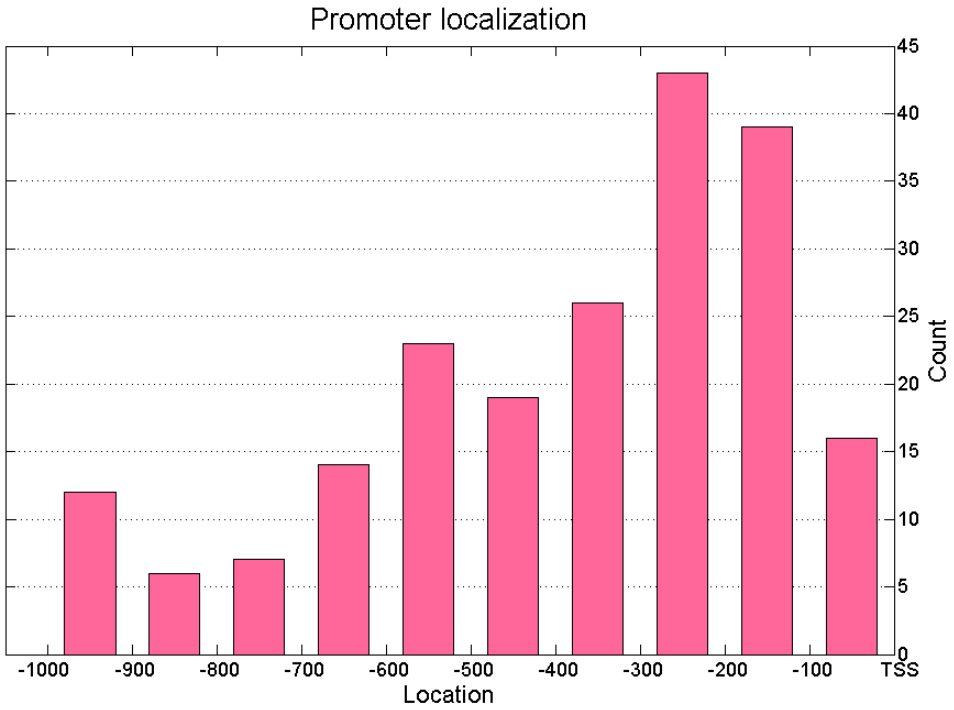
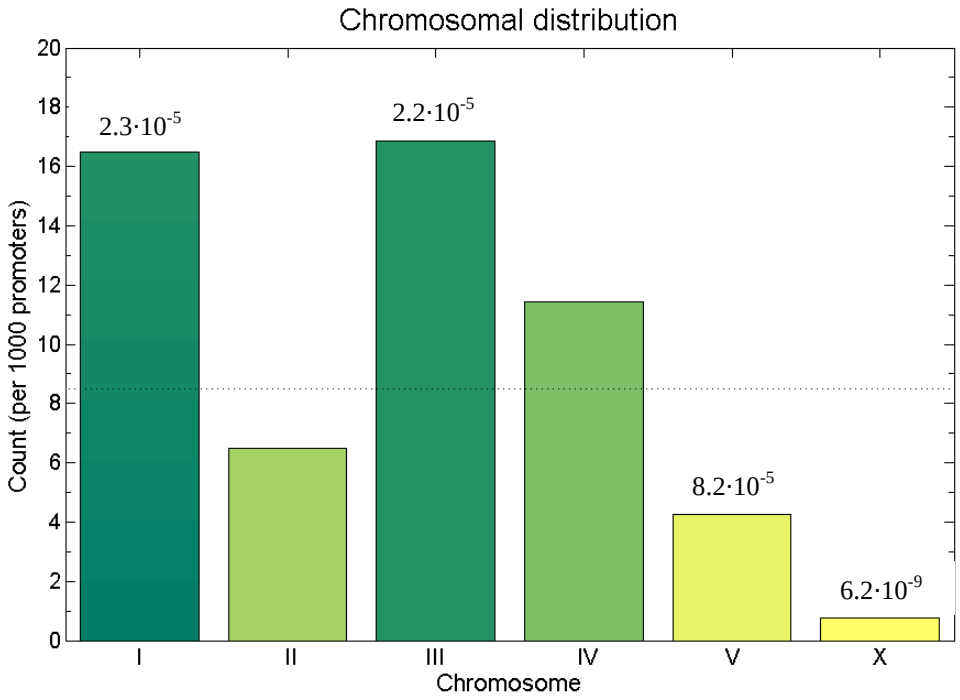
Supplementary Figures

