However, CDS sites performed much worse in explaining target enrichment compared to sites in 5' and 3' UTRs in GLD-1 RNA immunoprecipitation experiments, suggesting that, if at all, they are bound with low affinity (Wright et al., 2011). Consistently, in GLD-1 iPAR-CLIP experiments, only a small percentage of crosslinked sites were found in CDSs, despite the CDS being, on average, the longest transcript unit in *C. elegans* (Jungkamp et al., 2011). Thus, to date, it is still unclear what determines the mode of GLD-1 regulation.

GLD-1 forms dimers, and many targets harbor more than one binding site (Chen et al., 1997; Jungkamp et al., 2011; Teplava et al., 2013). Different models of how GLD-1 dimerization may connect to function have been proposed (Jungkamp et al., 2011; Lee and Schedl, 2001). One intriguing idea is that GLD-1 dimers bind 5' and 3' UTRs of the same transcript, thereby connecting the molecule's ends to form a regulatory loop structure. For QKI (Quaking), the GLD-1 homolog in human, it has been



proposed that dimer-mediated looping might aid in the generation of circular RNAs (Conn et al., 2015). However, in our studies, we found no evidence for cooperativity in translational regulation by multiple sites, arguing against the looping model. On the contrary, we showed that single sites within the “multi-site” GLD-1 targets *mcm-5* and *oma-2* individually confer strong repression and, thus, are capable of masking the presence of additional sites.

Why does GLD-1 dimerize then, and why do targets acquire multiple sites? Dimerization might be involved in other regulatory processes that we could not assess with our assay, such as joint coordination of localization of GLD-1 targets, or it could play a role in the dynamics within RNA granules. Presence of multiple sites, likewise, could be explained in several ways: First, individual sites might act via different mechanisms of translational inhibition, e.g., depend on different co-factors. In fact, target-specific co-factors have also been identified for GLD-1, and GLD-1 mutants have been shown to affect regulation of targets by different degrees (reviewed in Lee and Schedl, 2010). Second, these sites could be involved in other aspects of post-transcriptional regulation that we did not address, e.g., transcript localization. Additionally, as put forward for the so-called “shadow enhancers,” which, in most reporter assays appear redundant, multiple sites could (1) fine-tune repression strength and sharpen boundaries and, more importantly, (2) confer robustness to target regulation (Perry et al., 2010, 2011). While, for instance, under lab conditions, the *oma-2* sites appear redundant, upon environmental perturbations, they might all be necessary to ensure target repression. Interestingly, mTOR signaling has recently been shown to modulate 5′ or 3′ UTR binding by the RBP LARP1, inducing different modes of translational regulation (Hong et al., 2017). Usage and functions of individual GLD-1 binding sites could likewise be modulated by signaling cues.

We note that both GLD-1 in *C. elegans* and SXL in *Drosophila* play important roles in sex determination and germ cell development (reviewed in Lee and Schedl, 2010; Salz and Erickson, 2010). Furthermore, it has been shown that translational control via a 5′ UTR is crucial for proper regulation of meiosis in budding yeast and that many 5′ UTRs increase in length in the course of budding yeast gametogenesis (Berchowitz et al., 2013). It is intriguing to speculate that 5′ UTRs play an ancient regulatory role in the germline and in development.

Taken together, we demonstrated that (1) single 5′ UTR sites can be strong *cis*-regulatory elements and that (2) 5′ and 3′ UTR sites repress translation independently for the targets investigated here. In particular, we demonstrate that sites in 5′ UTRs can mask functional sites in 3′ UTRs. We suggest that 5′ UTRs should be tested systematically for regulatory activities and argue that, to understand *in vivo* regulation, both endogenous 5′ and 3′ UTRs should be taken into consideration.

## EXPERIMENTAL PROCEDURES

### *C. elegans* Maintenance

*C. elegans* strains were cultivated using standard procedures (Brenner, 1974). Unless indicated otherwise, worms were maintained at 24°C on *E. coli* OP50-seeded nematode growth medium (NGM) plates (Stiernagle, 2006). MosSCI injection strains EG4322 and EG6699 were kept at 16°C for maintenance.

