Section 1: Input Datasets

Ceanorhabditis elegans pharyngeal muscle expressed gene dataset

Gene Public Name	Sequence Name	Comment
unc-89	C09D1.1	UNCoordinated, encodes a Ca^{2+}/calmodulin dependent protein kinase
sma-6	C32D5.2	Encodes a serine/threonine proteine kinase
hen-1	C36B7.7	HEsitatioN behaviour, encodes a secretory protein affecting integration of sensory stimuli and behavioral plasticity
C49F8.2	C49F8.2	-
exp-2	F12F3.1	EXPulsion defective, encodes a member of a family of potassium channels required for normal feeding behavior
mkk-4	F42G10.2	MAP Kinase Kinase homolog, required in presynaptic neurons
nucb-1	F44A6.1	NUCLeoBindin homolog
dyb-1	F47G6.1	DYstroBrevin homolog
gas-1	K09A9.5	General Anaesthetic Sensitivity abnormal, encodes a subunit of mitochondrial complex I
rga-3	K09H11.3	Rho GTPase Activating protein
unc-68	K11C4.5	UNCoordinated, encodes a ryanodine receptor ortholog
egl-30	M01D7.7	Egg Laying defective, encodes an ortholog of the G protein alpha subunit Gq
nud-2	R11A5.2	Aspergillus Nuclear Division related, encodes an ortholog of human NDE1 and NDEL1
ric-3	T14A8.1	Resistance to Inhibitors of Cholinesterase, necessary for the function of at least 4 nicotinic acetylcholine receptors
myo-2	T18D3.4	MYOsin heavy chain structural class, encodes a muscle-type specific myosin heavy chain isoform
tni-4	W03F8.1	TropoNin I, encodes a troponin I protein
ctl-3	Y54G11A.13	CaTaLase, encodes one of three *C. elegans* catalases
cua-1	Y76A2A.2	CU (copper) ATPase, a cation transport ATPase
unc-52	ZC101.2	UNCoordinated, encodes prelecan, orthologous to human HSPG2
obr-3	ZK1086.1	Oxysterol Binding protein Related

Ciona intestinalis muscle dataset specific gene dataset

Gene name	Ensembl Transcript ID	Comment
Ci-TnC2	ENSCINT00000012854	Troponin C 2
Ci-MA5	ENSCINT00000002686	Muscle Actin 5
Ci-MA4	ENSCINT00000014519	Muscle Actin 4
Ci-MA1	ENSCINT00000015480	Muscle Actin 1
Ci-MRLC4	ENSCINT00000014813	Myosin Regulatory Light Chain 4
Ci-MRLC3	ENSCINT00000024572	Myosin Regulatory Light Chain 3
Ci-MRLC2	ENSCINT00000005614	Myosin Regulatory Light Chain 2
Ci-MA2	ENSCINT00000004215	Muscle Actin 2
Ci-TnT	ENSCINT00000010963	Troponin T
Ci-TPM1	ENSCINT00000004783	Tropomyosin 1
Ci-TPM2	ENSCINT00000009059	Tropomyosin 2
Ci-TPM3	ENSCINT00000008881	Tropomyosin 3
Ci-MLC2	ENSCINT00000009951	Myosin Light Chain 2
Ci-MLC3	ENSCINT00000009959	Myosin Light Chain 3
Ci-MLC4	ENSCINT00000009977	Myosin Light Chain 4
Ci-MHC2	ENSCINT00000002395	Myosin Heavy Chain 2
Ci-MHC4	ENSCINT00000019909	Myosin Heavy Chain 4
Ci-MHC5	ENSCINT00000019346	Myosin Heavy Chain 5
Ci-MHC6	ENSCINT00000019915	Myosin Heavy Chain 6

Section 2: Motif Prediction Parameters

For both datasets 4 sequence sets were made. They consist of:

1. the first 500 bases upstream of the translation start site of each gene
2. the first 1000 bases upstream of the translation start site of each gene
3. the 2000 bases upstream of position -1000 relative to the translation start site of each gene
4. the entire sequence (3000 bases) upstream of each gene.

On each of these 4 sets three motif prediction programs were applied.

For MEME [1] motifs were predicted using the mode allowing any number of sites for a motif in each sequence. Only motifs with at least 5 predicted sites, and with a width of 6 to 15 base pairs were investigated. Different runs were done considering only one strand, and considering both strands. A background model based on the genome wide set of promoter sequences was used to model oligonucleotide frequencies. A more detailed description can be found in the MEME manual.

For Weeder [2] motifs were predicted using the "medium" analysis type, which evaluates motifs of lengths 6, 8, and 10 base pairs allowing up to 3 mutations. The 10 highest-scoring motifs of each run are saved, and for each motif any number of sites per sequence was allowed. Different runs were done considering only one strand, and considering both strands. A background model based on the genome wide set of promoter sequences was used to model oligonucleotide frequencies.

For AlignACE [3] motifs were predicted using standard parameter settings. The GC content of the input sequences was used as background model.

The top 10 motifs of each run for each program are collected, and a variable order Bayestian network model was made for each motif based on the sites reported by the prediction programs. Subsequently, the group specificity as described by Hughes *et al.* [3] was calculated for each motif using the input sequences and a set of 1000 randomly selected promoter sequences from the genomic set of promoter sequences. For the *C. elegans* set the 7 most over-represented motifs, and for *C. intestinalis* the 10 most over-represented motifs were used for further analysis.

Section 3: Predicted Motifs and Position Dependencies

Position dependencies in motifs

C. elegans pharyngeal muscle dataset motifs

In this dataset we introduced only one dependency. In motif Cel_PM5 position 11 has a slight preference for the base T. However, when the base on position 12 is A, position 11 tends to prefer A as well.

C. intestinalis muscle dataset motifs

In this dataset we introduced 5 dependencies in 4 motifs.

In motif Cin_M1, position 2 prefers T, except when position 3 is A position 2 tends to be G.

In motif Cin_M4, position 8 tends to be A, but when position 9 is C position 8 shows a strong preference for G.

In motif Cin_M5 we detected 2 dependencies. Position 11 tends to be G but prefers A when position 7 is A. Secondly, position 8 tends to be T but prefers C when position 5 is C.

In motif Cis_M6, position 7 shows a strong preference for A, but when position 4 is A, position 7 tends to be T.

Of the 6 dependencies we found in total, 3 are between adjacent positions. The other 3 dependencies as well are between relatively proximal base pairs (one for a distance of 2, 3 and 4 bps distance each).

Motif logos

Below are the motif logos for the motifs used in the *C. elegans* pharyngeal muscle promoter architecture model, and the *C. intestinalis* muscle promoter architecture model. Note that motif logos do not consider dependencies between positions, and thus give an incomplete representation of the motif.

