Supplemental Information
Molecular Cell, *Volume 44*

In Vivo and Transcriptome-wide Identification of RNA-Binding Protein Target Sites

Anna-Carina Jungkamp, Marlon Stoeckius, Desirea Mecenas, Dominic Grün, Guido Mastrobuoni, Stefan Kempa, Nikolaus Rajewsky

Figure S1

(Related to Figure 1)

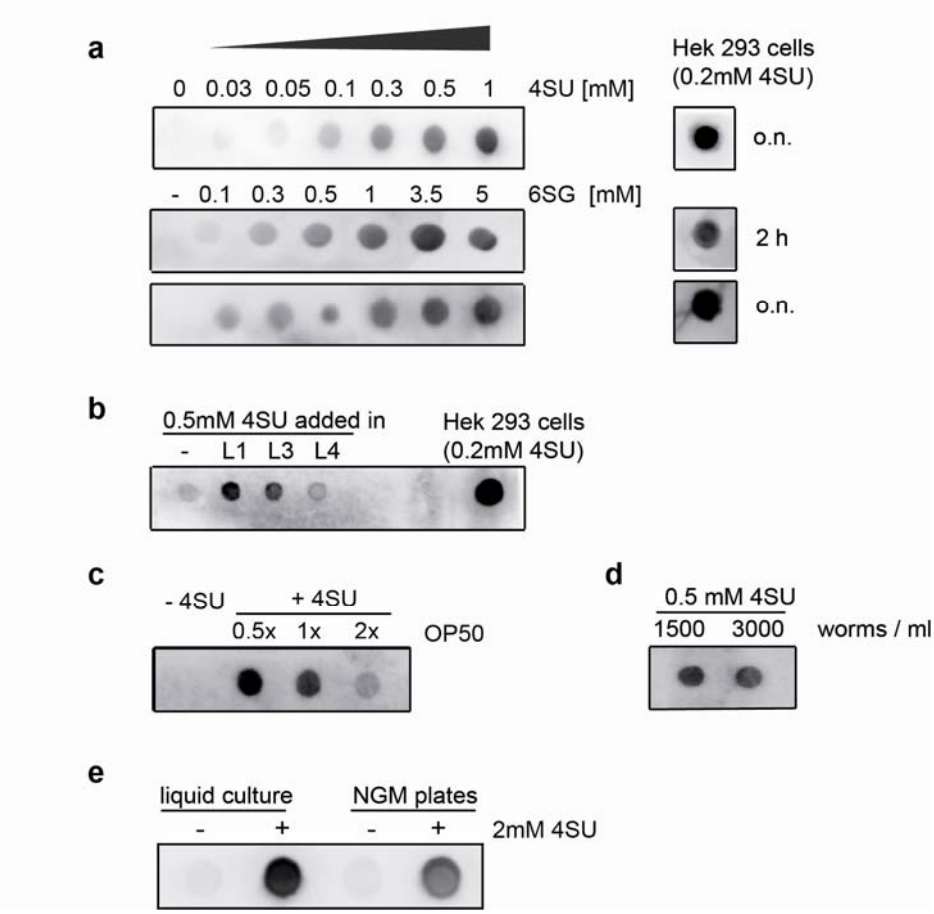

Figure S1. Labeling of Nascent RNA with Photoreactive Nucleosides in *C. elegans*

(A-E) Dotblots with thiol-specifically biotinylated total RNA extracted from adult worms labeled with 4-thiouridine (4SU) or 6-thioguanosine (6SG).

(A) RNA extracted from worms labeled with different concentrations of 4SU or 6SG. Right panels: equal amounts of RNA extracted from HEK293 cells labeled with 0.2mM 4SU for 2 hours or overnight (o.n.) were spottet on the same membrane.

(B) 4SU was added to worms in liquid culture at larval stages L1, L3 or L4. For comparison, equal amounts of RNA extracted from HEK293 cells labeled with 0.2mM 4SU were spottet on the same membrane.

(C) Labeling efficiency depends on the amount of food (*E.coli* OP50) in the liquid culture. 1x: 1ml bacteria OD 2.5/ 1000 worms.

(D) The number of worms per ml liquid culture does not influence labeling efficiency.

(E) Labeling of *C. elegans* RNA in liquid culture compared to labeling on NGM plates. For labeling on plates, 4SU concentrations were calculated according to the volume of NGM plates (40 ml for a 15 cm dish).

Figure S2

(Related to Figure 5)

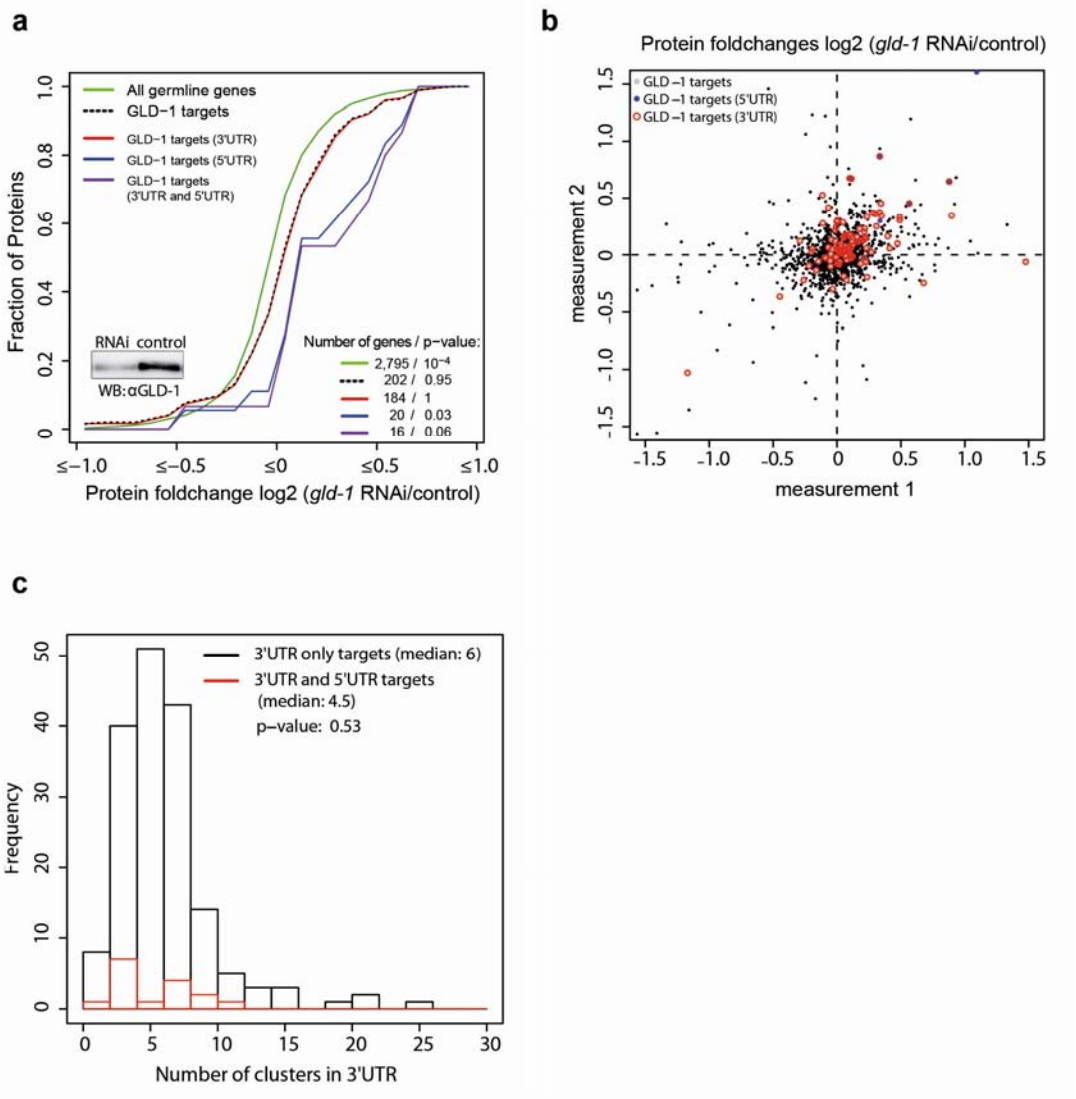

Figure S2. Targets with 5' UTR Sites Show Stronger Derepression upon GLD-1 Knockdown

(A) Cumulative fractions of fold-changes in protein expression after GLD-1 knockdown. Protein fold changes for 3,874 genes including 202 of the germline-expressed, reproducible iPAR-CLIP targets were measured by using SILAC in *C. elegans*. Compared to all targets that contain 3' UTR sites, targets with 5' UTR target sites show a significantly stronger de-repression upon GLD-1 knockdown (p value <0.03).