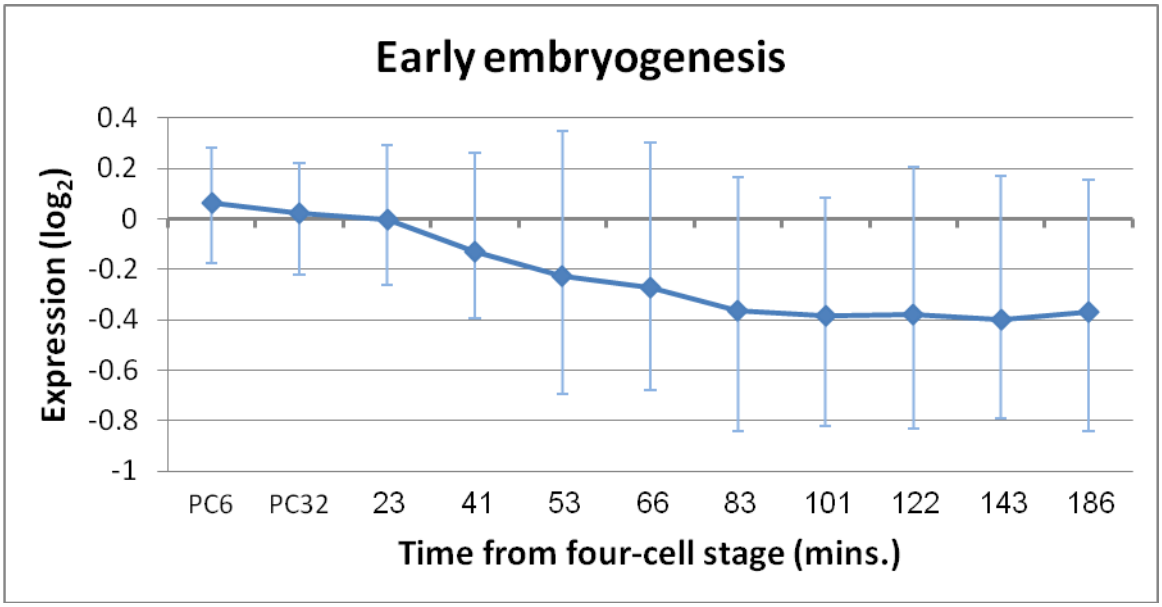


Supp. Fig 1. Location distribution of occurrences of the motif pair in promoters of *C. elegans*. All occurrences of M1→M2 with distance <100 between the motifs were considered (a total of 205 promoters). The promoter regions spanning 1,000 bases upstream of the TSS were analyzed. The bars show (from left to right) the number of occurrences between 1000 and 900 bases upstream of the TSS, the number between 900 and 800 bases upstream of the TSS, and so on. The location of the motif-pair base furthest from the TSS was used in calculating the histogram. The motif pair tends to appear close to the TSS, like many known *cis*-regulatory elements. Other spatial arrangements of the motif pair (especially r(M2)→M1, for which there is a sufficient number of occurrences), as well as the individual motifs, exhibit a similar location bias. In contrast, permuted versions of the motifs have a uniform distance from the TSS (data not shown).