C. elegans pharyngeal muscle dataset motif logos

Motif name	Motif logo
Cel_PM1	
Cel_PM2	

Cel_PM3	
Cel_PM4	
Cel_PM5	
Cel_PM6	
Cel_PM7	

C. intestinalis muscle dataset motif logos

Motif name	Motif logo

Cin_M1	
Cin_M2	
Cin_M3	
Cin_M4	
Cin_M5	
Cin_M6	

Cin_M7	
Cin_M8	
Cin_M9	
Cin_M10	
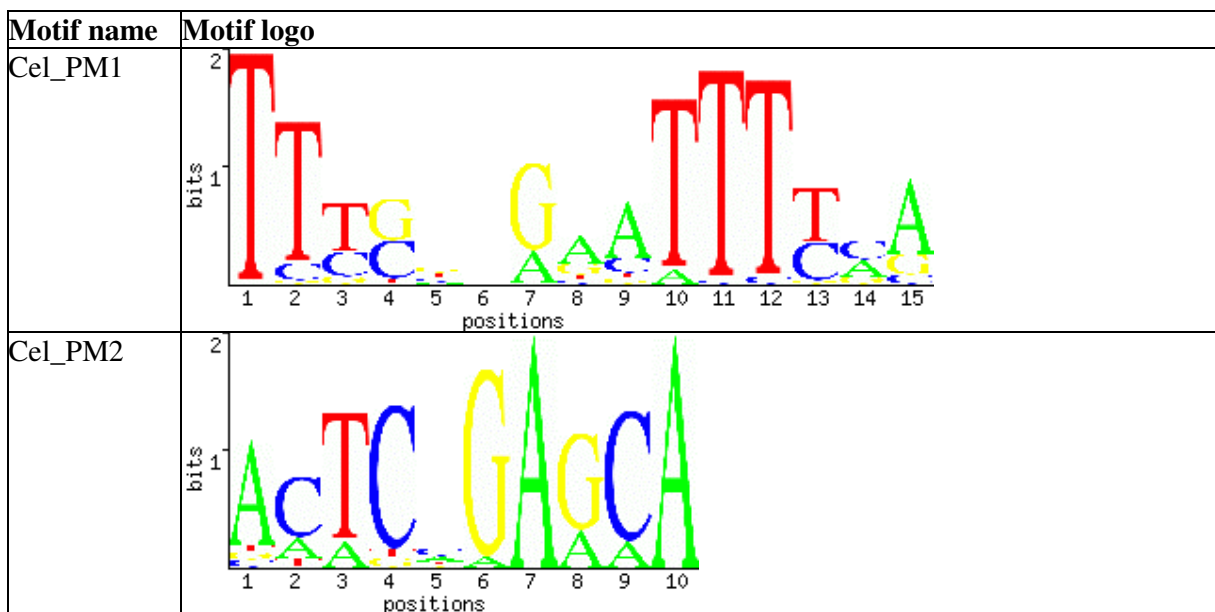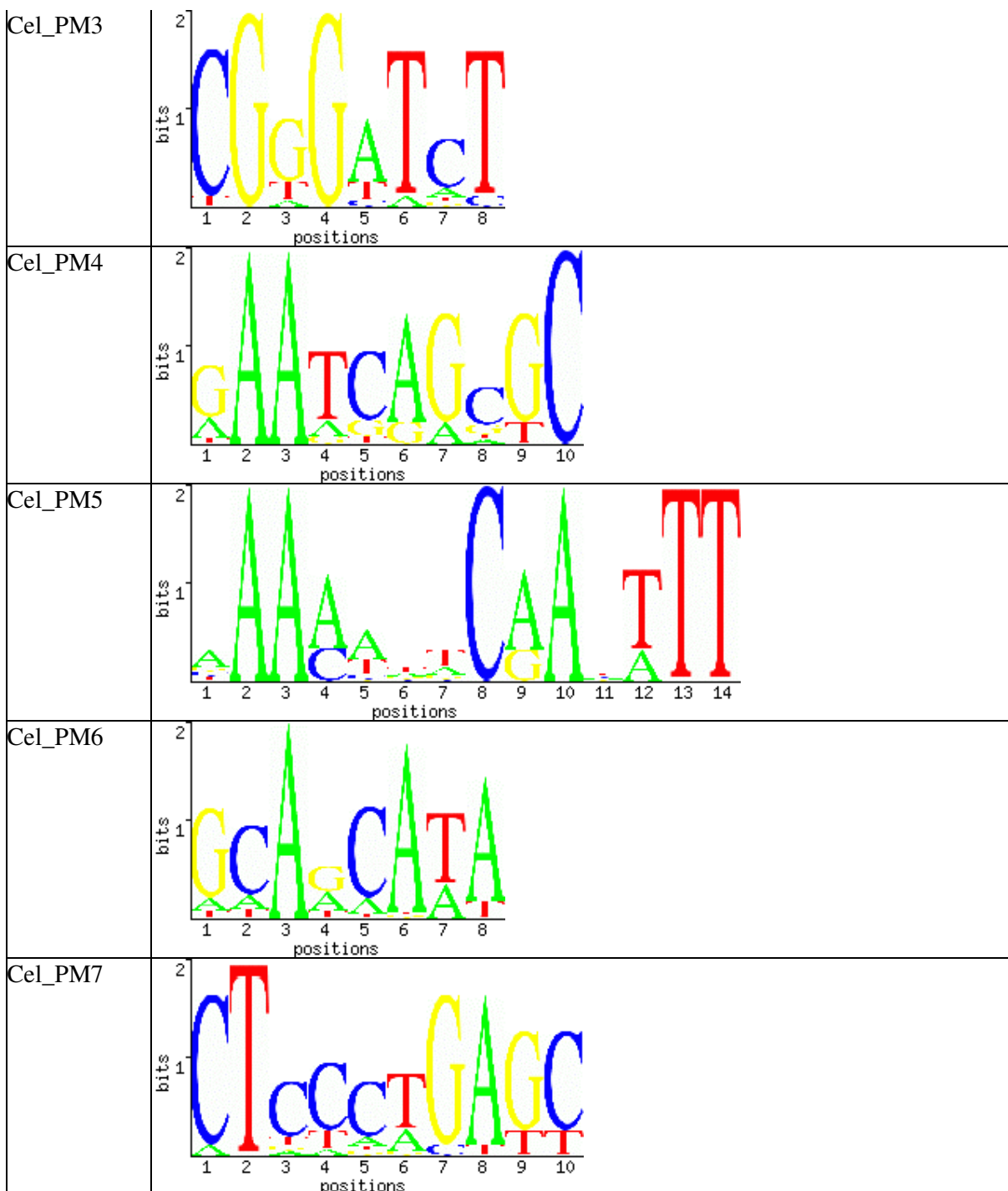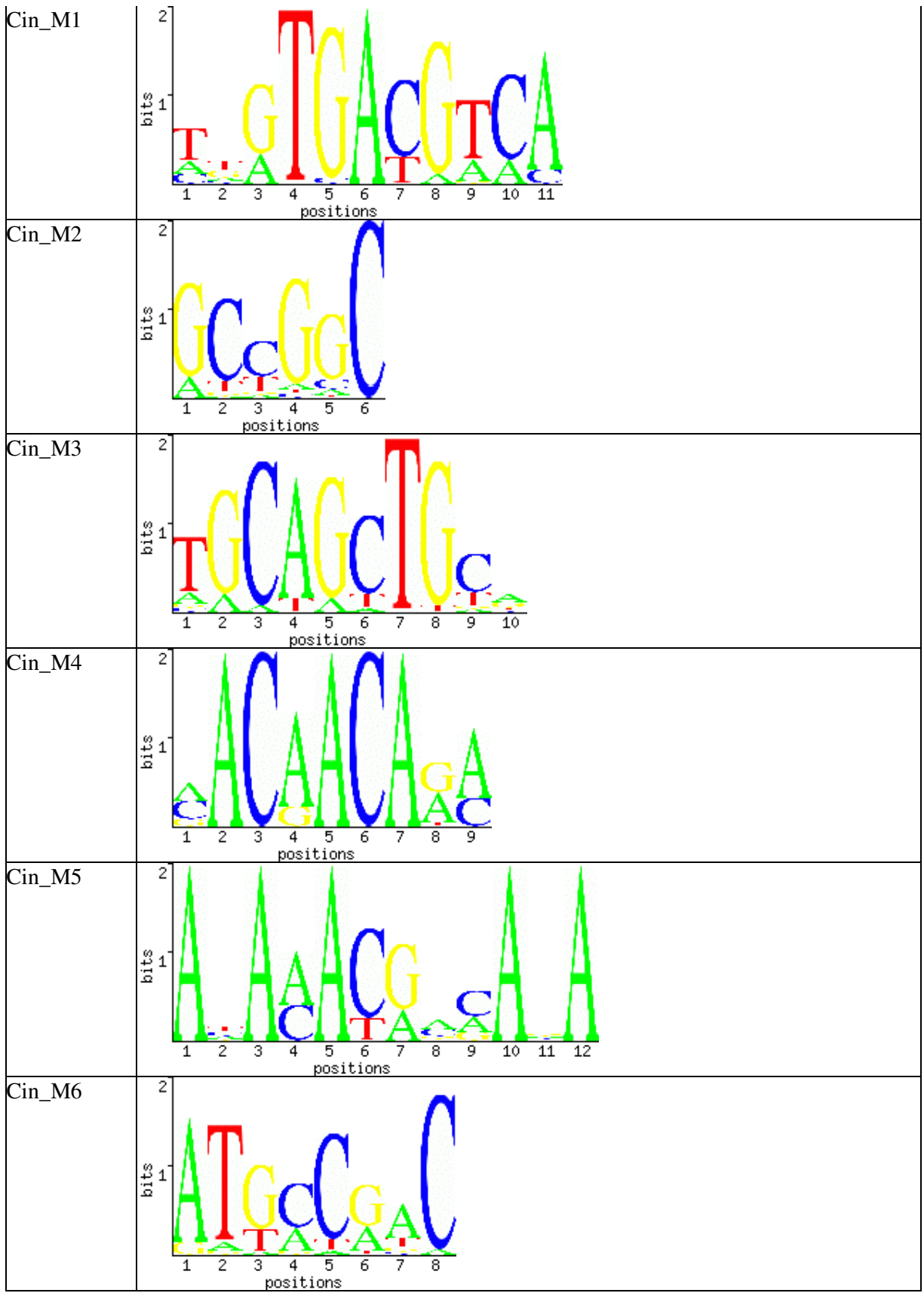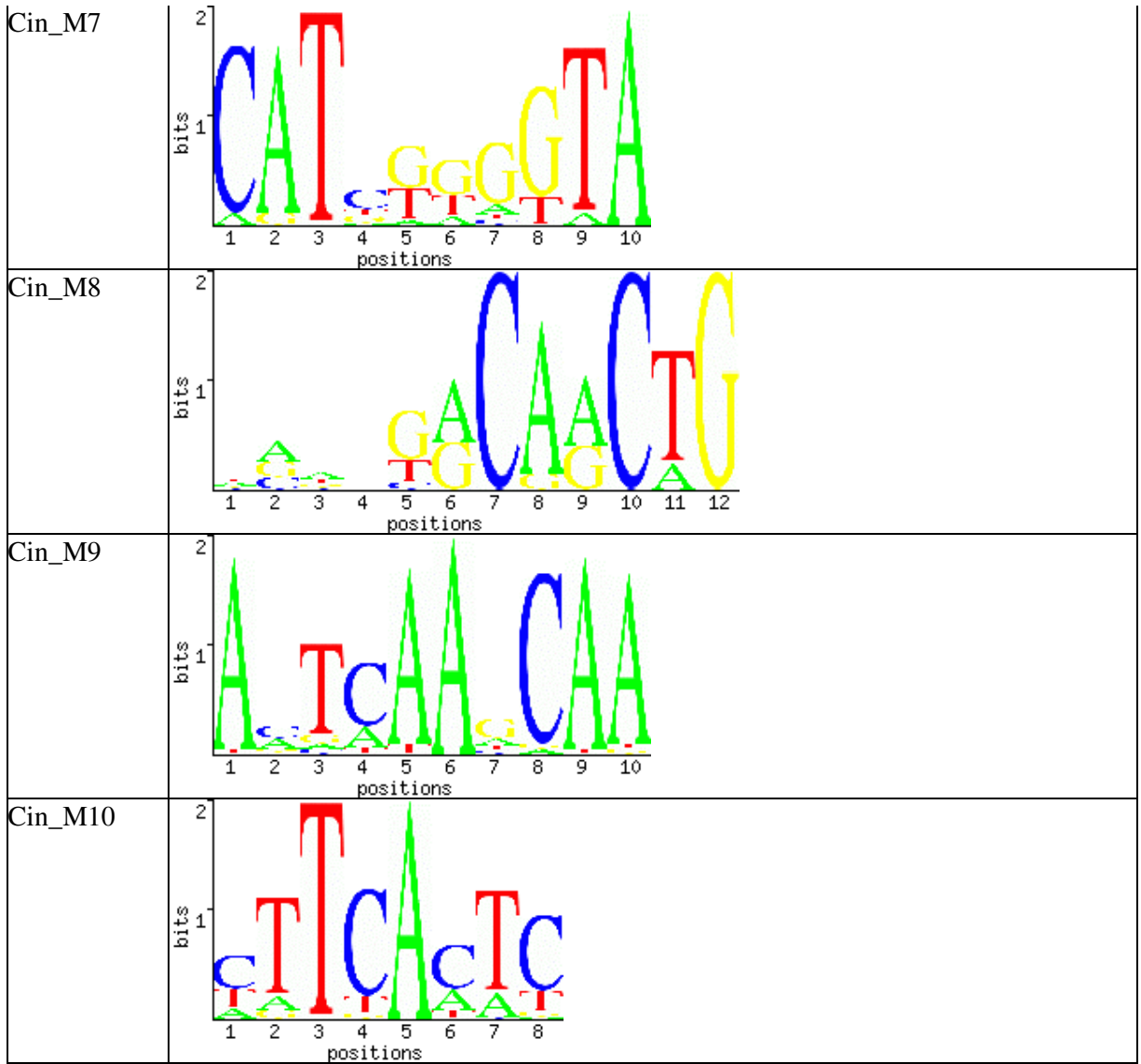
Section 4: Constructed Promoter Models

