Supporting Information

Mapes et al. 10.1073/pnas.1004513107

SI Methods

Mass Spectroscopy Analysis. The protein bands of interest excised from silver-stained gels were destained by 1% potassium ferricyanide and 1.6% sodium thiosulfate, subjected to reduction and alkylation by 10 mM DTT/55 mM iodoacetamide in 25 mM NH_4HCO_3, and then in-gel digested with trypsin (20 μg/mL in 25 mM NH_4HCO_3) at 37 °C for 16 h. The tryptic peptides were then analyzed by mass spectrometry. For MALDI-TOF mass spectrometric analysis, tryptic peptides were mixed with α-cyano-4-hydroxycinnamic acid (CHCA) matrix containing 2 fmol internal standards and analyzed on an Ultraflex MALDI-TOF mass spectrometer (Bruker Daltonics). For liquid chromatography-tandem mass spectrometry (LC MS/MS) analysis, each peptide mixture was reconstituted in HPLC buffer A (0.1% formic acid), loaded across a trap column (Zorbax 300SB-C18, 0.3 × 5 mm, Agilent Technologies), and separated on a 10-cm analytical C18 column (inner diameter, 75 μm; New Objective). The peptides were eluted using a linear gradient of 0–95% HPLC buffer B (99.9% acetonitrile containing 0.1% formic acid). The LC apparatus was coupled with a 2D linear ion trap mass spectrometer (LTQ-Orbitrap, Thermo Fisher). The resulting spectra from both mass spectrometers were used in searches of the National Center for Biotechnology Information nr database with taxonomy set on *Caenorhabditis elegans* (26,209 sequences) assuming the digestion enzyme trypsin. The MASCOT search engine (http://www.matrixscience.com; v.2.2.03 Matrix Science) was used, allowing one missed cleavage site with charge states from 1+ to 3+. MS mass tolerance was set to be 10 ppm for LC-MS/MS (50 ppm for MALDI-TOF MS), and MS/MS tolerance was set to be 0.5 Da for fixed modification for carbamidomethyl cysteines and variable oxidation of methionine residues. Protein identification was performed using Mowse scores ($P < 0.05$) and the MudPIT algorithm of the MASCOT search engine. Peptide identification was performed using the Scaffold 2 search engine (v.2.06.01; http://www.proteomesoftware.com).

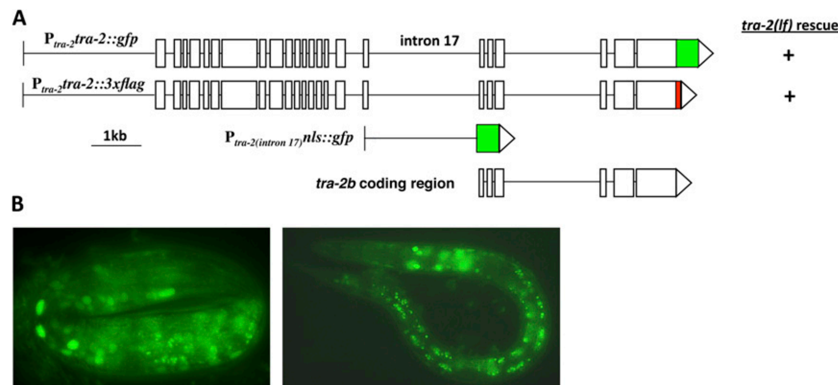
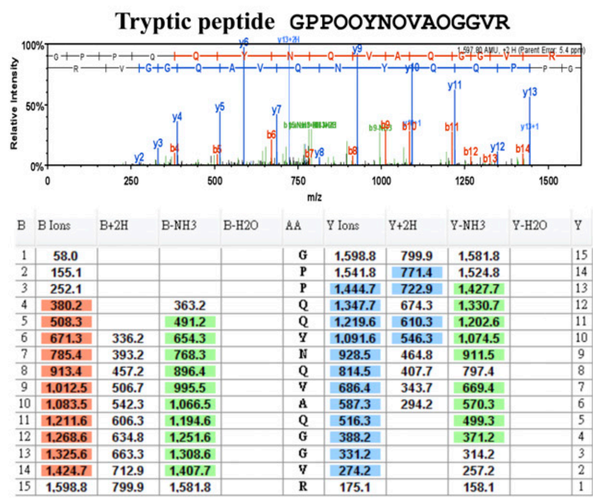