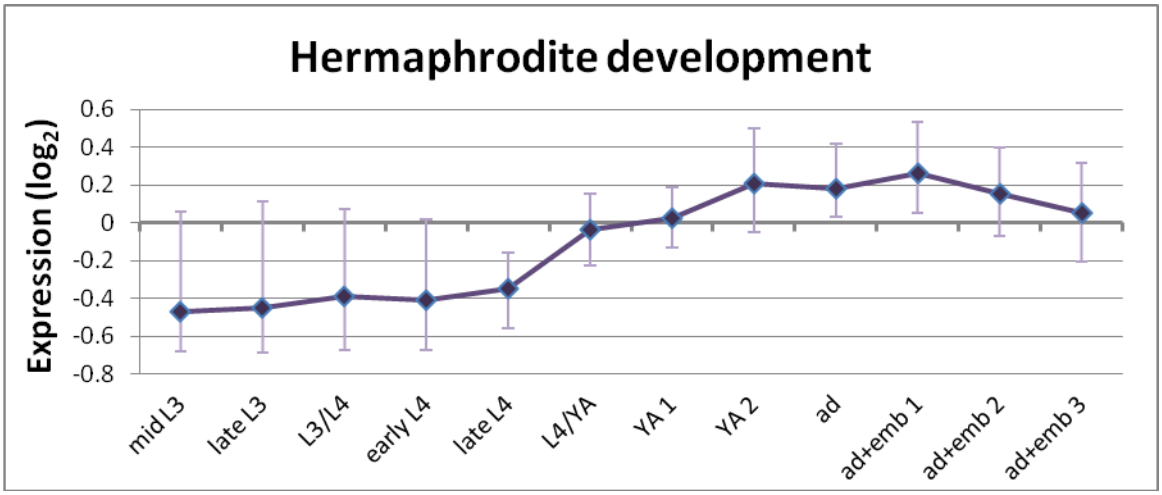


Supp. Fig 2. Chromosomal distribution of the 154 promoters that contain the motif pair M1→M2 with distance <100 between the motifs. The counts are normalized by the number of promoters in each chromosome. Colors indicate the statistical significance of the number of occurrences – over-representation (dark green) or under-representation (yellow). Significant *p*-values are indicated above the bars. The horizontal dotted line indicates the expected count per 1000 promoters (8.6).

(A)

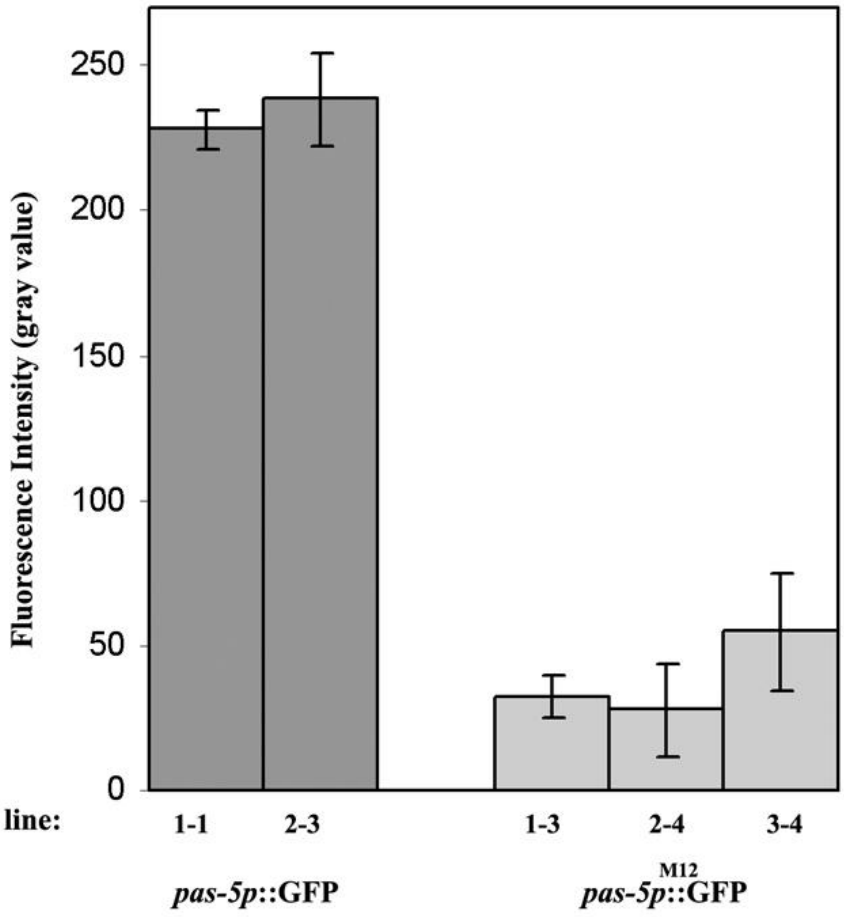


(B)