Below are the log likelihood ratios (LLRs) for the transitions between motif site pairs, for the proximal and distal regions of both promoter architecture models. The LLR values are calculated as described in the main text. The first column shows the first motif and its orientation, the top row shows the upstream adjacent motif and its orientation. The values in the table are the LLR scores for a transition between the first motif and the adjacent upstream motif. Also, the start and end state of each region is included in these tables.

C. elegans pharyngeal muscle promoter architecture model

Proximal region LLR values

	Cel_PM3+	Cel_PM3-	Cel_PM1+	Cel_PM1-	Cel_PM7+	Cel_PM7-	Cel_PM4+	Cel_PM4-	Cel_PM6+	Cel_PM6-	Cel_PM5+	Cel_PM5-	Cel_PM2+	Cel_PM2-	end
Cel_PM3+	0.09	0.33	0.22	-0.20	0.55	0.02	0.00	0.08	0.04	0.07	-0.05	0.21	0.13	0.40	-0.94
Cel_PM3-	0.23	-0.14	0.36	-0.23	-0.03	0.29	0.55	-0.04	-0.07	0.27	0.18	0.17	-0.10	-0.03	-0.65
Cel_PM1+	0.18	0.30	0.27	0.43	-0.17	-0.19	0.07	-0.07	-0.12	-0.16	0.14	0.28	0.20	-0.14	-0.80
Cel_PM1-	0.05	0.37	0.41	0.05	0.33	-0.05	0.33	0.05	0.28	0.30	-0.20	-0.28	0.27	0.35	-1.01
Cel_PM7+	0.25	0.42	-0.16	0.15	0.11	-0.25	0.42	0.41	0.44	0.24	0.09	-0.17	-0.09	0.00	-0.69
Cel_PM7-	0.07	0.08	0.51	-0.15	0.27	-0.06	0.02	0.09	0.16	0.40	0.05	0.08	0.05	0.21	-0.82
Cel_PM4+	0.27	0.25	0.12	0.20	0.14	-0.12	0.00	-0.07	0.25	0.15	0.13	0.18	0.36	0.16	-0.58
Cel_PM4-	0.15	0.14	0.03	-0.04	0.01	0.16	0.49	0.03	0.05	0.13	0.35	0.04	0.42	-0.07	-0.82
Cel_PM6+	0.13	0.46	0.25	-0.09	-0.04	0.30	0.04	0.04	-0.01	-0.03	-0.01	-0.12	0.00	0.12	-0.24
Cel_PM6-	0.01	-0.04	-0.07	-0.04	0.33	-0.06	0.08	0.03	0.23	0.21	-0.02	0.22	0.26	0.08	-0.27
Cel_PM5+	0.10	0.15	-0.16	-0.32	0.09	0.00	0.37	0.37	0.34	0.07	-0.26	0.26	0.35	0.60	-0.50
Cel_PM5-	0.08	0.12	-0.19	-0.23	0.35	-0.01	0.49	0.01	0.28	0.32	-0.05	0.09	0.34	0.44	-0.48
Cel_PM2+	0.58	0.42	0.37	0.14	-0.30	0.24	0.02	0.05	0.03	-0.02	0.21	-0.06	-0.04	-0.08	-0.48
Cel_PM2-	0.03	0.04	0.07	-0.12	0.41	0.15	0.51	0.00	0.02	0.20	-0.05	-0.14	0.41	0.06	-0.65
start	0.33	0.13	0.07	-0.07	0.32	-0.22	0.32	0.08	0.07	0.07	0.13	-0.39	0.14	0.43	-0.98

Distal region LLR values

	Cel_PM3+	Cel_PM3-	Cel_PM1+	Cel_PM1-	Cel_PM7+	Cel_PM7-	Cel_PM4+	Cel_PM4-	Cel_PM6+	Cel_PM6-	Cel_PM5+	Cel_PM5-	Cel_PM2+	Cel_PM2-	end
Cel_PM3+	0.45	0.53	0.23	0.38	0.04	0.10	0.14	0.36	0.19	0.25	0.30	0.30	0.43	0.40	-0.87
Cel_PM3-	-0.03	0.02	0.42	0.21	0.14	0.08	0.32	0.39	0.28	0.30	0.56	0.16	0.14	0.22	-0.50
Cel_PM1+	0.23	0.39	0.04	0.00	0.53	0.07	0.43	0.28	0.18	0.37	0.21	0.53	0.43	0.31	-0.68
Cel_PM1-	0.41	0.32	0.06	-0.10	0.26	0.27	0.54	0.56	0.53	0.30	0.19	0.23	0.70	0.37	-0.89
Cel_PM7+	0.04	0.28	0.42	0.14	0.11	-0.37	0.36	0.26	0.47	0.49	0.12	0.03	0.17	0.03	-0.45
Cel_PM7-	0.30	0.40	0.07	0.36	0.28	-0.03	0.30	0.54	0.44	0.31	0.46	0.24	0.34	-0.02	-0.89
Cel_PM4+	0.29	0.48	0.14	-0.01	0.17	0.04	0.29	0.42	0.48	0.40	0.55	0.15	0.17	0.27	-0.63
Cel_PM4-	0.19	0.35	0.00	0.28	0.36	0.23	0.17	0.43	0.35	0.53	0.19	0.19	0.14	-0.08	-0.60
Cel_PM6+	0.43	0.24	0.26	0.09	0.31	0.11	0.18	0.26	0.10	0.26	0.18	0.21	0.23	0.29	-0.45
Cel_PM6-	0.40	0.31	0.33	0.02	0.28	0.21	0.35	0.19	0.11	-0.07	0.48	0.26	0.45	0.18	-0.46
Cel_PM5+	0.26	0.49	0.01	-0.03	0.14	0.44	0.67	0.45	0.31	0.26	0.13	0.20	0.14	0.38	-1.01
Cel_PM5-	0.15	0.44	0.30	0.05	0.23	0.09	0.58	0.28	0.14	0.38	-0.18	0.20	0.25	0.30	-0.53
Cel_PM2+	0.41	0.39	0.13	0.33	0.02	0.43	0.07	0.31	0.33	0.43	0.20	0.22	0.41	0.07	-0.85
Cel_PM2-	0.39	0.31	0.17	0.11	0.17	-0.02	0.24	0.48	0.18	0.35	0.37	0.13	0.29	0.26	-0.58
start	0.33	0.70	0.11	0.05	0.81	0.21	0.30	0.59	0.22	0.27	0.78	0.73	0.24	0.64	-0.57