Fig. S1. *tra-2* transgenes. (*A*) Schematic of the $P_{tra\text{-}2}$*tra-2::gfp* and $P_{tra\text{-}2}$*tra-2::3xflag* translational fusions and the $P_{tra\text{-}2(intron\ 17)}$*nls::gfp* transcriptional reporter. Boxes represent exons, lines indicate introns or promoter, and the triangle represents the 3′ untranslated region. The *tra-2b* transcript is also shown. The ability of the translational fusions to rescue the defects of the *tra-2(lf)* mutants is indicated. The *tra-2a* transcript is trans-spliced to the SL2 splice leader, which is consistent with its being a downstream gene in an operon. The *tra-2b* transcript is trans-spliced to the SL1 splice leader, indicating that it is not expressed as a part of the operon (1, 2). The gene upstream of *tra-2* in the operon is *ppp-1*. Green box indicates GFP. Red box indicates the 3xFLAG tag. (*B*) Images of transgenic embryo and larva carrying the $P_{tra\text{-}2(intron\ 17)}$*nls::gfp* transgene, which directs nucleus-localized GFP expression under the control of the 2071-bp intron 17 of the *tra-2a* isoform.

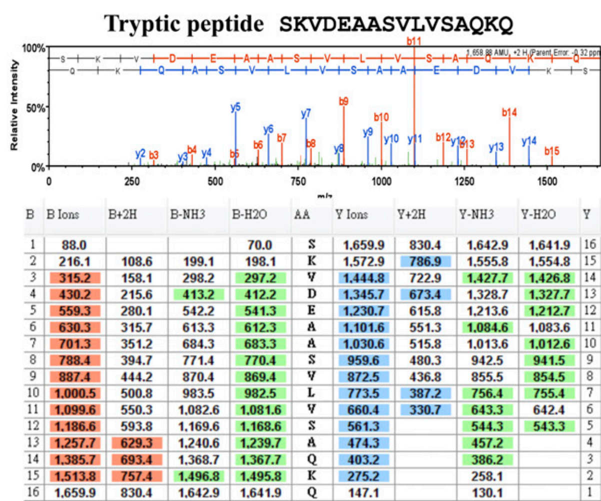
1. Blumenthal T (2005) Trans-splicing and operons. *WormBook* 25:1–9.
2. Kuwabara PE, Okkema PG, Kimble J (1998) Germ-line regulation of the *Caenorhabditis elegans* sex-determining gene *tra-2*. *Dev Biol* 204:251–262.

A PolyA Binding protein family member (PAB-1) [*Caenorhabditis elegans*] (NCBI Reference Sequence: NP_001021711.1)

```
  1 MEMNVAAPAA AVAGAAAPQP GQNQTGSSYT MASLYVGDLH PDVNESILFE KFSAAGPVLS
 61 IRVCRDNATR LSLGYAYVNF QQPADAERAM DTMNFEALHG KPMRIMWSQR DPAMRRSGAG
121 NIFIKNLDKV FVGKFQPRAQ RNRELGETAK QFTNVYVKNF GDHYNKETLE KLFAKFGNIT
181 SCEVMTVEGK SKGFGFVAFA NPEEAETAVQ ALHDSTIEGT DLKLHVCRAQ KKSERHAELK
241 KKHEQHKAER MQKYQGVNLY VKNLDETVDD DGLKKQFESY GNITSAKVMT DENGRSKGFG
301 FVCFEKPEEA TSAVTEMNSK MVCSKPLYVA IAQRKEDRRA QLASQYMQRL ASMRMHGNVP
361 GAAMYNPTQP GPGYYVANPM QQQRNFAGGQ QMVRPGGRWG MQNQYPVQNQ YMMAQGPGVY
421 QNRMGRPQNQ QGGPRGPPQQ YNQVAQGGVR MQGPPRTQNP GVQQQNVPRP PQQQQQQRPA
481 PTGPKAPPQP YQAYQQRPQG IVIGGQEPLT SAMLAAAAPQ EQKQLLGERI YALIEKLYPG
541 HKDAGKITGM MLEIDNSELI MMLQDSELFR SKVDEAASVL VSAQKQ
```