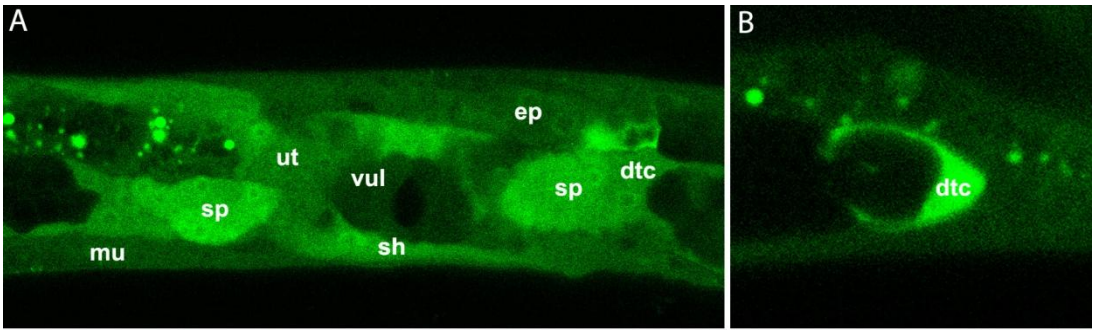


Supp. Fig 3. Temporal expression profiles of the 290 genes containing the motif pair in their promoter (see also Supp. Table 3 for a statistical analysis). (A) Early embryogenesis. Baugh et al. measured transcript abundance profiles during the first 3.5 hours of embryogenesis – from the zygote and into mid-gastrulation (Baugh et al. 2003). For each gene, we normalized its expression profile by computing the $\log_2$ of the ratio between the value at each time point and the value at time 0 (the first time point of the four-cell stage). The graph shows the normalized expression profile, averaged over the candidate target genes regulated by the motif pair. Vertical bars correspond to the first and third quartiles (i.e., 25^{th} and 75^{th} percentiles). Time points are minutes after the four-cell stage (23, 41, …, 186), except the first two time points, which are minutes after pseudocleavage (PC6, PC32). The expression of the motif pair's candidate targets decreases during the first two hours of embryogenesis. (B) Hermaphrodite development.

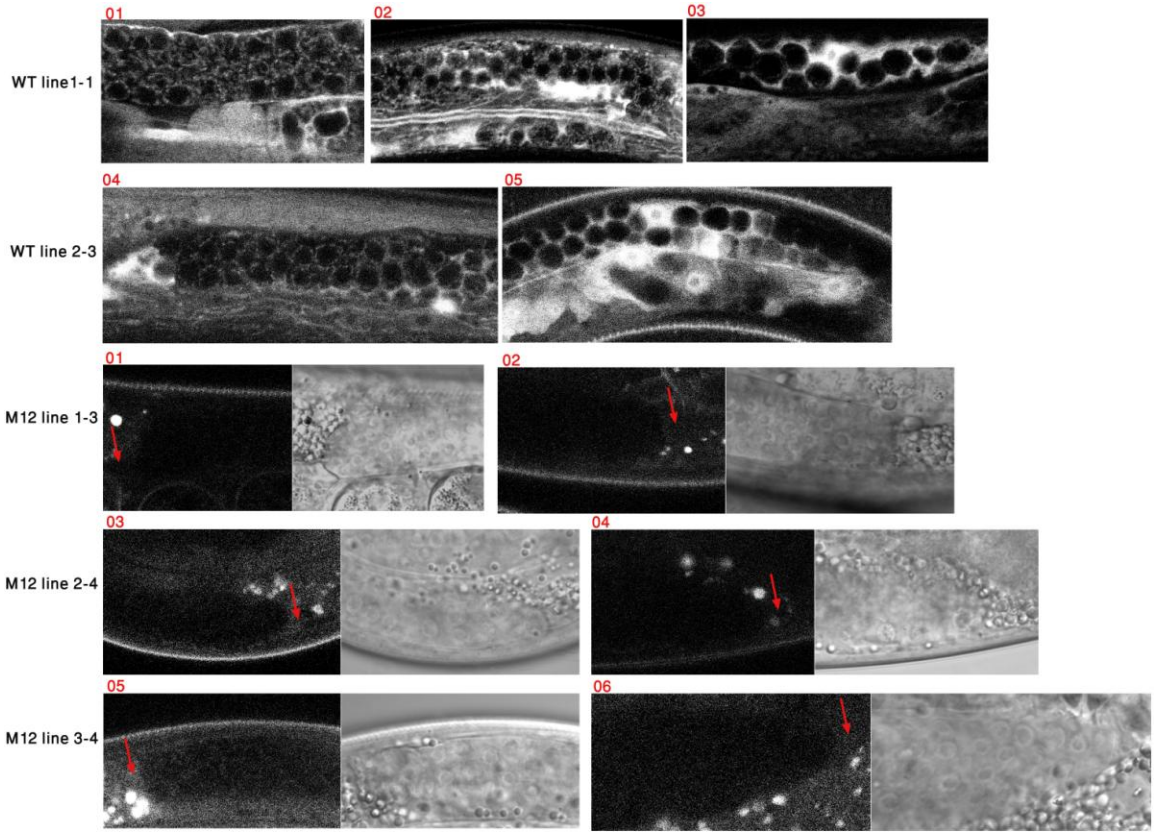
Reinke et al. performed a temporal analysis of gene expression during wild-type hermaphrodite development by measuring gene expression of staged hermaphrodites at 3-hour intervals, beginning in the middle of the third larval stage (L3), through the fourth larval stage (L4), and during adulthood – young adults (YA), adults (ad), and post-reproductive mature adults (ad+emb) (Reinke et al. 2004). Expression values for each gene are $\log_2$ ratio with respect to wild-type reference from mixed developmental stages. Shown in the graph is the average profile of the motif pair's candidate target genes. Vertical bars correspond to the first and third quartiles. The expression is low during L3, and increases sharply from late L4 until adulthood.