C. intestinalis muscle specific promoter model architecture model

Proximal region LLR values

	Cin_M1+	Cin_M1-	Cin_M2+	Cin_M2-	Cin_M7+	Cin_M7-	Cin_M5+	Cin_M5-	Cin_M10+	Cin_M10-	Cin_M3+	Cin_M3-	Cin_M4+	Cin_M4-	Cin_M9+	Cin_M9-	Cin_M6+	Cin_M6-	Cin_M8+	Cin_M8-	end
Cin_M1+	-0.20	0.02	0.41	0.27	0.47	0.07	0.05	-0.08	0.38	0.06	0.42	0.48	0.39	0.05	0.44	0.34	0.24	0.20	0.43	0.70	-0.66
Cin_M1-	-0.12	-0.17	0.48	0.53	0.86	0.44	0.46	0.17	0.31	0.31	0.59	0.48	0.32	0.22	0.27	0.36	0.19	-0.28	0.46	0.42	-0.77
Cin_M2+	0.11	-0.04	-0.07	0.41	0.29	0.36	0.40	0.15	0.62	0.32	0.34	0.54	0.11	0.34	0.29	0.34	0.54	0.36	0.46	0.19	-1.08
Cin_M2-	0.45	0.11	0.21	0.49	0.31	0.29	0.08	0.00	0.24	0.49	0.34	0.49	0.34	0.45	0.49	0.59	0.34	0.24	0.21	0.17	-1.08
Cin_M7+	0.05	0.07	0.29	0.46	0.21	0.46	0.29	0.27	0.21	0.19	0.25	0.62	0.21	0.48	0.34	0.37	0.75	0.16	0.27	0.04	-0.68
Cin_M7-	0.27	0.25	0.39	0.42	0.22	0.25	0.20	0.17	0.35	0.44	0.44	0.33	0.50	0.53	0.50	0.20	0.42	0.39	0.17	0.31	-0.80
Cin_M5+	0.16	0.28	0.76	0.67	0.25	0.11	-0.09	0.20	0.28	0.21	0.57	0.39	0.18	0.35	0.24	0.17	0.44	0.30	0.25	0.25	-0.54
Cin_M5-	0.16	0.18	0.34	0.56	0.76	0.46	0.38	-0.03	0.43	0.34	0.61	0.61	0.48	-0.10	0.35	0.29	0.73	0.27	0.27	0.15	-1.08
Cin_M10+	0.54	0.35	0.25	0.71	0.14	0.14	0.35	-0.09	0.46	0.05	0.30	0.30	0.09	0.08	0.53	0.21	0.44	0.38	0.54	0.23	-1.22
Cin_M10-	-0.03	0.32	0.42	0.58	0.52	0.32	0.26	0.10	0.27	0.27	0.40	0.55	0.36	0.27	0.32	0.15	0.47	0.27	0.34	0.22	-1.04
Cin_M3+	0.23	0.38	0.46	0.76	0.43	0.35	0.29	0.01	0.61	0.35	0.13	0.40	0.35	0.49	0.40	0.23	0.35	0.91	-0.29	0.21	-0.72
Cin_M3-	0.43	0.19	0.40	0.53	0.67	0.25	0.23	0.25	0.43	0.43	0.30	0.37	0.51	0.50	0.11	0.13	0.46	0.88	-0.29	0.14	-0.75

Cin_M4+	0.35	0.14	0.32	0.44	0.41	0.15	0.29	0.10	0.19	0.36	0.36	0.32	0.20	0.41	0.05	0.27	0.25	0.41	0.35	0.14	-0.84
Cin_M4-	0.23	0.48	0.27	0.69	0.46	0.29	0.29	0.13	0.32	0.16	0.32	0.52	0.29	0.32	0.44	0.30	0.49	0.49	0.25	0.00	-0.84
Cin_M9+	0.42	0.05	0.37	0.32	0.32	0.16	0.15	0.22	0.43	0.12	0.37	0.34	0.03	0.32	0.56	0.15	0.42	0.13	0.35	0.16	-0.75
Cin_M9-	0.07	0.15	0.25	0.22	0.49	0.23	0.35	0.03	0.31	0.18	0.30	0.55	0.22	0.28	0.09	0.35	0.79	0.31	0.13	0.53	-0.84
Cin_M6+	0.27	0.58	0.18	0.44	0.34	0.34	0.12	0.04	0.16	0.25	0.21	0.25	0.76	0.23	0.25	0.31	0.21	0.14	0.42	0.10	-1.15
Cin_M6-	0.21	0.15	0.10	0.16	0.19	0.45	0.46	0.13	0.37	0.22	0.31	0.61	0.31	0.46	0.26	0.18	0.37	0.22	0.29	0.11	-1.08
Cin_M8+	0.45	0.09	0.14	0.09	0.29	0.39	0.12	-0.03	0.42	0.37	0.63	0.55	0.38	0.39	0.14	0.13	0.07	0.07	-0.04	0.30	-0.73
Cin_M8-	0.53	0.47	0.29	0.54	0.34	0.21	0.27	0.49	0.24	0.36	-0.32	-0.46	0.27	-0.23	0.23	0.71	0.38	0.39	0.15	0.36	-1.06
start	0.41	-0.17	0.02	0.15	0.26	0.30	-0.15	-0.20	0.52	-0.09	0.34	0.36	-0.07	-0.02	0.35	0.30	0.42	0.42	0.25	0.11	-1.21