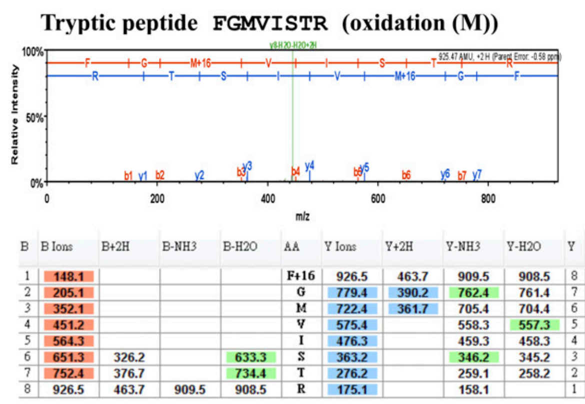
B	B Ions	B+2H	B-NH3	B-H2O	AA	Y Ions	Y+2H	Y-NH3	Y-H2O	Y
1	58.0				G	1,598.8	799.9	1,581.8		15
2	155.1				P	1,541.8	771.4	1,524.8		14
3	252.1				P	1,444.7	722.9	1,427.7		13
4	380.2		363.2		Q	1,347.7	674.3	1,330.7		12
5	508.3		491.2		Q	1,219.6	610.3	1,202.6		11
6	671.3	336.2	654.3		Y	1,091.6	546.3	1,074.5		10
7	785.4	393.2	768.3		N	928.5	464.8	911.5		9
8	913.4	457.2	896.4		Q	814.5	407.7	797.4		8
9	1,012.5	506.7	995.5		V	686.4	343.7	669.4		7
10	1,083.5	542.3	1,066.5		A	587.3	294.2	570.3		6
11	1,211.6	606.3	1,194.6		Q	516.3		499.3		5
12	1,268.6	634.8	1,251.6		G	388.2		371.2		4
13	1,325.6	663.3	1,308.6		G	331.2		314.2		3
14	1,424.7	712.9	1,407.7		V	274.2		257.2		2
15	1,598.8	799.9	1,581.8		R	175.1		158.1		1