Supp. Fig 4. Quantitative analysis of GFP fluorescence in transgenic worms. Quantification of the average fluorescent intensity in the DTCs of 10 animals from each line is presented. Two lines expressing the WT motif pair and three lines expressing the mutant (M12) motif were analyzed. The measured DTC area is marked with the red line in Figure 3E. Mean values and standard deviations (vertical bars) are shown.



Supp. Fig 5. *pas*-5*p*::GFP transgene expressed from an extrachromosomal array. (A) Spermatheca (sp), sheath cells (sh), vulva (vul), uterus (ut), muscle (mu), epidermis (ep). (B) Distal tip cell (dtc).



Supp. Fig 6. Expression pattern of additional WT *pas*-5*p*::GFP and *pas-5p*^{M12}::GFP worms from each of the transgenic lines tested (see Figure 3 and Supp. Fig. 4). The gonad distal tip cells (DTCs) in the mutant transgenes are marked with red arrows.

Supplementary Tables

Supp. Table 1. List of genes containing the motif pair in their promoter. (A) List of 205 genes whose promoter contains the pair M1→M2 with distance <100 between the two motifs. Promoters span 1000 bases upstream of the TSS. For each gene, the location is the pair's base furthest from the TSS. For example, -105 means that the more upstream between the two motifs (M1 or M2) has its most upstream base at a distance of 105 bases from the TSS. Strand is "+" if the occurrence is on the coding strand, and "-" otherwise. The table also specifies which genes are included in the relevant germline and oogenesis sets identified by Reinke et al. (Reinke et al. 2004) (see also Supp. Table 3). (B) List of 85 additional genes that reside in operons regulated by the motif pair (that is, operons whose first gene appears in (A)). Genes marked in yellow (137/290) are known to have central roles in germline control and gametogenesis according to WormBase, or are expressed in the germline (WormBase and NEXTDB, last two columns). Genes marked in orange (97/290) have a sterile and/or maternal sterile phenotype (RNAi data). Genes marked in green (83/290) have embryonic lethal phenotype (RNAi data). Genes marked in blue (72/290) have a function or are expressed in the soma (WormBase).

Species	# of bases	# of occurrences					*p*-value of M1→M2
		M1→M2	r(M1)→r(M2)	r(M1)→M2	M1→r(M2)	Total	
Caenorhabditis elegans	88987435	371	48	125	19	563	$1.5 \cdot 10^{-92}$
Caenorhabditis briggsae	78844299	163	46	44	45	298	$8.6 \cdot 10^{-28}$
Caenorhabditis brenneri	124807846	109	21	56	40	226	$4.4 \cdot 10^{-14}$
Caenorhabditis remanei	99197757	86	25	22	31	164	$5.1 \cdot 10^{-14}$
Caenorhabditis japonica	75125289	58	19	14	27	118	$1.4 \cdot 10^{-8}$
Pristionchus pacificus	146628039	37	32	33	31	133	0.25
Trichuris muris	114200694	35	41	31	28	135	0.43
Strongyloides ratti	387925567	27	28	24	28	107	0.51
Brugia malayi	481095535	49	40	51	60	200	0.59
Teladorsagia circumcincta	335788578	141	134	173	141	589	0.74
Ancylostoma caninum	75461709	17	11	29	19	76	0.74
Heterorhabditis bacteriophora	77636677	7	7	7	16	37	0.85
Ascaris lumbricoides	202943898	19	25	37	14	95	0.90
Globodera pallida	61096087	18	21	35	26	100	0.96
Meloidogyne incognita	400791202	28	39	50	39	156	0.99
Trichinella spiralis	427191962	15	24	40	22	101	1.00
Haemonchus contortus	171387606	66	98	115	81	360	1.00