(B) Protein fold-changes of GLD-1 targets in GLD-1 knockdown versus control samples in two measurements of the same worm samples.

(C) Targets that harbor 5' UTR target sites have on average fewer 3' UTR clusters compared to targets that do not have 5' UTR target sites.

Figure S3

(Related to Figure 7)

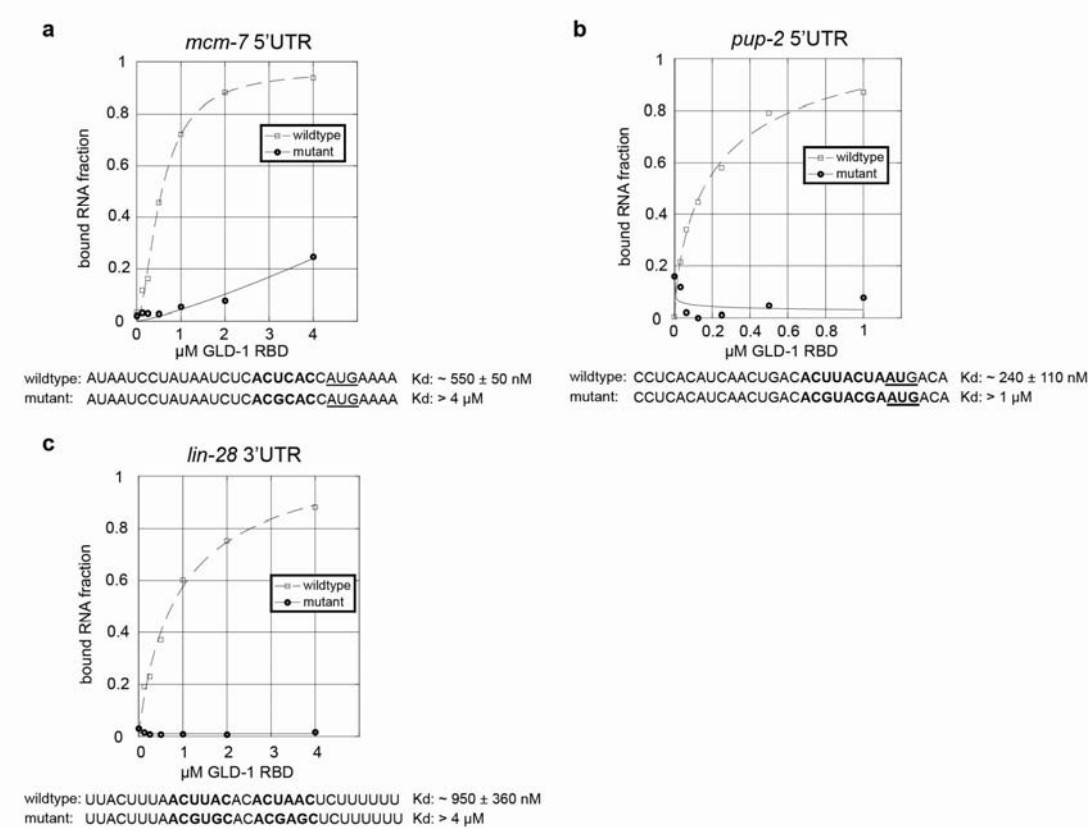

Figure S3. Affinity of GLD-1 for Identified 3' UTR and 5' UTR Binding Sites

Results for direct titration assays shown in Figure 7b. Fractions of bound and free RNA were quantified using Image Gauge (version 4.0) and equilibrium dissociation constants were derived from a non-linear least squares fit (KaleidaGraph software, see Supplemental Experimental Procedures).

Figure S4

(Related to Figure 7)

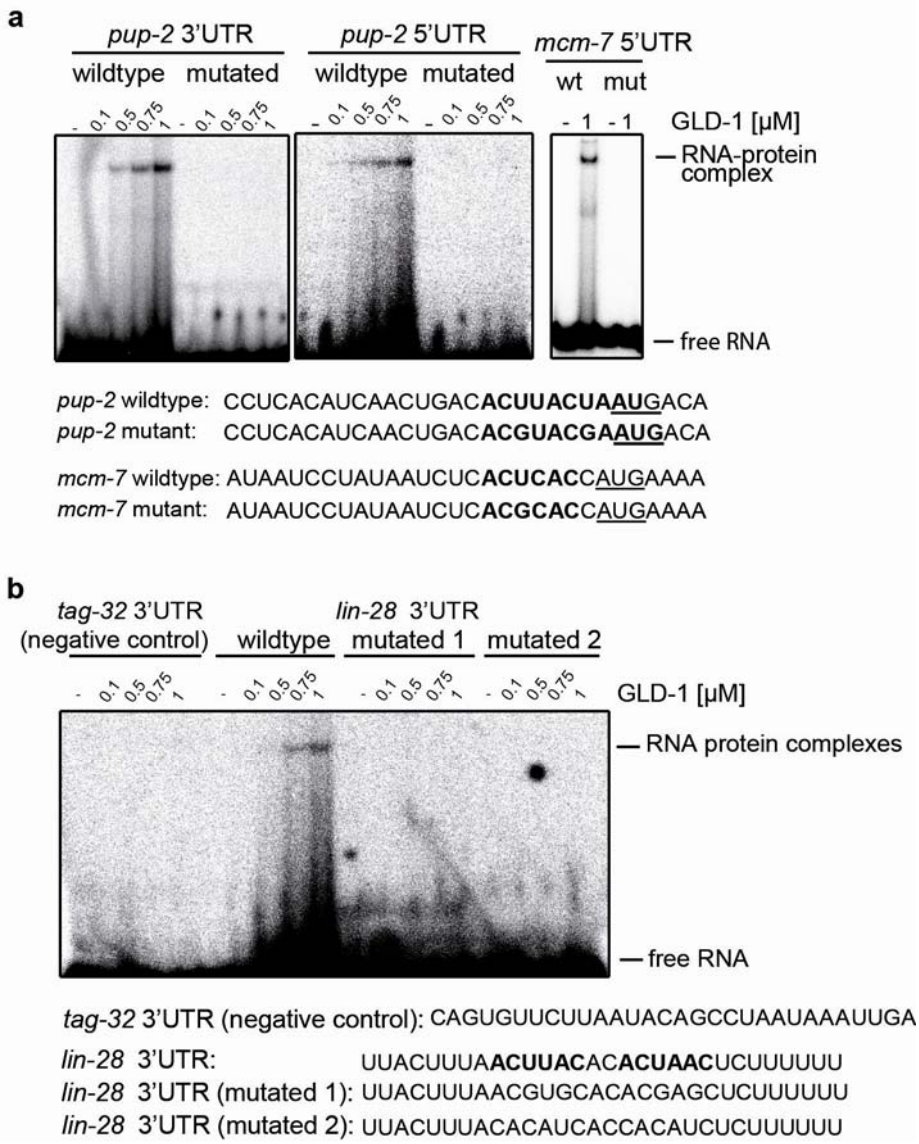

Figure S4. Immunoprecipitated GLD-1 Binds to Identified 3' UTR and 5' UTR Target Sites In Vitro

Gelshift assays demonstrating binding of GLD-1 to identified target sites *in vitro*, depending on the GLD-1 binding motif. 1nM radiolabeled RNA was incubated with increasing concentrations of immunoprecipitated GLD-1::GFP::FLAG protein. A sequence of the *tag-32* 3' UTR which was not identified as a GLD-1 target by iPAR-CLIP served as a negative control.

Figure S5

(Related to Figure 7)