Distal region LLR values

	Cin_M1+	Cin_M1-	Cin_M2+	Cin_M2-	Cin_M7+	Cin_M7-	Cin_M5+	Cin_M5-	Cin_M10+	Cin_M10-	Cin_M3+	Cin_M3-	Cin_M4+	Cin_M4-	Cin_M9+	Cin_M9-	Cin_M6+	Cin_M6-	Cin_M8+	Cin_M8-	end
Cin_M1+	-0.58	-0.06	0.17	0.27	0.15	0.40	0.05	0.03	0.05	0.04	0.36	0.34	0.11	0.44	0.16	0.04	0.32	0.21	0.32	-0.01	-0.82
Cin_M1-	-0.71	-0.28	0.13	0.33	0.23	0.10	0.34	0.04	0.47	0.20	0.42	0.27	0.38	0.13	0.03	0.06	0.12	-0.07	0.15	0.43	-0.48
Cin_M2+	-0.18	-0.23	0.29	0.08	-0.05	-0.11	0.05	0.08	0.18	-0.11	0.43	0.34	0.22	0.01	0.28	-0.16	-0.07	0.35	0.11	-0.25	-0.49
Cin_M2-	-0.04	0.00	0.04	0.13	0.17	0.13	-0.04	-0.05	0.10	0.16	0.16	0.16	0.24	0.24	0.15	0.10	0.30	0.01	0.18	0.24	-0.76
Cin_M7+	-0.14	-0.13	0.33	0.10	-0.07	0.30	-0.13	-0.07	0.28	0.00	0.32	0.15	0.07	0.25	0.02	-0.07	0.13	0.24	0.05	-0.10	-0.33
Cin_M7-	-0.15	-0.16	0.53	0.21	0.05	0.03	-0.09	-0.18	-0.06	-0.01	0.01	0.34	0.35	0.29	-0.06	-0.04	0.14	-0.01	0.03	-0.14	-0.19
Cin_M5+	-0.27	-0.21	0.51	0.06	0.13	-0.03	0.16	-0.15	0.43	-0.04	0.61	0.07	0.11	0.21	-0.21	-0.10	0.01	0.26	-0.13	0.05	-0.55
Cin_M5-	-0.01	0.25	0.06	0.25	0.15	0.08	0.03	0.04	-0.02	0.02	0.52	0.29	0.05	0.05	0.28	-0.09	0.17	0.14	0.12	-0.13	-0.58
Cin_M10+	0.20	-0.10	0.01	0.12	0.02	0.08	-0.05	-0.09	0.21	-0.03	0.15	0.22	0.03	0.13	-0.02	0.21	0.49	0.29	0.35	0.16	-0.66
Cin_M10-	-0.10	-0.07	0.07	0.23	0.17	0.05	0.30	0.04	0.03	-0.05	0.21	0.27	0.00	0.43	0.29	0.20	0.17	0.18	0.09	0.04	-0.87
Cin_M3+	0.12	0.08	0.18	0.49	0.46	0.16	0.12	0.00	0.07	0.15	-0.02	0.34	0.16	0.18	0.04	0.31	0.22	0.23	-0.37	0.00	-0.80
Cin_M3-	0.25	-0.03	0.31	0.10	0.24	0.01	0.06	0.04	0.35	0.12	0.11	0.09	0.04	0.22	0.05	0.07	0.25	0.18	-0.07	0.19	-0.84
Cin_M4+	-0.24	0.30	0.25	0.29	0.09	0.22	0.24	-0.07	-0.03	0.32	0.10	0.21	-0.11	0.20	-0.02	0.03	0.20	0.05	-0.24	0.39	-0.58
Cin_M4-	-0.05	-0.07	0.08	0.17	0.36	0.16	-0.03	-0.11	0.10	-0.03	0.26	0.54	0.10	0.02	0.36	0.27	0.25	0.05	-0.03	-0.10	-0.59
Cin_M9+	0.17	0.16	0.13	0.27	0.19	0.12	-0.10	0.20	0.02	-0.04	0.21	0.14	0.10	0.23	0.10	0.04	0.45	0.05	0.01	0.19	-0.90
Cin_M9-	-0.09	-0.07	0.41	0.24	0.11	0.12	-0.06	0.17	0.07	0.08	0.17	0.32	0.14	0.21	-0.02	-0.14	0.26	0.15	0.04	0.01	-0.39
Cin_M6+	-0.08	0.38	0.10	0.23	0.08	0.25	0.03	-0.04	0.12	0.36	0.14	0.14	0.32	0.15	-0.06	0.00	0.05	-0.04	0.09	0.02	-0.88
Cin_M6-	0.02	-0.15	0.19	0.09	0.53	0.42	0.27	-0.07	0.21	0.01	0.17	0.21	0.11	0.04	0.04	0.04	0.14	0.02	-0.03	-0.16	-0.89
Cin_M8+	-0.18	-0.01	-0.09	0.03	0.28	0.07	0.35	-0.10	-0.01	0.03	0.02	0.69	-0.04	0.10	0.34	0.01	0.07	0.05	-0.03	0.17	-0.87
Cin_M8-	-0.06	0.30	-0.11	0.15	0.56	0.40	0.52	0.07	0.12	0.42	-0.50	-0.21	0.39	-0.06	0.02	0.08	0.23	-0.10	-0.14	-0.10	-0.79
start	0.14	-0.30	0.32	-0.02	-0.18	0.36	0.40	0.14	-0.18	0.03	-0.04	0.30	-0.18	0.19	-0.27	-0.24	0.27	0.30	-0.18	-0.29	-0.62

Section 5: Lists of Top Scoring Genes

In this section we show the top 100 scoring genes for both models. The rank, gene name/ID and transcript name/ID, and score (calculated as described in the main text) is given. For some genes a note is given. Notice how alternative transcripts of genes can have the same score in case their promoter region is the same. Alternative transcripts of input genes are all considered as input data. In the list of the *C. intestinalis* model the 4 genes that were experimentally verified by *in situ* hybridization are indicated as well.