### Cloning of Reporter Constructs and Generation of Transgenic Lines

To analyze GLD-1 regulation mediated via 5′ and 3′ UTR binding sites, stable transgenic *in vivo* reporters were generated using the MosSCI technique (Frøkjær-Jensen et al., 2008). Sequences of target 5′ and 3′ UTRs were retrieved from transcriptome annotations of the modENCODE Consortium (Gerstein et al., 2010). Reporter constructs were generated by conventional restriction-ligation cloning. All final constructs contained the *gld-1* promoter, the target 5′ UTR, the GFP::H2B CDS, and the target 3′ UTR within the backbone of the MosSCI vector pCFJ151. To preserve sequence context upstream of the start codon, we fused the target 5′ UTR to the GFP::H2B CDS by PCR fusion, as described previously (Hobert, 2002). Mutations and shifts of GLD-1 binding motifs were generated either during PCR fusion or by site-directed mutagenesis according to the QuikChange protocol (Agilent), using PfuUltra II Fusion HS DNA polymerase on final reporter constructs. Plasmids were isolated using the ZR Plasmid Miniprep kit (Zymo Research). All constructs were validated by analytical restriction digest (HindIII) and Sanger sequencing (LGC Genomics). Oligonucleotides (Eurofins MWG Operon) used for amplification, mutagenesis, and sequencing are listed in the Supplemental Experimental Procedures.

Reporter constructs were injected into gonads of young adult worms of strain EG4322 (for *oma-2* constructs: strain EG6699), together with co-injection plasmids (Frøkjær-Jensen et al., 2012). Transgene-positive worms were identified by wild-type movement and the absence of co-marker *mCherry* expression. Transgene integrity was validated by presence of GFP expression and by single-worm PCR (<http://genetics.wustl.edu/tslab/protocols/genomic-stuff/single-worm-pcr/>), using transgene- and insertion site-specific primers (Supplemental Experimental Procedures).

### Imaging and Relative Quantification of Fluorescence

Young adult worms (with  $\leq 7$  eggs) were mounted by picking clean worms into a drop of 1 mM levamisole in M9 (22 mM KH<sub>2</sub>PO<sub>4</sub>, 42 mM Na<sub>2</sub>HPO<sub>4</sub>, 86 mM NaCl, 1 mM MgSO<sub>4</sub>) on a 2% agarose pad. Pictures of GFP expression in gonads were taken on an inverted fluorescence microscope (BZ-X710; Keyence) with a 40× objective (Plan Apo  $\lambda$  40×/0.95; gain disabled; 2 × 2 binning). Reporter strains for the same gene were monitored using same exposure times. Pictures were processed and analyzed using ImageJ in an identical manner. Germline GFP fluorescence was quantified by measuring pixel intensity profiles with ImageJ along a segmented line (line width: 35 pixels) from the distal tip to the bend of each gonad arm, as described previously (Wright et al., 2011). Background mean gray values were subtracted, and values were binned into 30 bins and normalized to the mean of the first 5% for each gonad to control for GLD-1-independent expression differences. Averages and SEM were calculated for all gonads analyzed per reporter strain (custom Python script). Estimations of repression for the different reporter constructs are described in the Supplemental Experimental Procedures.

### Relative Quantification of Reporter RNA Levels

Reporter strains were synchronized by bleaching (Stiernagle, 2006) and were grown at 24°C on *E. coli* OP50-seeded NGM plates until they reached young adulthood (typically, approximately 45 hr). Worms were harvested, washed in 0.1 M NaCl, and resuspended in 1 mL TRIzol reagent (Thermo Fisher Scientific). After homogenization with SILbeads (Sigmund Lindner) in a tissue lyser (2×10 s; Precellys 24 homogenizer; Bertin Technologies), RNA was extracted according to the TRIzol protocol. One microgram of RNA was DNase treated (RQ1 DNase, Promega), and one-fifth was used as input for reverse transcription with random hexamer primers and Maxima H Minus Reverse Transcriptase (Thermo Fisher Scientific). qPCR was done with Maxima SYBR Green/ROX qPCR Master Mix (Thermo Fisher Scientific). Reporter RNA levels were normalized to C<sub>t</sub> averages of three reference genes (*pmp-3*, *tba-1*, and *Y45F10D.4*; Zhang et al., 2012). Corresponding oligonucleotides (Eurofins MWG Operon) are listed in the Supplemental Experimental Procedures.