Supp. Table 2. Number of occurrences of the motif pair in various nematode genomes. "# of bases" is the total number of bases scanned. The "# of occurrences" columns specify the number of times the motifs occur in each of the four possible arrangements. The "*p*-value" column is the *p*-value of the abundance of M1→M2, assuming equal probabiliy for all four arrangements. See Methods for more details. Note that the number of motifs occurrences detected in each species may depend on its unique sequence statistics, such as (di-) nucleotide composition, and also on the number of bases scanned. However, the specific significance test that we performed (checking motif order and orientation) is insensitive to these variables.

Gene set		Overlap with putative targets of motif pair	
Description	**# genes**	**# genes**	***p*-value**
Maternal transcripts that rapidly decay ((Baugh et al. 2003), Fig. 7)			
Cluster 1: sharp decrease during first 1.5 hours	499	27	$3.8 \cdot 10^{-9}$
Cluster 3: small increase in first 30 minutes, then decrease	200	4	x
Cluster 6: like cluster 1, but followed by small increase	122	7	10^{-3}
Sex- and germline-enriched gene sets ((Reinke et al. 2004), Fig. 1)			
Set I: hermaphrodite germline (vs. somatic)	3029	108	$5.5 \cdot 10^{-21}$
Set II: male germline (vs. somatic)	1065	24	x
Set IIIa: hermaphrodite (vs. male)	1880	83	$2.7 \cdot 10^{-21}$
Set IIIb: male (vs. hermaphrodite)	1231	6	x
Set IVa: adult soma in hermaphrodite (vs. male)	447	2	x
Set IVb: adult some in male (vs. hermaphrodite)	417	2	x
Set Va: adult hermaphrodite oogenesis (vs. spermatogenesis)	1614	73	$3.6 \cdot 10^{-19}$
Set Vb: adult hermaphrodite spermatogenesis (vs. oogenesis)	1268	8	x
Temporal expression profiles during hermaphrodite development ((Reinke et al. 2004), Fig. 5)			
Cluster A: Low expression in adult	592	3	x
Cluster B: High from mid-L3 to end of L4, low in adult	713	2	x
Cluster C: High from mid-L3 to young adult	821	9	x
Cluster D: High during L4	696	2	x
Cluster E: Low from mid-L3 to end of L4, high in adult	1147	57	10^{-16}
Cluster F: Similar to E, but gradual increase from L4	685	38	$9.2 \cdot 10^{-13}$

Supp. Table 3. Overlap between gene sets defined by Baugh et al. (Baugh et al. 2003) and Reinke et al. (Reinke et al. 2004) and the 290 putative targets of the motif pair. "# genes" of each gene set is the number of genes for which we obtained a promoter sequence (original sizes of gene sets in (Baugh et al. 2003) and in (Reinke et al. 2004) are slightly larger). We used the HG test in order to check whether the overlap is significantly higher than expected by chance. *p*-values greater than 0.001 are marked with "x".

Maternal transcripts that rapidly decay: Baugh et al. profiled embryonic expression in *C. elegans* embryos during the first 3.5 hours of embryogenesis. They then identified 106 clusters of genes with similar expression patterns. Out of the 20 largest clusters, three clusters (1, 3 and 6) showed a rapid decrease of expression. The set of 290 candidate targets of our motif pair is enriched in cluster 1, and to a lesser extent in cluster 6, but it is not statistically over-represented in any of the other clusters. Clusters 1 and 6 show a clear pattern of maternal transcripts that are degraded; cluster 3 shows an initial rise in transcription followed by degradation. One possible explanation of the delayed decrease is a contribution of a short-time zygotic transcription of the same genes after fertilization - just as a short time boost of the maternal pool that are degraded quite rapidly. This explanation would distinguish our putative target genes as strictly maternal as they do not show enrichment in the maternal + zygotic expression pattern of cluster 3.

Sex- and germline-enriched gene sets: Reinke et al. analyzed gene expression of various mutants of *C. elegans* to identify sets of genes that are sex- and/or germline-regulated.