C. elegans pharyngeal muscle promoter architecture model

Rank	Sequence name	Transcript Name	Score	Note
1	F47G6.1	F47G6.1	5.29	input
2	C14A6.8	C14A6.8	3.68	
3	Y38C1AB.1	Y38C1AB.1	3.52	
4	Y38C1AB.5	Y38C1AB.5	3.52	
5	W03F8.1	W03F8.1.2	3.23	input
6	W03F8.1	W03F8.1.1	3.23	input
7	C49F8.2	C49F8.2	3.04	input
8	Y41E3.13	Y41E3.13	2.88	
9	F12F3.1	F12F3.1a.2	2.87	input
10	C18H2.2	C18H2.2	2.81	
11	K11C4.5	K11C4.5	2.79	input
12	C46C2.5	C46C2.5	2.54	
13	ZK1086.1	ZK1086.1b	2.53	input
14	ZK1086.1	ZK1086.1a	2.53	input
15	C36B7.7	C36B7.7	2.48	input
16	F40A3.7	F40A3.7	2.39	
17	F46H5.3	F46H5.3b.1	2.36	
18	F46H5.3	F46H5.3b.2	2.36	
19	R11A5.2	R11A5.2	2.35	input
20	T06A10.4	T06A10.4	2.34	
21	M03C11.4	M03C11.4	2.34	
22	Y73B6A.1	Y73B6A.1	2.32	
23	K12G11.1	K12G11.1	2.30	
24	F32D1.5	F32D1.5.2	2.25	
25	C09D1.1	C09D1.1f	2.24	input
26	C09D1.1	C09D1.1e	2.24	input
27	C09D1.1	C09D1.1g	2.24	input
28	C09D1.1	C09D1.1a	2.24	input
29	C09D1.1	C09D1.1b	2.24	input
30	Y24D9A.4	Y24D9A.4a.2	2.23	
31	Y60A3A.5	Y60A3A.5	2.23	
32	R02D5.6	R02D5.6	2.23	
33	W03G9.1	W03G9.1.2	2.16	
34	W03G9.1	W03G9.1.1	2.16	
35	F18G5.5	F18G5.5	2.12	
36	T24B8.4	T24B8.4	2.06	
37	Y43F8B.12	Y43F8B.12	2.06	
38	F21C3.6	F21C3.6	2.06	
39	Y39B6A.1	Y39B6A.1.2	2.06	
40	Y39B6A.1	Y39B6A.1.1	2.06	
41	F46C5.4	F46C5.4	2.04	
42	R90.1	R90.1	2.02	
43	C03B8.3	C03B8.3	2.02	
44	ZC247.1	ZC247.1	2.00	
45	B0554.4	B0554.4	1.99	

46	F25D7.5	F25D7.5	1.95	
47	F52C9.8	F52C9.8e	1.94	
48	C18B2.1	C18B2.1	1.94	
49	K01A12.2	K01A12.2	1.92	
50	T13C5.1	T13C5.1a	1.91	
51	C06E1.1	C06E1.1	1.91	
52	T05C1.2	T05C1.2	1.90	
53	C25G6.3	C25G6.3	1.90	
54	F44A6.1	F44A6.1b	1.89	input
55	F44A6.1	F44A6.1a	1.89	input
56	F32H2.5	F32H2.5	1.87	
57	F13H8.9	F13H8.9	1.87	
58	C48C5.3	C48C5.3	1.86	
59	R09H3.1	R09H3.1	1.84	
60	Y97E10AR.6	Y97E10AR.6	1.83	
61	W03D2.6	W03D2.6	1.81	
62	K09E3.5	K09E3.5	1.80	
63	C17F3.1	C17F3.1	1.80	
64	F36F12.3	F36F12.3	1.80	
65	C08G5.2	C08G5.2	1.78	
66	F13D12.6	F13D12.6.1	1.78	
67	F13D12.6	F13D12.6.2	1.78	
68	F25H8.6	F25H8.6	1.78	
69	M01D7.7	M01D7.7b	1.76	input
70	K02B2.6	K02B2.6	1.75	
71	Y47G6A.15	Y47G6A.15b	1.74	
72	D1081.2	D1081.2	1.73	
73	H32C10.2	H32C10.2	1.72	
74	M02A10.2	M02A10.2	1.72	
75	F22B7.9	F22B7.9	1.71	
76	C54F6.6	C54F6.6	1.71	
77	T20H4.5	T20H4.5.2	1.71	
78	T20H4.5	T20H4.5.1	1.71	
79	C53C11.3	C53C11.3	1.69	
80	F10E9.6	F10E9.6a.1	1.68	
81	F10E9.6	F10E9.6a.2	1.68	
82	C06G4.4	C06G4.4	1.68	
83	ZK652.8	ZK652.8	1.66	
84	C36E6.5	C36E6.5.2	1.63	
85	C36E6.5	C36E6.5.1	1.63	
86	R07B1.1	R07B1.1	1.63	
87	Y53C12B.5	Y53C12B.5b	1.63	
88	Y37E11AL.9	Y37E11AL.9	1.63	
89	C32E8.11	C32E8.11	1.62	
90	F13B12.5	F13B12.5	1.62	
91	Y69A2AR.9	Y69A2AR.9	1.62	
92	Y51A2A.1	Y51A2A.1	1.59	
93	H22D14.1	H22D14.1	1.59	
94	E04F6.6	E04F6.6	1.58	
95	F55A11.6	F55A11.6b	1.58	
96	F26B1.7	F26B1.7	1.58	
97	Y65B4BL.1	Y65B4BL.1	1.58	
98	C04G2.10	C04G2.10	1.57	
99	C36B7.1	C36B7.1	1.57	
100	Y48G1C.12	Y48G1C.12	1.57	