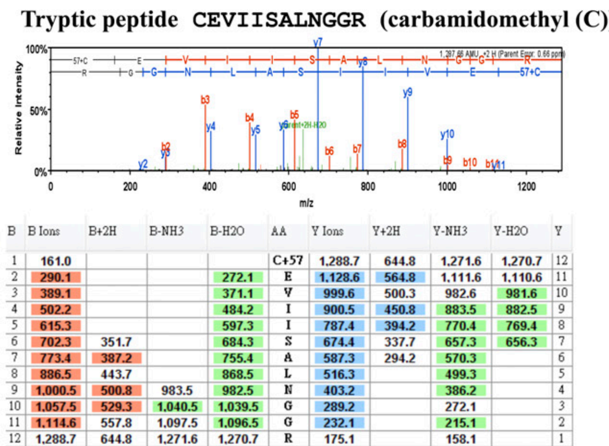
B	B Ions	B+2H	B-NH3	B-H2O	AA	Y Ions	Y+2H	Y-NH3	Y-H2O	Y
1	88.0			70.0	S	1,659.9	830.4	1,642.9	1,641.9	16
2	216.1	108.6	199.1	198.1	K	1,572.9	786.9	1,555.8	1,554.8	15
3	315.2	158.1	298.2	297.2	V	1,444.8	722.9	1,427.7	1,426.8	14
4	430.2	215.6	413.2	412.2	D	1,345.7	673.4	1,328.7	1,327.7	13
5	559.3	280.1	542.2	541.3	E	1,230.7	615.8	1,213.6	1,212.7	12
6	630.3	315.7	613.3	612.3	A	1,101.6	551.3	1,084.6	1,083.6	11
7	701.3	351.2	684.3	683.3	A	1,030.6	515.8	1,013.6	1,012.6	10
8	788.4	394.7	771.4	770.4	S	959.6	480.3	942.5	941.5	9
9	887.4	444.2	870.4	869.4	V	872.5	436.8	855.5	854.5	8
10	1,000.5	500.8	983.5	982.5	L	773.5	387.2	756.4	755.4	7
11	1,099.6	550.3	1,082.6	1,081.6	V	660.4	330.7	643.3	642.4	6
12	1,186.6	593.8	1,169.6	1,168.6	S	561.3		544.3	543.3	5
13	1,257.7	629.3	1,240.6	1,239.7	A	474.3		457.2		4
14	1,385.7	693.4	1,368.7	1,367.7	Q	403.2		386.2		3
15	1,513.8	757.4	1,496.8	1,495.8	K	275.2		258.1		2
16	1,659.9	830.4	1,642.9	1,641.9	Q	147.1		130.1		1

B *C. elegans* protein R10E4.2a (SUP-26a)[*Caenorhabditis elegans*]. GenBank: CAA90772.1

```
  1 MNASSAPQQQ QQQQQPQQGQ APQQQQQGPP QHQYQQQRPF HGRVQNHMRG SGPFGNSNGY
 61 GRYTAPRGDQ QHHDSTPLSS TNLYIRGLMP NTNDDLLREM CSKYGNIAST KAIMDKATNN
121 CKGYGFVDFE SPQAAAAAVD GLNTEGVQAQ MAKLQQQEQD PTNLYIANLP LDFTEQMLET
181 ELNKFGMVIS TRILRTPDNQ SRGVGFARMD SKEKCEVIIS ALNGGRFDTM SKEGPALLIK
241 QADTGRKSKH SMNNPEMLQR MQYPQVYQSY YGYHPAAVYQ QHYDVNSLAS QMGGMHVGGG
301 NPQANGGGDM YGAHMYGQNA GGQGGQGVGQ GGQGGNQQFY NPNNGRNKKY FQPMNPQ
```



B	B Ions	B+2H	B-NH3	B-H2O	AA	Y Ions	Y+2H	Y-NH3	Y-H2O	Y
1	148.1				F+16	926.5	463.7	909.5	908.5	8
2	205.1				G	779.4	390.2	762.4	761.4	7
3	352.1				M	722.4	361.7	705.4	704.4	6
4	451.2				V	575.4		558.3	557.3	5
5	564.3				I	476.3		459.3	458.3	4
6	651.3	326.2		633.3	S	363.2		346.2	345.2	3
7	752.4	376.7		734.4	T	276.2		259.1	258.2	2
8	926.5	463.7	909.5	908.5	R	175.1		158.1		1