Temporal expression profiles: Reinke et al. performed a temporal analysis of gene expression during wild-type hermaphrodite development. They collected 12 samples at 3-hour intervals, beginning in the middle of the third larval stage (L3) and extending through adulthood. They then identified 5,083 genes that showed a significant alteration in gene expression levels between two or more time points, and applied hierarchical clustering to define six clusters of genes with similar temporal expression profiles.

Gene set	Single motif		Motif pair	
	M1	M2	r(M1)→M2	M1→M2
Gene Ontology categories				
GO:0009790: embryonic development	$4\cdot10^{-13}$	x	x	10^{-26}
GO:0000003: reproduction	$4\cdot10^{-7}$	x	$3\cdot10^{-8}$	10^{-22}
GO:0010467: gene expression	x	x	x	$2\cdot10^{-9}$
GO:0048513: organ development	x	x	x	10^{-8}
GO:0009791: post-embryonic development	$4\cdot10^{-11}$	x	x	x
GO:0040007: growth	$2\cdot10^{-9}$	x	x	x
SAGE – Serial Analysis of Gene Expression dataset				
SW040: dissected gonad	$6\cdot10^{-8}$	$4\cdot10^{-12}$	$5\cdot10^{-7}$	$8\cdot10^{-23}$
SWYA1: N2 Young Adults	x	x	x	$4\cdot10^{-7}$
SW602: fer-15 6 day	x	x	x	$8\cdot10^{-7}$
SW039: FACS sorted pharyngeal gland cells	x	x	x	$3\cdot10^{-6}$
SW032: purified oocytes	$6\cdot10^{-7}$	x	x	x
Maternal transcripts that rapidly decay ((Baugh et al. 2003), Fig. 7)				
Cluster 1: sharp decrease during first 1.5 hours	x	$4\cdot10^{-7}$	$5\cdot10^{-6}$	$4\cdot10^{-9}$
Cluster 3: small increase in first 30 minutes, then decrease	x	x	x	x
Cluster 6: like cluster 1, but followed by small increase	x	x	x	10^{-3}
Sex- and germline-enriched gene sets ((Reinke et al. 2004), Fig. 1)				
Set I: hermaphrodite germline (vs. somatic)	10^{-6}	$3\cdot10^{-10}$	$4\cdot10^{-5}$	$5\cdot10^{-21}$
Set II: male germline (vs. somatic)	x	x	x	x
Set IIIa: hermaphrodite (vs. male)	$3\cdot10^{-4}$	$5\cdot10^{-8}$	x	$3\cdot10^{-21}$
Set IIIb: male (vs. hermaphrodite)	x	x	x	x
Set IVa: adult soma in hermaphrodite (vs. male)	x	x	x	x
Set IVb: adult some in male (vs. hermaphrodite)	x	x	x	x
Set Va: adult hermaphrodite oogenesis (vs. spermatogenesis)	$7\cdot10^{-8}$	$4\cdot10^{-4}$	x	$4\cdot10^{-19}$
Set Vb: adult hermaphrodite spermatogenesis (vs. oogenesis)	x	x	x	x
Temporal expression profiles during hermaphrodite development ((Reinke et al. 2004), Fig. 5)				
Cluster A: Low expression in adult	x	x	x	x
Cluster B: High from mid-L3 to end of L4, low in adult	x	x	x	x
Cluster C: High from mid-L3 to young adult	x	x	x	x
Cluster D: High during L4	x	x	x	x
Cluster E: Low from mid-L3 to end of L4, high in adult	$2\cdot10^{-5}$	$2\cdot10^{-6}$	$7\cdot10^{-4}$	10^{-16}
Cluster F: Similar to E, but gradual increase from L4	$2\cdot10^{-6}$	$2\cdot10^{-6}$	x	$9\cdot10^{-13}$

Supp. Table 4. Functional analysis of genes whose promoter contains only one of the motifs or the motif pair in different configurations. The single motif columns refer to genes that contain one of the motifs but not the other - M1 (1,727 genes), M2 (792 genes). The last two columns contain the results for genes with the motif pair in the orientations r(M1)→M2 (53 genes) and M1→M2 (the original configuration, 290 genes). The two other orientations, M2→M1 and M2→r(M1), appear in a small number of genes (15 and 11, respectively), and we did not detect any enrichment for them. Statistical *p*-values were computed as described in Supp. Table 3. The sets of genes with the single motifs or with the pair r(M1)→M2 exhibit similar functional trends to those of the pair M1→M2, though with considerably lower statistical significance.