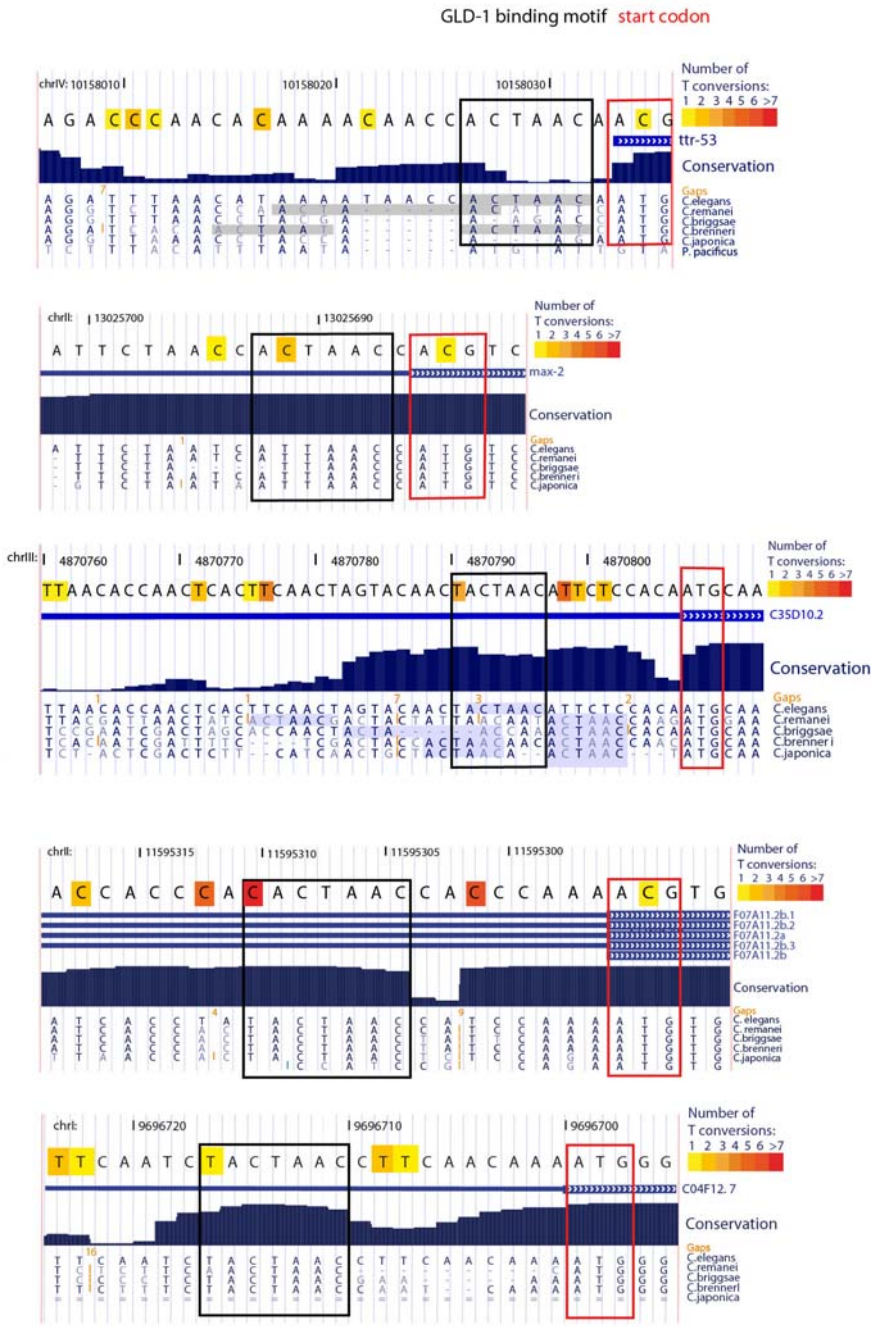

Figure S5. GLD-1 Binds Close to the Start Codon

Examples of 5' UTR binding sites showing highly conserved GLD-1 binding motifs directly upstream of start codons. The number of T conversions observed in one iPAR-CLIP experiment is color-coded.

Figure S6

(Related to Figure 7)

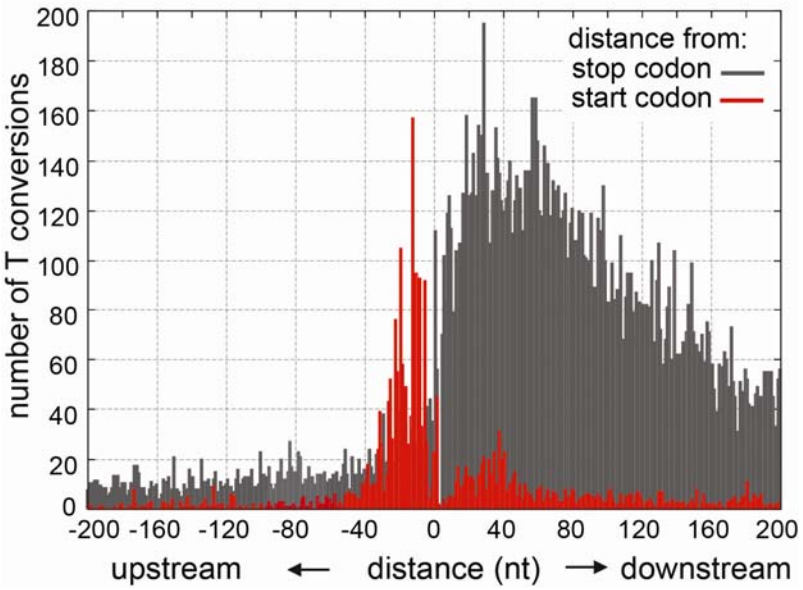

Figure S6. GLD-1 Binds near the Start Codon
Histogram of the distance of T conversions from stop- and start-codons. T conversions pile up nearby start codons.

Figure S7

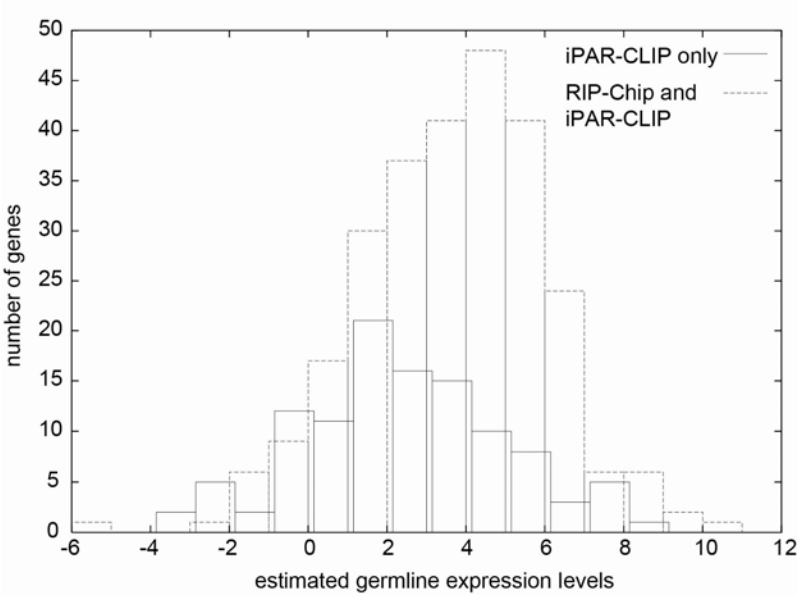

Figure S7. RIP-CHIP and iPAR-CLIP Agree on Highly Expressed Targets

Histograms of germline expression levels (log units) (see Supplemental Experimental Procedures) for 439 iPAR-CLIP targets and 948 targets identified by RIP-Chip (Wright et al., 2011) with a IP cutoff of 3 fold. ~ 70% of the identified iPAR-CLIP targets can be confirmed by RIP-CHIP. Differences come mostly from genes which are expressed at low or mid-range levels.

Table S1. Known GLD-1 Targets

mRNA targets	Sequence ID	references	#conversions (#clusters) (cut-off: 2 conversions)			
			4SU(1)	4SU(2)	4SU(3)	6SG
tra-2	C15F1.3	(Jan et al., 1999)	-	-	3(1)	6(2)
rme-2	T11F8.3	(Lee and Schedl, 2001)	36(3)	42(3)	85(4)	76(3)
gna-2	T23G11.2	(Lee and Schedl, 2004)	16(2)	23(2)	24(2)	32(2)
oma-1	C09G9.6	(Lee and Schedl, 2004)	14(2)	24(5)	40(5)	38(4)
oma-2	ZC513.6	(Lee and Schedl, 2004)	22(3)	23(3)	24(3)	121(3)
mes-3	F54C1.3	(Xu et al., 2001)	2(1)	3(1)	4(2)	3(1)
pal-1	C38D4.6	(Mootz et al., 2004)	6(3)	5(1)	15(3)	15(5)
glp-1	F02A9.6	(Marin and Evans, 2003)	7(2)	6(3)	17(5)	23(3)
cep-1	F52B5.5	(Schumacher et al., 2005)	2(1)	4(1)	11(1)	3(1)
cye-1	C37A2.4	(Biedermann et al., 2009)	25(4)	34(3)	59(5)	21(3)
puf-5	F54C9.8	(Lee and Schedl, 2001, 2010)	4(2)	3(1)	12(2)	42(3)
lin-45	Y73B6A.5	(Lee and Schedl, 2001, 2004, 2010)	7(2)	12(2)	24(4)	7(3)
bir-1	T27F2.3	(Wright et al., 2011)	5(1)	10(1)	17(2)	15(1)
rmd-1	T05G5.7	(Lee and Schedl, 2001; Wright et al., 2011)	10(2)	12(2)	34(2)	30(1)
dpf-3	K02F2.1	(Wright et al., 2011)	18(2)	30(2)	38(3)	21(3)
C01G8.1	C01G8.1	(Wright et al., 2011)	8(1)	9(1)	16(1)	16(1)
C36B1.11	C36B1.11	(Wright et al., 2011)	5(1)	9(2)	9(2)	15(1)
F59A3.4	F59A3.4	(Wright et al., 2011)	3(1)	7(1)	6(2)	-

Table S1. Known GLD-1 Targets
The table lists 18 previously identified GLD-1 targets compiled from the literature that have been studied in detail and served as positive controls for iPAR-CLIP.