B	B Ions	B+2H	B-NH3	B-H2O	AA	Y Ions	Y+2H	Y-NH3	Y-H2O	Y
1	161.0				C+57	1,288.7	644.8	1,271.6	1,270.7	12
2	290.1			272.1	E	1,128.6	564.8	1,111.6	1,110.6	11
3	389.1			371.1	V	999.6	500.3	982.6	981.6	10
4	502.2			484.2	I	900.5	450.8	883.5	882.5	9
5	615.3			597.3	I	787.4	394.2	770.4	769.4	8
6	702.3	351.7		684.3	S	674.4	337.7	657.3	656.3	7
7	773.4	387.2		755.4	A	587.3	294.2	570.3		6
8	886.5	443.7		868.5	L	516.3		499.3		5
9	1,000.5	500.8	983.5	982.5	N	403.2		386.2		4
10	1,057.5	529.3	1,040.5	1,039.5	G	289.2		272.1		3
11	1,114.6	557.8	1,097.5	1,096.5	G	232.1		215.1		2
12	1,288.7	644.8	1,271.6	1,270.7	R	175.1		158.1		1

Fig. S2. Representative MS/MS spectra from LTQ Orbitrap used to confirm the identification of poly(A)-binding protein PAB-1 and SUP-26 in the SUP-26::GFP-containing immunoprecipitated complexes. Tryptic peptides of protein bands 1 and 2 in SDS/PAGE (Fig. 5*A*) were analyzed by LC-MS/MS using LTQ Orbitrap. The amino acid sequences of peptides identified by MS/MS analysis and matched to the amino acid sequences of PAB-1 (*A*) and SUP-26 (*B*) are underlined. The representative MS/MS spectra of two selected peptides (indicated in red) for each protein are shown below the amino acid sequences. The assignments of the fragmented ions observed to specific amino acid residues were performed using the Scadffold 2 search engine, and the search results are shown below the MS/MS spectra.

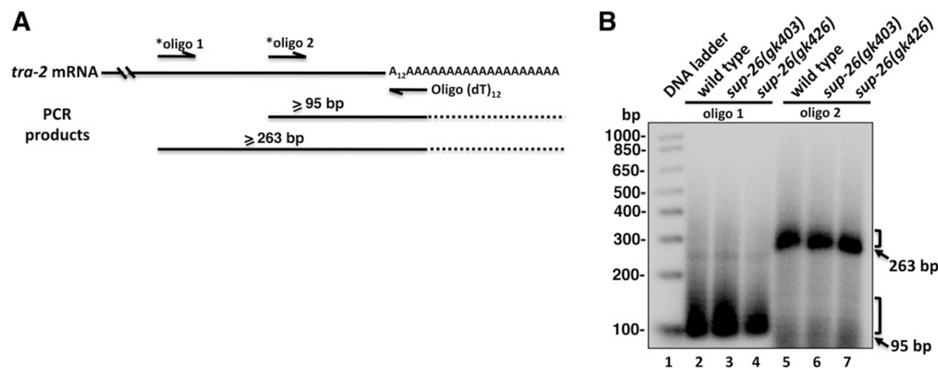


Fig. S3. Analysis of *tra-2* mRNAs using the poly(A) tail length assay (PAT). (*A*) A schematic of RT-PCR products generated by two different primer pairs. Oligo 1 and oligo 2 correspond to different sequences in the *tra-2* 3′ UTR, respectively. An asterisk indicates ^{32}P end-labeling of the two oligos by the polynucleotide kinase. The indicated PCR product sizes represent amplicons with a minimum poly(A) tail length of 12 nucleotides. PCR products larger than 95 bp or 263 bp suggest that the corresponding mRNAs have poly(A) tails longer than 12As. (*B*) RT-PCR products of *tra-2* mRNAs from the indicated genotypes. The PAT assay products were resolved on a 2.5% agarose gel and visualized by autoradiography. Lane 1 contains 100-bp DNA ladders that were end-labeled by ^{32}P. Lanes 2–4 contain RT-PCR products derived from the oligo 1 and oligo$(dT)_{12}$ primer pair, and lanes 5–7 contain products derived from the oligo 2 and oligo$(dT)_{12}$ primer pair.