C. intestinalis muscle specific promoter model architecture model

Rank	Gene ID	Transcript ID	Score	Note
1	ENSCING00000007550	ENSCINT00000015480	7.03	input Ci-MA1
2	ENSCING00000007219	ENSCINT00000014813	6.21	input Ci-MRLC4
3	ENSCING00000009772	ENSCINT00000019903	6.19	similar to input Ci-MRL4
4	ENSCING00000009772	ENSCINT00000019909	6.19	input Ci-MHC4
5	ENSCING00000009772	ENSCINT00000019905	5.74	similar to input Ci-MHC4
6	ENSCING00000004818	ENSCINT00000009977	4.90	input Ci-MLC4
7	ENSCING00000004811	ENSCINT00000009959	4.55	input Ci-MLC3
8	ENSCING00000002068	ENSCINT00000004215	4.32	input Ci-MA2
9	ENSCING00000004811	ENSCINT00000009951	3.91	input Ci-MLC2
10	ENSCING00000012638	ENSCINT00000017888	3.70	similar to input Ci-MA4
11	ENSCING00000012638	ENSCINT00000014519	3.70	input Ci-MA4
12	ENSCING00000002749	ENSCINT00000005614	3.69	input Ci-MRLC2
13	ENSCING00000002749	ENSCINT00000024572	3.26	input Ci-MRLC3
14	ENSCING00000001218	ENSCINT00000002395	3.06	similar to input Ci-MHC2
15	ENSCING00000001218	ENSCINT00000002399	3.06	input Ci-MHC2
16	ENSCING00000001218	ENSCINT00000002389	3.06	similar to input Ci-MHC2
17	ENSCING00000004282	ENSCINT00000008881	3.01	input Ci-TPM3
18	ENSCING00000006130	ENSCINT00000017917	2.78	
19	ENSCING00000006130	ENSCINT00000012677	2.74	
20	ENSCING00000014374	ENSCINT00000026270	2.51	tested by ISH
21	ENSCING00000005333	ENSCINT00000010963	2.45	input Ci-TnT
22	ENSCING00000005333	ENSCINT00000024866	2.45	similar to input Ci-TnT
23	ENSCING00000004818	ENSCINT00000009978	2.14	similar to input Ci-MLC4
24	ENSCING00000015577	ENSCINT00000027760	2.01	
25	ENSCING00000013146	ENSCINT00000024481	1.99	
26	ENSCING00000001377	ENSCINT00000002686	1.98	input Ci-MA5
27	ENSCING00000011503	ENSCINT00000022194	1.92	
28	ENSCING00000011776	ENSCINT00000022592	1.89	
29	ENSCING00000004103	ENSCINT00000022966	1.83	
30	ENSCING00000017023	ENSCINT00000029353	1.80	
31	ENSCING00000003268	ENSCINT00000006714	1.72	tested by ISH
32	ENSCING00000004994	ENSCINT00000010274	1.72	
33	ENSCING00000002305	ENSCINT00000004703	1.70	
34	ENSCING00000002305	ENSCINT00000024273	1.70	
35	ENSCING00000015038	ENSCINT00000027164	1.70	
36	ENSCING00000014058	ENSCINT00000025813	1.63	
37	ENSCING00000010768	ENSCINT00000009260	1.61	
38	ENSCING00000000960	ENSCINT00000001733	1.58	
39	ENSCING00000009408	ENSCINT00000019138	1.56	
40	ENSCING00000008083	ENSCINT00000016536	1.55	
41	ENSCING00000003631	ENSCINT00000007469	1.55	tested by ISH
42	ENSCING00000006130	ENSCINT00000023457	1.53	
43	ENSCING00000013183	ENSCINT00000024546	1.53	
44	ENSCING00000000583	ENSCINT00000001070	1.51	
45	ENSCING00000004434	ENSCINT00000009154	1.48	
46	ENSCING00000005715	ENSCINT00000011803	1.46	
47	ENSCING00000014525	ENSCINT00000018109	1.44	
48	ENSCING00000016494	ENSCINT00000028789	1.43	
49	ENSCING00000012527	ENSCINT00000023589	1.41	
50	ENSCING00000003992	ENSCINT00000008325	1.41	tested by ISH
51	ENSCING00000005571	ENSCINT00000011559	1.40	
52	ENSCING00000011769	ENSCINT00000022587	1.38	
53	ENSCING00000007847	ENSCINT00000016086	1.37	
54	ENSCING00000002349	ENSCINT00000004783	1.36	input TPM1

55	ENSCING00000017243	ENSCINT00000009056	1.36	similar to input Ci-TPM1
56	ENSCING00000017243	ENSCINT00000009054	1.36	similar to input Ci-TPM1
57	ENSCING00000016534	ENSCINT00000028828	1.36	
58	ENSCING00000007682	ENSCINT00000015757	1.35	
59	ENSCING00000006235	ENSCINT00000012854	1.33	input Ci-TnC2
60	ENSCING00000002478	ENSCINT00000005015	1.30	
61	ENSCING00000000231	ENSCINT00000000411	1.29	
62	ENSCING00000006972	ENSCINT00000014308	1.29	
63	ENSCING00000003092	ENSCINT00000009429	1.28	
64	ENSCING00000014909	ENSCINT00000010165	1.28	
65	ENSCING00000015814	ENSCINT00000005395	1.23	
66	ENSCING00000016759	ENSCINT00000029082	1.22	
67	ENSCING00000015195	ENSCINT00000027332	1.22	
68	ENSCING00000007638	ENSCINT00000015655	1.21	
69	ENSCING00000009244	ENSCINT00000018782	1.21	
70	ENSCING00000007031	ENSCINT00000022807	1.20	
71	ENSCING00000014877	ENSCINT00000026958	1.18	
72	ENSCING00000001706	ENSCINT00000003438	1.18	
73	ENSCING00000015680	ENSCINT00000027867	1.16	
74	ENSCING00000012882	ENSCINT00000024108	1.14	
75	ENSCING00000008148	ENSCINT00000016638	1.12	
76	ENSCING00000008082	ENSCINT00000016530	1.11	
77	ENSCING00000009410	ENSCINT00000019143	1.11	
78	ENSCING00000008422	ENSCINT00000017161	1.10	
79	ENSCING00000006339	ENSCINT00000013066	1.10	
80	ENSCING00000006339	ENSCINT00000025838	1.10	
81	ENSCING00000006339	ENSCINT00000025839	1.10	
82	ENSCING00000010210	ENSCINT00000020348	1.09	
83	ENSCING00000016710	ENSCINT00000029027	1.08	
84	ENSCING00000001030	ENSCINT00000023342	1.08	
85	ENSCING00000009771	ENSCINT00000028021	1.07	
86	ENSCING00000009745	ENSCINT00000019797	1.07	
87	ENSCING00000009556	ENSCINT00000019441	1.06	
88	ENSCING00000009483	ENSCINT00000019315	1.05	
89	ENSCING00000004449	ENSCINT00000009192	1.05	
90	ENSCING00000005640	ENSCINT00000011687	1.04	
91	ENSCING00000006115	ENSCINT00000012634	1.04	
92	ENSCING00000005361	ENSCINT00000011013	1.04	
93	ENSCING00000005361	ENSCINT00000004019	1.04	
94	ENSCING00000016259	ENSCINT00000006242	1.03	
95	ENSCING00000014032	ENSCINT00000009663	1.03	
96	ENSCING00000013643	ENSCINT00000025207	1.03	
97	ENSCING00000005896	ENSCINT00000012192	1.02	
98	ENSCING00000014797	ENSCINT00000026832	1.00	
99	ENSCING00000006856	ENSCINT00000014081	1.00	
100	ENSCING00000011462	ENSCINT00000022134	0.99	

References

1. Bailey, T. L. and Elkan, C., 1994, Fitting a mixture model by expectation maximization to discover motifs in biopolymers, *Proc Int Conf Intell Syst Mol Biol*, 2, 28-36.

2. Pavesi, G., Mauri, G. and Pesole, G., 2001, An algorithm for finding signals of unknown length in DNA sequences, *Bioinformatics*, 17 Suppl 1, S207-14.

3. Hughes, J. D., Estep, P. W., Tavazoie, S. and Church, G. M., 2000, Computational identification of cis-regulatory elements associated with groups of functionally related genes in Saccharomyces cerevisiae, *J Mol Biol*, 296, 1205-14.