### Statistical Methods

All statistical details including sample sizes and displayed data are given in the figure legends. Data were not randomized. All quantification plots were generated with the R software package.

## SUPPLEMENTAL INFORMATION

Supplemental Information includes Supplemental Experimental Procedures and four figures and can be found with this article online at <https://doi.org/10.1016/j.celrep.2018.02.094>.

## ACKNOWLEDGMENTS

We are indebted to Christian Frøkjær-Jensen for his practical guidance in setting up MosSCI injections in our lab. K.T. thanks all members of the Rajewsky lab for discussions and support, particularly Benedikt Obermayer, Filippos Klironomos, and Sebastian Mackowiak for advice concerning computation, Lena von Oertzen for technical assistance, and Christine Kocks and Anna-Carina Rajewsky for discussions. MosSCI injection strains were provided by the *Caenorhabditis* Genetics Center (CGC), which is funded by the NIH Office of Research Infrastructure Programs (P40 OD010440). This work was partially supported by the e:Bio program of the Federal Ministry of Education (0316164B), the Human Frontier Science Program (RGP0009/2014), and grant BIH CRG 2a TP7.

## AUTHOR CONTRIBUTIONS

N.R. conceived and supervised the project. K.T. designed and performed all experiments and analyzed the data. M.H. assisted by performing all MosSCI injections. K.T. and N.R. interpreted the data and wrote the paper.

## DECLARATION OF INTERESTS

The authors declare no competing interests.