Table S1. ***sup-26* mutations and their suppression of the *her-1(n695gf)* Tra defect**

Genotype	% CEM surviving*	% HSN surviving*
Wild type (male)†	100	0
Wild type	0	100
sup-26(n1091)	0	100
her-1(n695)	86	0
sup-26(n1091); her-1(n695)	14	96
sup-26(n1091); her-1(n695); smIs336[P_{sup-26}sup-26::gfp]	89	0
sup-26(gk403) ; her-1(n695)	24	100
sup-26(ct49) ; her-1(n695)	31	60
sup-26(gk426) ; her-1(n695)	39	48

*The presence of CEMs and HSNs was determined in young adult hermaphrodites (except noted otherwise) with the aid of the integrated *smIs26* transgene that carries both P_{pkd-2}*gfp* and P_{tph-1}*gfp* reporters, which direct GFP expression in CEMs and HSNs, respectively. The percentage of surviving CEMs was calculated by dividing the total number of CEMs observed by the presumptive number of CEMs [(the number of CEMs observed)/(4 × the number of animals scored)]. The percentage of surviving HSNs was calculated by dividing the number of animals with at least one HSN by the number of animals scored. For all genotypes, 25 animals were scored.
†The wild-type male strain contained the *him-5(e1490)* mutation that results in an increased frequency of X chromosome nondisjunction and thus an increased frequency of XO males. CEM, cephalic companion neurons; HSN, hermaphrodite-specific neurons.

Table S2. ***sup-26* mutations enhance feminization of males when TRA-2 is overexpressed**

Genotype*	% CEM surviving	% HSN surviving	% abnormal male tail
Wild type	100	0	0
sup-26(n1091)	97	0	0
sup-26(gk426)	99	0	0
smIs350/+	91	0	16
sup-26(gk426); smIs350/+	70	0	40

The presence of CEMs and HSNs was determined in young adult males with the aid of the *smIs26* transgene. The percentage of surviving CEMs or HSNs was calculated as described in Table S1. The *smIs350* integrated transgene carries P_{tra-2}*tra-2::3xflag*, which directs expression of the TRA-2::3xFLAG protein from the native *tra-2* promoter (*Methods*). For all genotypes, at least 15 animals were scored.
*All strains contain the *him-5(e1490)* mutation that results in an increased frequency of X chromosome nondisjunction and thus an increased frequency of XO males. CEM, cephalic companion neurons; HSN, hermaphrodite-specific neurons.

Table S3. MALDI-TOF mass spectroscopic analysis of protein bands detected in the SUP-26::GFP-containing immunoprecipitated complexes

Band no./ protein name	M.W. (kDa)	Mascot score	Sequence coverage rate (%)	Accession no.	Peptide: start–end	Mass expected	Mass calculated	ΔMass (ppm)	Peptide sequence matched
1/PAB-1	71.945	59	21%	gi|71993203	52–62	1116.6366	1116.6291	7	K.FSAAGPVLSIR.V
					71–88	2040.9812	2040.9905	−5	R.LSLGYAYVNFQQPADAER.A
					89–104	1863.8457	1863.8430	1	R.AMDTMNFEALHGKPMR.I Oxidation (M)
					180–189	1029.5899	1029.6005	−10	K.VNGMLLAGKK.V
					211–226	1972.8880	1972.9432	−28	K.QFTNVYVKNFGDHYNK.E
					381–394	1634.8166	1634.8636	−29	K.MVCSKPLYVAIAQR.K
					399–409	1350.6494	1350.6826	−25	R.RAQLASQYMQR.L
					400–409	1194.5858	1194.5815	4	R.AQLASQYMQR.L
					445–458	1473.7263	1473.7259	0	R.NFAGGQQMVRPGGR.W
					484–495	1324.6403	1324.6418	−1	R.MGRPQNQQGGPR.G
					496–510	1597.7986	1597.7961	2	R.GPPQQYNQVAQGGVR.M
2/TAG-310 (SUP-26)	39.455	76	25%	gi|17554534	50–62	1268.5529	1268.5534	0	R.GSGPFGNSNGYGR.Y
					68–86	2168.0329	2168.0247	4	R.GDQQHHDSTPLSSTNLYIR.G
					87–98	1357.6623	1357.6660	−3	R.GLMPNTNDDLLR.E
					87–98	1373.6593	1373.6609	−1	R.GLMPNTNDDLLR.E oxidation (M)
					87–103	1992.8794	1992.9067	−14	R.GLMPNTNDDLLREMCSK.Y
					185–192	909.4837	909.4742	10	K.FGMVISTR.I
					185–192	925.4767	925.4691	8	K.FGMVISTR.I oxidation (M)
					213–226	1544.7862	1544.7981	−8	K.EKCEVIISALNGGR.F
					215–226	1287.6580	1287.6605	−2	K.CEVIISALNGGR.F
					233–240	839.5134	839.5116	2	K.EGPALLIK.Q
					248–260	1570.7263	1570.7344	−5	K.SKHSMNNPEMLQR.M
					250–260	1355.6149	1355.6074	6	K.HSMNNPEMLQR.M
					250–260	1371.6174	1371.6023	11	K.HSMNNPEMLQR.M oxidation (M)