Table S2. Suggested GLD-1 Targets

mRNA targets	Sequence ID	references	#conversions (#clusters) (cut-off: 2 conversions)			
			4SU(1)	4SU(2)	4SU(3)	6SG
cpg-1	C07G2.1	(Lee and Schedl, 2001)	42(1)	39(1)	53(1)	178(2)
cpg-2	B0280.5	(Lee and Schedl, 2001)	58(3)	69(4)	86(5)	224(6)
pie-1	Y49E10.14	(Ryder et al., 2004)	3(1)	5(1)	11(3)	16(2)
H02I12.5	H02I12.5	(Lee and Schedl, 2001)	2(1)	2(1)	4(1)	3(1)
puf-6	F18A11.1	(Lee and Schedl, 2001)	-	3(1)	5(1)	2(1)
puf-7	F18A11.1	(Lee and Schedl, 2001)	-	-	2(1)	2(1)
puf-10		(Lee and Schedl, 2001)	-	-	-	-
Y75B12B.1	Y75B12B.1	(Lee and Schedl, 2001)	-	3(1)	4(1)	14(1)
egg-1	B0244.8	(Lee and Schedl, 2001)	-	7(1)	7(1)	38(3)
exo-3	R09B3.1	(Lee and Schedl, 2001)	-	-	-	3(1)
tra-1	Y47D3A.6	(Ryder et al., 2004)	-	-	2(1)	2(1)
mes-4	Y2H9A.1	(Ryder et al., 2004)	-	-	-	2(1)
gln-5	F26D10.10	(Lee and Schedl, 2001)	-	-	-	7(1)
spn-4	ZC404.8	(Mootz et al., 2004)	20(4)	18(3)	38(4)	101(6)
mex-3	F53G12.5	(Mootz et al., 2004)	8(3)	20(5)	43(8)	26(4)
mex-5	W02A2.7	(Mootz et al., 2004)	11(2)	17(4)	29(5)	36(5)
mex-6	AH6.5	(Mootz et al., 2004)	8(2)	9(1)	32(4)	24(1)

Table S2. Suggested GLD-1 Targets

The table lists genes that were suggested, but not validated, GLD-1 targets and the respective number of conversions and clusters in our iPAR-CLIP experiments.

Table S3. Top 30 Targets (4SU iPAR-CLIP 1)

mRNA targets	# T conversions
sip-1	116
cpg-2	58
cpg-1	42
rme-2	36
gld-1	33
pup-2	31
ima-2	28
cbd-1	26
cye-1	25
plk-3	22
oma-2	22
R02F2.1	20
spn-4	20
dpf-3	18
CE16308	18
CE40981	17
pgl-1	17
CE00867	17
gln-6	17
unc-66	17
rskn-1	16
gna-2	16
mesp-1	16
tdc-1	16
act4	15
F27C8.6.2	15
sqd-1	15
pos-1	14
oma-1	14
pqn-45	14

Table S3. Top 30 GLD-1 Targets
The table lists the top 30 targets of GLD-1 obtained in 4SU iPAR-CLIP experiment 1. Highlighted in grey: previously suggested or validated GLD-1 targets.

Table S5. Top-Enriched GO-Terms of the Identified 439 Targets

Term	Count	%	P-value	Benjamini corrected p-value
cell division	37	8.4	$2.7 \cdot 10^{-10}$	$1.9 \cdot 10^{-7}$
cytokinesis	28	6.4	$1.3 \cdot 10^{-8}$	$4.5 \cdot 10^{-6}$
cell fate commitment	15	3.4	$3.1 \cdot 10^{-7}$	$7.4 \cdot 10^{-5}$
embryonic development ending in birth or egg hatching	165	37.5	$8.1 \cdot 10^{-7}$	$1.4 \cdot 10^{-4}$
embryonic pattern specification	9	2	$2.0 \cdot 10^{-5}$	$2.9 \cdot 10^{-3}$
cell cycle	40	9.1	$3.1 \cdot 10^{-5}$	$3.6 \cdot 10^{-3}$
cell fate specification	11	2.5	$3.9 \cdot 10^{-5}$	$4.0 \cdot 10^{-3}$
pattern specification process	10	2.3	$5.5 \cdot 10^{-5}$	$4.8 \cdot 10^{-3}$
multicellular organism reproduction	36	8.2	$7.4 \cdot 10^{-5}$	$5.8 \cdot 10^{-3}$
reproductive process in a multicellular organism	36	8.2	$7.4 \cdot 10^{-5}$	$5.8 \cdot 10^{-3}$
reproductive developmental process	65	14.8	$1.5 \cdot 10^{-4}$	$1.1 \cdot 10^{-2}$
DNA metabolic process	22	5.0	$2.7 \cdot 10^{-4}$	$1.7 \cdot 10^{-2}$
DNA replication	13	3.0	$3.2 \cdot 10^{-4}$	$1.9 \cdot 10^{-2}$
cell cycle process	34	7.7	$6.1 \cdot 10^{-4}$	$3.3 \cdot 10^{-2}$
gamete generation	21	4.8	$7.3 \cdot 10^{-4}$	$3.6 \cdot 10^{-2}$
oogenesis	15	3.4	$7.5 \cdot 10^{-4}$	$3.5 \cdot 10^{-2}$
female gamete generation	15	3.4	$8.6 \cdot 10^{-4}$	$3.7 \cdot 10^{-2}$
sexual reproduction	25	5.7	$8.9 \cdot 10^{-4}$	$3.6 \cdot 10^{-2}$

Table S5. Top-Enriched GO-Terms of the Identified 439 Targets
The table lists the top-scoring categories of the GO-Term-Analysis for the 439 candidates. All germline-expressed genes were used as background.

Supplemental Experimental Procedures

I. Supplemental Experimental Procedures

Maintenance of *C. elegans*

Strains were maintained using standard methods *(Stiernagle, 2006)* on OP50 seeded NGM plates at 20°C unless otherwise noted. Liquid culture of *C. elegans* was modified from *(Stiernagle, 2006)*. Worms were cultivated in S-Basal (100mM NaCl, 6mM K_2HPO_4, 44mM KH_2PO_4, 5mg/L Cholesterol) supplemented with 3mM $MgCl_2$, 3mM CaCl and 10mM K-Citrate (pH6) on a rotary shaker at 180 rpm. The liquid culture medium (S-Medium) had a pH of approximately 6 and an osmolarity of around 370 mOsmol/kg.

Strains

Wildtype *C. elegans* (N2 Bristol) were used for testing the incorporation of photoreactive nucleosides. The BS1080 (*ozIs5* [GLD-1::GFP/FLAG, pMMO16 (unc-119(+))] (I)) strain used for *in vivo* PAR-CLIP experiments expresses a rescuing GLD-1::GFP::FLAG fusion protein and was kindly provided by Tim Schedl. BS1080 and *glp-4 (bn2ts)* strains (Beanan and Strome, 1992) were used for mRNA sequencing. The RNAi hypersensitive *eri-1(mg366)* strain (Kennedy et al., 2004) was used for RNAi experiments. The EG4322 strain was used for integration of reporter constructs according to the MosSCI direct insertion protocol (Frokjaer-Jensen et al., 2008).

RNAi

gld-1 RNAi was performed by feeding as described previously using the *gld-1* clone from the Ahringer RNAi library (Fraser et al., 2000; Kamath et al., 2003). The empty L4440 vector was used as the negative RNAi control. For proteomics, L1 stage *eri-1*(mg366) worms were plated onto RNAi plates and grown at 20°C for three days. For reporter strains, L1 stages of the respective lines were plated onto RNAi plates and grown at 25°C.

Labeling of *C. elegans* with photoreactive nucleosides

Arrested L1 worms were typically grown in liquid culture supplemented with 2mM 4-thiouridine (4SU) or 6-thioguanosine (6SG) and harvested at the adult stage. Liquid cultures usually contained 3,000 synchronously growing worms per ml and 1ml *E.coli* OP50 (OD_{600} 2.3) per 1,000 worms. Alternatively, photoreactive nucleosides were mixed with *E.coli* OP50 and added on NGM plates. For labeling on plates, 4SU concentrations were calculated according to the volume of NGM agar (add e.g. 80µl [1mM] 4SU mixed with bacteria on a 15 cm dish that contains 40 ml NGM agar). HEK293 cells were labeled as described previously (Hafner et al., 2010).