Received: September 24, 2017

Revised: November 20, 2017

Accepted: February 23, 2018

Published: March 20, 2018

## REFERENCES

- Aeschimann, F., Kumari, P., Bartake, H., Gaidatzis, D., Xu, L., Ciosk, R., and Großhans, H. (2017). LIN41 post-transcriptionally silences mRNAs by two distinct and position-dependent mechanisms. *Mol. Cell* **65**, 476–489.
- Allen, M.A., Hillier, L.W., Waterston, R.H., and Blumenthal, T. (2011). A global analysis of *C. elegans* trans-splicing. *Genome Res.* **21**, 255–264.
- Baltz, A.G., Munschauer, M., Schwanhäusser, B., Vasile, A., Murakawa, Y., Schueler, M., Youngs, N., Penfold-Brown, D., Drew, K., Milek, M., et al. (2012). The mRNA-bound proteome and its global occupancy profile on protein-coding transcripts. *Mol. Cell* **46**, 674–690.
- Bashaw, G.J., and Baker, B.S. (1997). The regulation of the *Drosophila msl-2* gene reveals a function for *Sex-lethal* in translational control. *Cell* **89**, 789–798.
- Beckmann, K., Grskovic, M., Gebauer, F., and Hentze, M.W. (2005). A dual inhibitory mechanism restricts *msl-2* mRNA translation for dosage compensation in *Drosophila*. *Cell* **122**, 529–540.
- Berchowitz, L.E., Gajadhar, A.S., van Werven, F.J., De Rosa, A.A., Samoylova, M.L., Brar, G.A., Xu, Y., Xiao, C., Futcher, B., Weissman, J.S., et al. (2013). A developmentally regulated translational control pathway establishes the meiotic chromosome segregation pattern. *Genes Dev.* **27**, 2147–2163.
- Blin, K., Dieterich, C., Wurmus, R., Rajewsky, N., Landthaler, M., and Akalin, A. (2015). DoRiNA 2.0—upgrading the doRiNA database of RNA interactions in post-transcriptional regulation. *Nucleic Acids Res.* **43**, D160–D167.
- Brenner, S. (1974). The genetics of *Caenorhabditis elegans*. *Genetics* **77**, 71–94.
- Brümmer, A., Kishore, S., Subasic, D., Hengartner, M., and Zavolan, M. (2013). Modeling the binding specificity of the RNA-binding protein GLD-1 suggests a function of coding region-located sites in translational repression. *RNA* **19**, 1317–1326.
- Chen, T., Damaj, B.B., Herrera, C., Lasko, P., and Richard, S. (1997). Self-association of the single-KH-domain family members Sam68, GRP33, GLD-1, and Qk1: role of the KH domain. *Mol. Cell. Biol.* **17**, 5707–5718.
- Conn, S.J., Pillman, K.A., Toubia, J., Conn, V.M., Salamanidis, M., Phillips, C.A., Roslan, S., Schreiber, A.W., Gregory, P.A., and Goodall, G.J. (2015). The RNA binding protein quaking regulates formation of circRNAs. *Cell* **160**, 1125–1134.
- Farley, B.M., and Ryder, S.P. (2012). POS-1 and GLD-1 repress *glp-1* translation through a conserved binding-site cluster. *Mol. Biol. Cell* **23**, 4473–4483.
- Frøkjær-Jensen, C., Davis, M.W., Hopkins, C.E., Newman, B.J., Thummel, J.M., Olesen, S.P., Grunnet, M., and Jorgensen, E.M. (2008). Single-copy insertion of transgenes in *Caenorhabditis elegans*. *Nat. Genet.* **40**, 1375–1383.
- Frøkjær-Jensen, C., Davis, M.W., Ailion, M., and Jorgensen, E.M. (2012). Improved Mos1-mediated transgenesis in *C. elegans*. *Nat. Methods* **9**, 117–118.
- Gerstein, M.B., Lu, Z.J., Van Nostrand, E.L., Cheng, C., Arshinoff, B.I., Liu, T., Yip, K.Y., Robilotto, R., Rechtsteiner, A., Ikegami, K., et al.; modENCODE Consortium (2010). Integrative analysis of the *Caenorhabditis elegans* genome by the modENCODE project. *Science* **330**, 1775–1787.
- Goossen, B., Caughman, S.W., Harford, J.B., Klausner, R.D., and Hentze, M.W. (1990). Translational repression by a complex between the iron-responsive element of ferritin mRNA and its specific cytoplasmic binding protein is position-dependent in vivo. *EMBO J.* **9**, 4127–4133.
- Graindorge, A., Carré, C., and Gebauer, F. (2013). Sex-lethal promotes nuclear retention of *msl2* mRNA via interactions with the STAR protein HOW. *Genes Dev.* **27**, 1421–1433.
- Gray, N.K., and Hentze, M.W. (1994). Iron regulatory protein prevents binding of the 43S translation pre-initiation complex to ferritin and eALAS mRNAs. *EMBO J.* **13**, 3882–3891.
- Hobert, O. (2002). PCR fusion-based approach to create reporter gene constructs for expression analysis in transgenic *C. elegans*. *Biotechniques* **32**, 728–730.
- Hong, S., Freeberg, M.A., Han, T., Kamath, A., Yao, Y., Fukuda, T., Suzuki, T., Kim, J.K., and Inoki, K. (2017). LARP1 functions as a molecular switch for mTORC1-mediated translation of an essential class of mRNAs. *eLife* **6**, e25237.
- Jones, A.R., and Schedl, T. (1995). Mutations in *glid-1*, a female germ cell-specific tumor suppressor gene in *Caenorhabditis elegans*, affect a conserved domain also found in Src-associated protein Sam68. *Genes Dev.* **9**, 1491–1504.
- Jungkamp, A.C., Stoeckius, M., Mecnas, D., Grün, D., Mastrobuoni, G., Kempa, S., and Rajewsky, N. (2011). In vivo and transcriptome-wide identification of RNA binding protein target sites. *Mol. Cell* **44**, 828–840.
- Kelley, R.L., Wang, J., Bell, L., and Kuroda, M.I. (1997). Sex lethal controls dosage compensation in *Drosophila* by a non-splicing mechanism. *Nature* **387**, 195–199.
- Kent, W.J., Sugnet, C.W., Furey, T.S., Roskin, K.M., Pringle, T.H., Zahler, A.M., and Haussler, D. (2002). The human genome browser at UCSC. *Genome Res.* **12**, 996–1006.
- Kühn, L.C. (2015). Iron regulatory proteins and their role in controlling iron metabolism. *Metallomics* **7**, 232–243.
- Lee, M.H., and Schedl, T. (2001). Identification of in vivo mRNA targets of GLD-1, a maxi-KH motif containing protein required for *C. elegans* germ cell development. *Genes Dev.* **15**, 2408–2420.
- Lee, M.H., and Schedl, T. (2004). Translation repression by GLD-1 protects its mRNA targets from nonsense-mediated mRNA decay in *C. elegans*. *Genes Dev.* **18**, 1047–1059.
- Lee, M.H., and Schedl, T. (2010). *C. elegans* star proteins, GLD-1 and ASD-2, regulate specific RNA targets to control development. *Adv. Exp. Med. Biol.* **693**, 106–122.
- Mangone, M., Manoharan, A.P., Thierry-Mieg, D., Thierry-Mieg, J., Han, T., Mackowiak, S.D., Mis, E., Zegar, C., Gutwein, M.R., Khivansara, V., et al. (2010). The landscape of *C. elegans* 3' UTRs. *Science* **329**, 432–435.