Tryptic peptide mass fingerprints of protein bands 1 and 2 in the SDS/PAGE (Fig. 5*A*) were analyzed by MALDI-TOF mass spectrometry. When searched in the National Center for Biotechnology Information nr database with taxonomy set on *C. elegans*, these observed peptide masses matched to 11 peptides of PAB-1 and to 13 peptides of TAG-310 (SUP-26), with a 21% and a 25% sequence coverage rate, respectively. The corresponding amino acid residue numbers (start to end) of matched peptides in PAB-1 and SUP-26 are denoted. The Mascot score used for protein identification is defined as $-10 \times \mathrm{Log}(P)$, where P is the probability that the observed match is a random event. Scores greater than 57 are considered to be significant ($P < 0.05$). ΔMass (in ppm) indicates the difference between expected (expt) mass and calculated (calc) mass, which is defined as [mass(expt) - mass(calc)]/mass(calc) × 1,000,000. M.W., molecular weight.

Table S4. Oligonucleotides used in this study

Oligonucleotide	Sequence
SUP-26proS	5′ GGGGACAAGTTTGTACAAAAAAGCAGGCTTAGCTAGCTCGCAGAGCCTAAAAACATTTC 3′
SUP-26NS	5′ GGGGACAAGTTTGTACAAAAAAGCAGGCTTAATGAACGCATCTTCGGCTCCACAACAAC 3′
SUP-26CAS	5′ GGGGACCACTTTGTACAAGAAAGCTGGGTGTTGTGGATTCATCGGCTGGAAATA 3′
TRA-2proS	5′ CGGGATCCCCATGGTGAGGCCT 3′
TRA-2proAS	5′ AAATCCCGGGTTTTATCTGAAATTGTTATTTGAACG 3′
TRA-2 3′S	5′ ATGAATTCAATGTCTGTTTCCTTTTTCAG 3′
TRA-2 3′AS	5′ ATATACTAGTTAAACTTATTACGTTTATTAAC 3′
TRA-2 3′T7 S	5′ TAATACGACTCACTATAGGGAATGTCTGTTTCCTTTTTCAGAATTTGG 3′
TRA-2 delta TGE S	5′ TGTACAATTTCCATTTCATATCGTCCACTCGAC 3′
TRA-2 delta TGE AS	5′ GTCGAGTGGACGATATGAAATGGAAATTGTACA 3′
oligo(dT)$_{12}$	5′ GCGAGCTCCGCGGCCGCGTTTTTTTTTTTT 3′
TRA-2 oligo 1	5′ AATGTCTGTTTCCTTTTTCAGAATTTGG 3′
TRA-2 oligo 2	5′ ATCGTCCACTCGACCTCAACTTGTAAT 3′