In vivo PAR-CLIP

Synchronized L1 worms were grown in liquid culture supplemented with 2mM 4SU or 6SG. 250,000 worms were sufficient for one iPAR-CLIP experiment. Living adult worms were transferred to NGM plates and crosslinked on ice using a Stratalinker (Stratagene) with customized 365nm UV-lamps (energy setting: 2J/cm^2). Worms were lysed on ice by douncing in NP40 lysis buffer (50 mM HEPES-K pH 7.5, 150 mM KCl, 2 mM EDTA, 0.5% (v/v) NP-40, 0.5 mM DTT, protease inhibitor cocktail (Roche)). Cleared lysates were treated with RNase T1 (Fermentas) (final concentration 1U/µl) for 15 min at 22ºC. GLD-1::GFP::FLAG fusion proteins were immunoprecipitated for 1h at 4ºC using anti-FLAG antibody (Sigma, F3165) coupled to Protein G magnetic beads (Invitrogen). For one iPAR-CLIP experiment (1ml cleared lysate obtained from 250,000 worms), 300µl beads and 150µg antibody were used. Immunoprecipitates were treated with RNase T1 (100U/µl) for exactly 12 min at 22 ºC. Subsequently, PAR-CLIP was carried out as described previously (Hafner et al., 2010). cDNA libraries were sequenced on a Genome Analyzer II (Illumina).

Isolation of labeled RNA and dot-blot assays

After labeling with photoreactive nucleosides (Experimental Procedures), worms were washed in M9 buffer and cleaned by floating on a sucrose gradient. Worms in Trizol LS reagent (Invitrogen) were homogenized in a precellys 24 Homogenizer (Bertin Technologies) and total RNA was isolated. Thiol-specific biotinylation, dot-blot assays and pull-down of labeled RNA using streptavidin-beads were carried out as described previously (Dolken et al., 2008). For dot-blot assays, typically 10 µg total biotinylated RNA was spotted on the membrane. After isolation of labeled RNA, reverse transcription polymerase chain reaction was performed to test the labeling of tissue-specific transcripts.

RT-PCRs

After labeling *C. elegans* with photoreactive nucleosides, labeled RNA was isolated (Experimental Procedures) and reverse transcription polymerase chain reaction (RT-PCR) was performed to examine the labeling of tissue-specific transcripts. The RT reaction was random primed. Primers for subsequent PCR reaction are listed below. 20-35 cycles of PCR were performed. Gel pictures were processed in Adobe Illustrator.

Gene name	Primer
myo-2	Fwd: agttcgagttccaggttgctgagg Rev: gctctctcttcagcggtttcaagg
myo-3	Fwd: cgctgtctctgatgaagcttaccg Rev: gtacctcccttctttccatccttgg

oma-1	Fwd: cggtgaaaacaacgagaagatcg Rev: ggttgttctggggaaaactctga
pie-1	Fwd: gccgtgattctcgttctagacg Rev: gtatgtcggtgccattggtgca
elt-2	Fwd: cgacttgtatcccgtttcctcagc Rev: ggatgttatcggcaggtcttaggc
unc-8	Fwd: gacttagagggtgcagtgtatcagc Rev: cagtccatacggaagttgggtagc
lag-2	Fwd: gacacgttcacgacgacatctgg Rev: attccgtcgttctcgcatgagc

Quantitative PCR

Reverse transcription quantitative PCRs were performed with random primed cDNA using the SYBR Green qPCR Master Mix according to manufacturer (Invitrogen). *Act-1* and *ama-1* were used as normalizing controls.

mRNA-seq

Poly(A) mRNA was purified from 1 µg of total RNA using the Dynalbeads mRNA Purification Kit (Invitrogen) according to the manufacturer's protocol and subsequently fragmented into approximately 250 nt fragments by chemical fragmentation (200 mM tris acetate pH 8.2, 500 mM potassium acetate, 150 mM magnesium acetate) at 94°C for exactly 3.5 min in a thermocycler. Fragmented RNAs were isolated with RNA Clean beads (Beckman Coulter) according to manufacturer's instructions. Fractionation was checked by capillary electophoresis in a RNA Pico 6000 chip using the Bioanalyzer (Agilent Technologies). First strand cDNA synthesis was accomplished using Superscript III Reverse Transcriptase and random primers, followed by second strand synthesis using DNA Polymerase I and RNaseH (Invitrogen). Double-stranded DNA was purified with Agencourt AMPure beads XP (Beckman Coulter) and quality was checked by capillary gelelectrophoresis on the Bioanalyzer with the Agilent DNA 1000 kit (Agilent Technologies). dsDNA libraries subsequently processed for sequencing using the Genomic DNA Sample Prep Kit (Illumina) according to the manufacturer's protocol and sequenced on HiSeq 2000 (Illumina).

Western blotting

Proteins were transferred to a PVDF membrane using a semi-dry blotting apparatus (BioRad) at 2mA/cm^2. The membrane was blocked in 5% non-fat milk and incubated with mouse anti-GFP antibody (Roche) for 1h at room temperature, washed three times with PBST and incubated with anti-mouse-HRP antibody for 1h at room temperature. The protein bands were visualized using ECL reagent (GE Healthcare) and the LAS-4000 CCD camera (GE Healthcare).

Transgenic reporters

Reporter constructs were made using the Multi Site Gateway Cloning system (Invitrogen).

The *gld-1* promoter was amplified from *C. elegans* N2 genomic DNA using DM126 and DM127 primers. Primer sequences are listed below with uppercase letters representing gateway recombination sites and lowercase letters matching genomic sequence. The PCR fragment was recombined into pDONRP4P1R using BP clonase.

The GFP::H2B reporter construct was amplified using the following primers: DM140 and DM142, which contain a SpeI and BglII restriction site, respectively. The PCR fragment was digested and cloned into a modified pDONR221 vector that contains these restriction sites.

The 3' UTR genomic sequences (3'GS) for *lin-28* and *cpg-2* were amplified from *C. elegans* N2 genomic DNA using the following primers: forward *lin-28* 3' UTR primer from the UTRome *(Mangone et al 2010)* and DM213 for *lin-28* 3'GS or DM148 and DM149 for *cpg-2* 3'GS. The PCR product was recombined into pDONRP2RP3 using BP clonase. These 3' UTR entry clones were used as a template to generate a mutant version, in which two bases of the GLD-1 binding motif were altered. For the mutagenesis PCR we used Pfu Turbo polymerase (Strategene) and the following primers: DM214 and DM215 for *lin-28* 3'GS mutant and DM163 and DM164 for *cpg-2* 3'GS mutant. Primers for mutagenesis PCRs are listed below, where mutations are indicated by uppercase letters. The PCR products were purified (Zymo Research DNA Clean & Concentrator - 5 Kit) and eluted in 15µl water. The eluate was digested with DpnI at 37°C for 2 hours and 2 µl were transformed into Invitrogen TOP10 cells. Colonies were selected and sequence verified for the correct mutations.

Three final entry clones: *gld-1* promoter, GFP::H2B and the gene 3'GS (either the wild-type or mutant version) were recombined with the MosSCI Destination vector, pCFJ151, using LR clonase. The final LR reaction reporter constructs containing either the wild-type or mutant 3'GS were injected into EG4322 for integration onto chromosome II according to the MosSCI direct insertion protocol *(Frokjaer-Jensen et al., 2008)*. Stable integrated lines were maintained on NGM plates seeded with OP50 at 25°C. Adult worms were then imaged for GFP expression.

Images were acquired with a Hamamatsu EM-CCD digital camera attached to an Improvision Yokogawa CSU-10 spinning disc Leica DMIRE2 confocal microscope. All images were taken in oil using a 20x lens in order to capture an entire arm of the *C. elegans* gonad in one field of view. Images were taken as a stack through approximately 20-30µm and then flattened to view nuclei through the gonad arm.