- Medenbach, J., Seiler, M., and Hentze, M.W. (2011). Translational control via protein-regulated upstream open reading frames. *Cell* **145**, 902–913.
- Merritt, C., Rasoloson, D., Ko, D., and Seydoux, G. (2008). 3' UTRs are the primary regulators of gene expression in the *C. elegans* germline. *Curr. Biol.* **18**, 1476–1482.
- Perry, M.W., Boettiger, A.N., Bothma, J.P., and Levine, M. (2010). Shadow enhancers foster robustness of *Drosophila* gastrulation. *Curr. Biol.* **20**, 1562–1567.
- Perry, M.W., Boettiger, A.N., and Levine, M. (2011). Multiple enhancers ensure precision of gap gene-expression patterns in the *Drosophila* embryo. *Proc. Natl. Acad. Sci. USA* **108**, 13570–13575.
- Pesole, G., Mignone, F., Gissi, C., Grillo, G., Licciulli, F., and Liuni, S. (2001). Structural and functional features of eukaryotic mRNA untranslated regions. *Gene* **276**, 73–81.
- Prasad, A., Porter, D.F., Kroll-Conner, P.L., Mohanty, I., Ryan, A.R., Crittenden, S.L., Wickens, M., and Kimble, J. (2016). The PUF binding landscape in metazoan germ cells. *RNA* **22**, 1026–1043.
- Rybak-Wolf, A., Jens, M., Murakawa, Y., Herzog, M., Landthaler, M., and Rajewsky, N. (2014). A variety of dicer substrates in human and *C. elegans*. *Cell* **159**, 1153–1167.
- Salz, H.K., and Erickson, J.W. (2010). Sex determination in *Drosophila*: The view from the top. *Fly (Austin)* **4**, 60–70.
- Scheckel, C., Gaidatzis, D., Wright, J.E., and Ciosk, R. (2012). Genome-wide analysis of GLD-1-mediated mRNA regulation suggests a role in mRNA storage. *PLoS Genet.* **8**, e1002742.
- Sood, P., Krek, A., Zavolan, M., Macino, G., and Rajewsky, N. (2006). Cell-type-specific signatures of microRNAs on target mRNA expression. *Proc. Natl. Acad. Sci. USA* **103**, 2746–2751.
- Stiernagle, T. (2006). Maintenance of *C. elegans*. *WormBook*, 1–11.
- Teplova, M., Hafner, M., Teplov, D., Essig, K., Tuschl, T., and Patel, D.J. (2013). Structure-function studies of STAR family Quaking proteins bound to their in vivo RNA target sites. *Genes Dev.* **27**, 928–940.
- Van Nostrand, E.L., Freese, P., Pratt, G.A., Wang, X., Wei, X., Xiao, R., Blue, S.M., Dominguez, D., Cody, N.A.L., Olson, S., et al. (2017). A large-scale binding and functional map of human RNA binding proteins. *bioRxiv*. <https://doi.org/10.1101/179648>.
- Vernet, C., and Artzt, K. (1997). STAR, a gene family involved in signal transduction and activation of RNA. *Trends Genet.* **13**, 479–484.
- Wright, J.E., Gaidatzis, D., Senften, M., Farley, B.M., Westhof, E., Ryder, S.P., and Ciosk, R. (2011). A quantitative RNA code for mRNA target selection by the germline fate determinant GLD-1. *EMBO J.* **30**, 533–545.
- Zhang, Y., Chen, D., Smith, M.A., Zhang, B., and Pan, X. (2012). Selection of reliable reference genes in *Caenorhabditis elegans* for analysis of nanotoxicity. *PLoS ONE* **7**, e31849.