Primer sequences:

5' UTR	
DM 126 f *gld-1* promoter	GGGGACAACTTTGTATAGAAAAGTTGattgagatacacaagtgtttttta

DM 127 r *gld-1* promoter	GGGGACTGCTTTTTTGTACAAACTTGTcattcttcgatggttaacctg
Reporter	
DM 140 speI-GFP::H2B f	GCGACTAGTatgagtaaaggagaagaacttttcactg
DM 142 bglII-GFP::H2B r	ATAAGATCTgccggctagtttacttgctggaa
3' UTR	
UTRome f *lin-28* 3'GS	GGGGACAGCTTTCTTGTACAAAGTGGGAagagaaaagaatagtaattcctctgatgaatag
DM 213 r *lin-28* 3'GS	GGGGACAACTTTGTATAATAAAGTTGCctctcgatttatttcagcgttcgcccgcaatagc
DM148 f *cpg-2* 3'GS	GGGGACAGCTTTCTTGTACAAAGTGGGAttatcatcagtgcactgccggacaa
DM149 r *cpg-2* 3'GS	GGGGACAACTTTGTATAATAAAGTTGCgagtggcacttttttcatctaaaac
Mutagenesis	
DM 214 f *lin-28* mut	gcatgcgttactttaacttacacacGaGctctttttttattcaaattttg
DM 215 r *lin-28* mut	caaaatttgaataaaaaaagagCtCgtgtgtaagttaaagtaacgcatgc
DM163 f cpg-2 mut	attacacactaattgacGaGccccgtgaaaaactattgcccc
CM164 r cpg-2 mut	ggggcaatagtttttcacggggCtCgtcaattagtgtgtaat

Electrophoretic mobility-shift assay using immunoprecipiated GLD-1

0.2 pmol γ-^{32}P-labeled RNA was incubated with different amounts (0.1 – 5 µg) of immunoprecipitated GLD-1::GFP::FLAG protein and 100 ng tRNA in 20 µl binding buffer (20 mM HEPES-KOH, pH 7.4, 330 mM KCl, 10 mM MgCl2, 0.1 mM EDTA and 0.01% IGEPAL CA630 (Sigma)). After addition of 6 µl loading dye (40% glycerol, bromophenol blue in binding buffer), the solution was loaded onto a native 10% (37.5:1) acrylamide gel, running at 200 V for 2 h at room temperature or in the cold, using Tris-glycine buffer as running buffer. Protein-bound RNA and free RNA were visualized using a phosphorimager.

For GLD-1::GFP::FLAG immunoprecipitation, adult worms were dounced on ice in NP40 lysis buffer (50 mM HEPES-K pH 7.5, 150 mM KCl, 2 mM EDTA, 0.5% (v/v) NP-40, 0.5 mM DTT, protease inhibitor cocktail (Roche)). GLD-1::GFP::FLAG fusion proteins were immunoprecipitated for 4h at 4 ºC using anti-FLAG antibody (Sigma, F3165) coupled to Protein G magnetic beads (Invitrogen). Beads were washed with NP40 buffer and protein was eluted by incubating with 3x FLAG peptide (150ng/µl in NP40 buffer) for 30 min at 4ºC.

Electrophoretic mobility-shift assay using the GLD-1 RNA binding domain

Expressed and purified GLD-1 binding domain as a fusion with N-terminal maltose binding protein (MBP) was kindly provided by James Williamson (Ryder et al., 2004). The protocol for direct titration and competition binding assays was adapted from (Ryder et al., 2008). For direct titration experiments, 100pM radiolabeled RNA was incubated with different protein concentrations in 20 µl 1x binding buffer (10mM Tris-HCl pH 8.0, 25mM NaCl, 0.1mM EDTA, 0.1mg/ml tRNA, 5µg/µl heparin) for at least 3 hours at room temperature. After addition of 4µl 6x loading dye (30% (v/v) glycerol, bromphenol blue in 1x binding buffer), 5 µl were loaded on a pre-run, pre-cleaned 6% (29:1) acrylamide gel containing up to 10% glycerol and 0.25x TBE, running at 250V for 30-45 min in the cold. For competition binding assays, a constant concentration of protein was included in the reaction: increasing concentrations of unlabeled competitor RNA were added to reactions containing 100pM labeled *pup-2* RNA (wildtype) and 1µM GLD-1 or 100 pM labeled *lin-28* RNA (wildtype) and 2µM GLD-1. For direct titration assays, fractions of bound and free RNA were quantified using Image Gauge (version 4.0) and dissociation constants were estimated using KaleidaGraph™ software (Version 3.6) by fitting the binding data to the following equation:

$$Y = \frac{m1}{\left(1+\left(\frac{X}{m2}\right)^{m3}\right)}$$

m1: maximal fraction of bound RNA, m2: dissociation constant, m3: cooperativity

Sample preparation for Mass spectrometry

100 µl worm pellets were resuspended in urea-containing buffer (8 M Urea, 100 mM TrisHCl, pH 8.25). 100 µl zirconium beads (SiLi, Germany) were added and worms were lysed in a precellys 24 Homogenizer (Bertin Technologies) two times for 10 seconds at 6,000 rpm. Beads and cell debris was removed by centrifugation (14,000*xg*, 5 min) and protein concentration was measured by Bradford colorimetric assay. 100 µg of each protein sample was mixed in a 1:1 ratio with the "heavy" reference sample (see *in vivo* SILAC, Experimental Procedures). Disulfide bridges of proteins were reduced in DTT (2mM) for 30 minutes at 25ºC and successively free cysteines alkylated in iodoacetamide (11 mM) for 20 minutes at room temperature in the dark. LysC digestion was then performed by adding 5 µg of LysC (Wako) to the sample and incubating it for 18 hours under gentle shaking at 30°C. After LysC digestion, samples were diluted 3 times with 50 mM ammonium bicarbonate solution, 7 µl immobilized trypsin (Applied Biosystems) was added and samples were rotated for 4 hours at 30ºC. 18 µg of the resulting peptide mixtures were desalted on STAGE Tips (Rappsilber et al., 2003) and the eluates dried and reconstituted to 20 µl of 0.5 % acetic acid in water.

In addition, other aliquots of the samples were fractionated in order to obtain a deeper analysis of the proteome. For fractionation, 250 µg control and GLD-1 knockdown sample were mixed 1:1 (w/w) with the heavy reference sample and subjected to digestion as described above. The resulting peptide mixtures were loaded on Empore cartridges (3M) following the instructions from the manufacturer and eluted with 70% acetonitrile. After removing the acetonitrile by evaporation, the peptides were fractionated by isoelectric focusing on a microrotofor device as described in (Adamidi et al., 2011). Briefly, the peptides were diluted to 2.5 ml with MilliQ water and 150 µl of ampholite solution (40% w/w) were added. After focusing for 3 hours following the manufacturer's instruction, 10 fractions for each sample were collected and desalted on 2 or 3 StageTips; eluates were dried and resuspended in 20 µl of 0.5 % acetic acid.

LC-MS/MS analysis

Each sample was analyzed in duplicate. 5 µl of each sample were injected on a LC-MS/MS system (Agilent 1200, Agilent Technologies and LTQ-Orbitrap Velos, Thermo), using a 240 minute gradient ranging from 5% to 45% of solvent B (80% acetonitrile, 0.1 % formic acid; solvent A= 5 % acetonitrile, 0.1 % formic acid). For the chromatographic separation a 45 cm long capillary (75 µm inner diameter) was packed with 3 µm C18 beads (Reprosil-AQ, Dr. Maisch). On one end of the capillary nanospray tip was generated using a laser puller, allowing fretless packing.

The nanospray source was operated with a spay voltage of 1.9 kV and an ion transfer tube temperature of 260ºC. Data were acquired in data dependent mode, with one survey MS scan in the Orbitrap mass analyzer (60,000 resolution at 400 m\z) (LTQ Velos, Thermo) followed by up to

20 MS\MS scans in the ion trap on the most intense ions. Once selected for fragmentation, ions were excluded from further selection for 30 seconds, in order to increase new sequencing events.
For the isoelectric focusing fractions the same conditions as above were employed, with gradient length shortened to 155 minutes and a 30 cm chromatographic column (75 um inner diameter) packed with 1.8 um C18 beads (Dr. Maisch).
Raw data were analyzed using the MaxQuant proteomics pipeline (v1.1.1.36 and 1.2.0.18) and the built in the Andromeda search engine (Cox et al., 2011). Carbamidomethylation of cysteines was chosen as fixed modification, oxidation of methionine and acetylation of N-terminus were chosen as variable modifications. The search engine peptide assignments were mapped to the modENCODE gene models (Gerstein et al., 2010) and filtered at 5% FDR (determined by mapping to inverted gene models) and the feature match between runs was enabled; other parameters were left as default.
337,382 MS/MS spectra (54%) could be mapped to 26,063 unique peptide sequences. These peptides allowed the identification of 3,484 different proteins with at least one unique peptide and 3,130 proteins with least one unique peptide and at least one other peptide.
For pre-fractionated samples, 389,338 MSMS spectra (34%) could be mapped to 41,223 unique peptide sequences and 5,372 (with at least one unique peptide) or 4,821 different proteins (with least one unique peptide and at least one other peptide) could be identified.

Quantification of 4SU incorporation

Total RNA was digested and dephosphorylated to single nucleosides for LC-MS analysis using a modified protocol similar as described by Andrus and Kuimelis (Andrus and Kuimelis, 2001). Briefly, 80 µg of total RNA were incubated for 16 hours at 37 °C with 0.45 U of bacterial alkaline phosphatase (Worthington Biochemical) and 0.45 U of snake venom phosphodiesterase (Worthington Biochemical) in 30 µl buffer containing 13 mM $MgCl_2$, 3 mM Zn Cl_2 and 63 mM Tris-HCl at pH 8.5.
To separate the nucleosides from the residual salts and enzymes 4 µl 3M sodium acetate and 100 µl ethanol were added to the reaction mixture and samples were chilled on dry ice for 10 minutes. After centrifugation (5 min at 14000 g) the clear supernatant was transferred to a new tube, 300 µl ethanol were added and samples were chilled on dry ice for 10 minutes. After centrifugation the supernatant containing the nucleosides was completely dried under vacuum and the samples were dissolved in 30 µl water prior LC-MS analysis.
The nucleoside mixture was separated by UPLC (Agilent 1290) on a Zorbax Eclipse Plus C18 (2.1 × 50mm, 1.8 micron particle size) reverse phase column (Agilent). Separation was performed using 5% methanol in water containing 0.1% formic acid with a flow rate of 0.1 ml/min isocratic for 9 minutes. For detection a TSQ Quantum Vantage triple quadrupole (Thermo) was used. The instrument was operated in SRM mode to monitor the transitions 245 -> 113 (positive mode) and

243 -> 110 + 245 -> 200 (negative mode) for uridine and the transitions 261 - > 129 (positive mode) and 259 -> 116 + 259 -> 216 (negative mode) for 4S-U.

For quantification 11 different concentrations of U and 4S-U, ranging from 12.5pg/µl to20 ng/µl, were used for calibration. Each sample was analyzed at 5 different dilutions in triplicate (ranging from 1:10000 to 1:1), to enable the quantification of each analyte in the linear range of the calibration curve.

II. Computational Analysis

Mapping Illumina small RNA reads

a) Adapter removal and preprocessing of reads

We first removed the 3' adapters by an iterative procedure. Since we did not attempt to map reads less than 18 nt long, we first searched the read sequence starting at position 18 (counted from 5') for matches to the first six nucleotides in the adapter sequence. If no matches are found, then matches to the next five nucleotides were checked, and so on. If a match was found, the corresponding sequence was removed from the read. If not match whatsoever was found, the read was retained. After removing adapters, identical reads where collapsed into single reads.

b) Mapping reads

We ran BWA version 0.5.8a *(Li and Durbin, 2009)* allowing a maximal edit distance of 2 (command line options "aln -n2"). We mapped against all mRNA isoforms that have been annotated by the ModEncode consortium (Gerstein et al., 2010). Reads that mapped to more than one gene model were discarded. The output was converted into SAM format using the BWA "samse" option. All further analyses were carried out after transforming SAM into an in-house developed human readable alignment format. Reads in edit 2 distance were mapping generally with equal probability to the mRNA and to its reverse complement and were thus discarded. All remaining reads were mapped, using exon/intron coordinates, to absolute genome coordinates [Genome release WS190].

c) Clustering reads

Only reads mapping to the + strand of mRNAs were used. The number of reads mapping antisense was vastly smaller when using edit distances of <= 1. Typically, all reads containg T->C conversions or T deletions were, if overlapping or immediately adjacent, grouped into clusters. All remaining reads (in edit 0 or edit 1 distance) that mapped within these clusters were considered part of the respective clusters. For the 6SG library we clustered on all edit 1 reads. Clusters with less than two T->C or T deletions (or, any two edit one reads in the case of the 6SG library) were discarded. Identical reads were only considered once because they might originate from PCR artifacts.

d) Annotating clusters

The midpoints of all clusters were mapped against the annotation. Since, for example, CDS and 3' UTR can overlap for different isoforms from the same locus, we used an iterative strategy based on the assumption that CDS are generally annotated with higher confidence than untranslated regions. If a midpoint mapped to a CDS of a transcript, the cluster was flagged as a "CDS cluster".

If not, the procedure was repeated for 3' UTRs and 5' UTRs. For visual inspection, we generated html files that show the position and sequence of all reads and how they align to the cluster sequence. These files will be available for download. After careful visual inspection of 70 CDS clusters we noticed that ~50% of the CDS clusters are in fact 3' UTR clusters and ~3% are in fact 5' UTR clusters (RefSeq and Wormbase WS220 gene annotations). We estimated the true number of CDS sites and corrected the number of 3' UTR, CDS und 5' UTR clusters in Fig 4d.

Motif Analyses

a) To define the gld-1 binding site motif, MEME (Bailey and Elkan, 1994) version 4.4.0 was run on the top 100 clusters from one 4SU iPAR-CLIP library.
This procedure yielded a significant motif of with relative nucleotide frequencies denoted in the following matrix (columns refer to A, C, G, U):

0.180000 0.150000 0.000000 0.670000
0.790000 0.130000 0.000000 0.080000
0.030000 0.680000 0.000000 0.290000
0.000000 0.010000 0.000000 0.990000
0.680000 0.150000 0.000000 0.170000
0.980000 0.010000 0.000000 0.010000
0.020000 0.650000 0.000000 0.330000

b) To define p-values for the occurrence of motifs defined by the frequency matrix, we used all clusters from a 4SU library (and separately all clusters from the 6SG library) to generate 1 million nucleotides of random background sequences by using a Markov model of order 2, respectively. For each library and each observed cluster length, 10,000 clusters of the same length were sampled from the appropriate background sequence. The frequency of maximum motif scores (scores were defined as 100 - [log of the sum of motif nucleotide frequencies specified by the matrix above]) was computed in this set of 10,000 clusters and used to compute a p-value that quantifies for a given maximum motif score S_M and cluster length L the probability to observe a maximum motif score S_BG at least as high as S_M in the background set of clusters of length L.

c) For comparing the fractions of reproduced and non-reproduced iPAR-CLIP clusters with strong or weak GLD-1 binding motifs, we considered a GLD-1 "affinity score" that was recently published based on classical RIP-CHIP experiments (Wright et al., 2011): after running the published GLD-1 Binding Motif (GBM) finder (Wright et al., 2011), motifs with an affinity score ≥1 were called strong GLD-1 motifs, motifs with a score <1 were called weak GLD-1 motifs.

Conservation Analysis

Multiple Alignments of 5 nematode species with *C. elegans* (ce6/ws190) were downloaded from the UCSC Genome Browser database (Fujita et al.). To infer evolutionary conservation of predicted GLD-1 motifs within iPAR-CLIP clusters, multiple alignments were extracted for these 7-mer motifs flanked by 2 nucleotides upstream and downstream, respectively, after determining the motif position in *C. elegans* by scanning iPAR-CLIP clusters with the GLD-1 Binding Motif (GBM) finder (Wright et al., 2011). Evolutionary conservation in 4 nematodes (*C. briggsae* (cb3), *C. remanei* (caeRem3), *C. brenner*i (caePb2), *C. japonica* (caeJap1)) was assessed after running the GBM finder on the aligned sequence for each of the species. Whenever no motif was found, a score of -10 was assigned. Motifs with a score greater than a minimum score threshold were then considered functional. This threshold was set to -10, after validating that more conservative score thresholds yield comparable outcomes. A motif present in *C. elegans* and at least two other nematodes was considered conserved. All other motifs present in *C. elegans* with available multiple alignments were flagged as non-conserved. Conserved motifs in three clusters (chrIII:4,870,780-4,870,805; chrII:11,595,294-11,595,320; chrII:273,885-273,991) were manually flagged as conserved since they were falsely characterized as non-conserved in the computational analysis.

Quantification of transcript expression in the *C. elegans* germline

Deep sequencing read libraries of WT and *glp-4* samples were mapped to the C. elegans genome sequence (ws190) by running TOPHAT v1.3.1 (Trapnell et al., 2009) (with parameters –i 30 –I 10000 -–mate-inner-dist 250 -–mate-std-dev 125) and using gene models from (Gerstein et al., 2010) to provide reference splice junctions. For WT and *glp-4* libraries 34,758,377 out of 86,776,631 (40%) and 29,286,210 out of 76,081,136 (38%) sequenced reads could be mapped. Subsequently, isoform expression was quantified by running CUFFLINKS v1.0.3 (Trapnell et al., 2010) using the same gene models as reference annotation. Expression of gene loci was estimated by summing up RPKM values of all isoforms and confidence intervals were determined by summing up lower and upper boundaries for isoform expression as determined by CUFFLINKS.

RPKM values were then used to compute an approximation of gene expression levels in the germline as follows. The concentration $c(i)_{WT}$ of molecule i in the WT sample is given by $c(i)_{WT} = (n(i)_S + n(i)_G)/(V_S + V_G)$, where S and G denote the soma and the germline, and V the respective volumes. Thus, to obtain the concentration $c(i)_G = n(i)_G/V_G$ of molecule i in the germline, we need to subtract $c(i)_S$ multiplied with a factor $V_S/(V_S + V_G)$ from $c(i)_{WT}$:

(1) $$c(i)_G \sim c(i)_{WT} - \text{alpha} * c(i)_S$$

with alpha = $(1/(1+V_G/V_S))$. We then estimate $c(i)_{WT}$ and $c(i)_S$ by the RPKM values for the WT and the *glp-4* samples, respectively (the sum of RPKM values was very similar for both samples). It remains to estimate lambda = V_G/V_S. We did this in four different ways. First, we approximated lambda by the ratio of the number of cells in the germline and soma which is known to be roughly 1/(2..3). Second, we used our RPKM values for genes which are known to be mostly expressed in the soma to compute alpha for each gene. The average of alpha was roughly 0.6, consistent with the estimate by cell number. Third, close inspection of Figure 4a revealed a population of genes with RPKM values that are off-diagonal and shifted by a constant value that again corresponded to alpha ~0.6. Fourth, one can interpret equation (1) to define alpha in the following way: alpha is the largest positive value for which (1) has non-negative $c(i)_G$'s. This optimization problem can be solved by using the measured RPKM value and yields optimal alpha values which are consistent with the above three estimations of alpha. We therefore used alpha =0.6 in all plots. However, we checked that our results were overall robust when varying alpha between 0.5 and 1 which corresponds to the range of alpha values for which $V_G \leq V_S$.

Genes with confidently quantified germline expression were identified as follows. Gene expression in WT and in *glp-4* (after multiplying with alpha), respectively, was modeled as a Gaussian distribution centered at the estimated expression for the gene locus with a standard deviation approximated by the difference between the upper boundary of the confidence interval and the mean. The absolute difference between mean expression in WT and *glp-4* (after multiplying by alpha) divided by the maximum standard deviation of the two yielded a z-value which was then converted into a p-value.

Quantification of protein expression

Protein expression was quantified by SILAC ratios computed by MaxQuant *(Cox and Mann, 2008)* against wild-type L4 to adult stage worms as reference. In short, SILAC ratios are first computed for peptides and transcripts encoding a common set of peptides are comprised in a protein group. For each group, the gene locus corresponding to the protein that contains all peptides of the group was selected and expression was quantified by the median SILAC ratio across all peptides. Protein fold changes between GLD-1 knockdown and control samples were computed by dividing the respective SILAC ratios. SILAC ratios of the two independent (with and without pre-fractionation) were combined, using the weighted average whenever a protein was detected in both runs. As weights we used the relative sequence coverage of the two quantifications as calculated by MaxQuant. We were able to obtain fold changes for 217 of the 439 GLD-1 targets. For comparing protein expression changes of GLD-1 targets and all germline genes upon GLD-1 knockdown, we averaged over 202 germline-expressed GLD-1 targets. However, results did not change, when we included soma-expressed GLD-1 targets and averaged over all 217 GLD-1 targets.

Supplemental References

Adamidi, C., Wang, Y., Gruen, D., Mastrobuoni, G., You, X., Tolle, D., Dodt, M., Mackowiak, S.D., Gogol-Doering, A., Oenal, P.*, et al.* (2011). De novo assembly and validation of planaria transcriptome by massive parallel sequencing and shotgun proteomics. Genome Research *21*, 1193-1200.

Andrus, A., and Kuimelis, R. (2001). Base composition analysis of nucleosides using HPLC. . In Curr Protoc Nucleic Acid Chem.

Bailey, T.L., and Elkan, C. (1994). Fitting a mixture model by expectation maximization to discover motifs in biopolymers. Proceedings / ... International Conference on Intelligent Systems for Molecular Biology ; ISMB *2*, 28-36.

Beanan, M.J., and Strome, S. (1992). Characterization of a germ-line proliferation mutation in C. elegans. Development (Cambridge, England) *116*, 755-766.

Biedermann, B., Wright, J., Senften, M., Kalchhauser, I., Sarathy, G., Lee, M.H., and Ciosk, R. (2009). Translational repression of cyclin E prevents precocious mitosis and embryonic gene activation during C. elegans meiosis. Developmental cell *17*, 355-364.

Cox, J., and Mann, M. (2008). MaxQuant enables high peptide identification rates, individualized p.p.b.-range mass accuracies and proteome-wide protein quantification. Nature biotechnology *26*, 1367-1372.

Cox, J., Neuhauser, N., Michalski, A., Scheltema, R.A., Olsen, J.V., and Mann, M. (2011). Andromeda: a peptide search engine integrated into the MaxQuant environment. Journal of proteome research *10*, 1794-1805.

Fraser, A.G., Kamath, R.S., Zipperlen, P., Martinez-Campos, M., Sohrmann, M., and Ahringer, J. (2000). Functional genomic analysis of C. elegans chromosome I by systematic RNA interference. Nature *408*, 325-330.

Frokjaer-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. Nature genetics *40*, 1375-1383.

Fujita, P.A., Rhead, B., Zweig, A.S., Hinrichs, A.S., Karolchik, D., Cline, M.S., Goldman, M., Barber, G.P., Clawson, H., Coelho, A.*, et al.* The UCSC Genome Browser database: update 2011. Nucleic acids research *39*, D876-882.

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.* (2010). Integrative analysis of the Caenorhabditis elegans genome by the modENCODE project. Science (New York, N.Y *330*, 1775-1787.

Hafner, M., Landthaler, M., Burger, L., Khorshid, M., Hausser, J., Berninger, P., Rothballer, A., Ascano, M., Jr., Jungkamp, A.C., Munschauer, M., *et al.* (2010). Transcriptome-wide identification of RNA-binding protein and microRNA target sites by PAR-CLIP. Cell *141*, 129-141.

Jan, E., Motzny, C.K., Graves, L.E., and Goodwin, E.B. (1999). The STAR protein, GLD-1, is a translational regulator of sexual identity in Caenorhabditis elegans. The EMBO journal *18*, 258-269.

Kamath, R.S., Fraser, A.G., Dong, Y., Poulin, G., Durbin, R., Gotta, M., Kanapin, A., Le Bot, N., Moreno, S., Sohrmann, M., *et al.* (2003). Systematic functional analysis of the Caenorhabditis elegans genome using RNAi. Nature *421*, 231-237.

Kennedy, S., Wang, D., and Ruvkun, G. (2004). A conserved siRNA-degrading RNase negatively regulates RNA interference in C. elegans. Nature *427*, 645-649.

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 & development *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 & development *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. Advances in experimental medicine and biology *693*, 106-122.

Li, H., and Durbin, R. (2009). Fast and accurate short read alignment with Burrows-Wheeler transform. Bioinformatics (Oxford, England) *25*, 1754-1760.

Marin, V.A., and Evans, T.C. (2003). Translational repression of a C. elegans Notch mRNA by the STAR/KH domain protein GLD-1. Development (Cambridge, England) *130*, 2623-2632.

Mootz, D., Ho, D.M., and Hunter, C.P. (2004). The STAR/Maxi-KH domain protein GLD-1 mediates a developmental switch in the translational control of C. elegans PAL-1. Development (Cambridge, England) *131*, 3263-3272.

Rappsilber, J., Ishihama, Y., and Mann, M. (2003). Stop and go extraction tips for matrix-assisted laser desorption/ionization, nanoelectrospray, and LC/MS sample pretreatment in proteomics. Analytical chemistry *75*, 663-670.

Ryder, S.P., Frater, L.A., Abramovitz, D.L., Goodwin, E.B., and Williamson, J.R. (2004). RNA target specificity of the STAR/GSG domain post-transcriptional regulatory protein GLD-1. Nature structural & molecular biology *11*, 20-28.

Ryder, S.P., MI, R., and Williamson, J.R. (2008). Quantitative analysis of protein-RNA interactions by gel mobility shift. . Methods Mol Biol. *488*, 99-115.

Schumacher, B., Hanazawa, M., Lee, M.H., Nayak, S., Volkmann, K., Hofmann, E.R., Hengartner, M., Schedl, T., and Gartner, A. (2005). Translational repression of C. elegans p53 by GLD-1 regulates DNA damage-induced apoptosis. Cell *120*, 357-368.

Stiernagle, T. (2006). Maintenance of C. elegans. WormBook, 1-11.

Trapnell, C., Pachter, L., and Salzberg, S.L. (2009). TopHat: discovering splice junctions with RNA-Seq. Bioinformatics (Oxford, England) *25*, 1105-1111.

Trapnell, C., Williams, B.A., Pertea, G., Mortazavi, A., Kwan, G., van Baren, M.J., Salzberg, S.L., Wold, B.J., and Pachter, L. (2010). Transcript assembly and quantification by RNA-Seq reveals unannotated transcripts and isoform switching during cell differentiation. Nature biotechnology *28*, 511-515.

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. The EMBO journal *30*, 533-545.

Xu, L., Paulsen, J., Yoo, Y., Goodwin, E.B., and Strome, S. (2001). Caenorhabditis elegans MES-3 is a target of GLD-1 and functions epigenetically in germline development. Genetics *159*, 1007-1017.