SUPPLEMENTARY MATERIALS:

-Supplementary Discussion

-Supplementary Materials and Methods

-Supplementary Figure Legends

-Supplementary Figures

-Supplementary Tables

-Supplementary References

SUPPLEMENTARY DISCUSSION:

The analyses presented in this study have inherent sensitivity/specificity trade-offs and threshold choices that affect the composition of gene groups with specific behaviours (e.g. HLH-1 occupied, muscle preferred expression, *unc-120* regulated, etc.). Statistical measures provided confidence that different distributions within a data type differ significantly from each other, while the major trends and comparisons involving multiple data types were stable to reasonable shifts in thresholds. In the case of HLH-1 occupancy, we also compared the set of regions detected in this ChIP-seq study with ChIP-chip using the same antibody (Lei et al. 2010) and with ChIP-seq performed with a tagged transgenic HLH-1 (Gerstein et al. 2010). The former study differs from ours because they over-expressed HLH-1 protein to elevate the ChIP signal, which might alter the occupancy map. The latter study apparently had lower sensitivity due to the low fraction of tissue that was BWM. When we processed their called regions identically to our data, the intersections were encouragingly shared at the prominent ChIP sites. As expected, concordance decreases for the more modest signals, and there is no expectation of agreement in

regimes where the less sensitive studies lacked signal. Similarly, our BWM expressed gene list had statistically significant overlaps with prior gene lists of embryonic muscle-enriched RNA datasets ($p<0.001$) (Fukushige et al. 2006; Fox et al. 2008). Differences between the datasets likely arise in part from differences in techniques and stage, as our embryos were during late differentiation and we did not over-express any transcription factors. Our non-BWM expressed gene list had a statistically significant overlap with existing non-muscle datasets ($p<0.01$) (Fox et al. 2005; Von Stetina et al. 2007).

SUPPLEMENTARY MATERIALS AND METHODS:

There are two genetic strategies known to increase the proportion of muscle, which was necessary to increase the signal. We avoided gene overexpression, due to our interest in the *hlh-1* mutant and concern it would alter the muscle differentiation regulatory circuit itself (Fox et al. 2008), leading to uncertain changes in the composition and behavior of target genes. Instead, we knocked-down early specification genes to permit muscle specification. Knocking down *mex-3* in the embryo causes each granddaughter of the AB lineage to divide like the C lineage, thus producing ~80 BWM cells rather than 1 BWM cell, 3 NSM cells, and most of the pharynx (Sulston et al. 1983; Draper et al. 1996; Hunter and Kenyon 1996). Knockdown of *skn-1* causes the EMS lineage to adopt a C-like fate (Bowerman et al. 1992; Blackwell et al. 1994), thus abolishing the 4th NSM cell and the M-cell lineage (the source of all post-embryonic muscle) (Sulston et al. 1983). This also converts MS-derived BWM into C-like BWM. Knocking down *elt-1* – the hypodermal master regulator – causes hypodermal cells in the C and C-like lineages to adopt a mesodermal muscle specification (Michaux et al. 2001).

Also, by driving muscle enrichment in an embryo rather than in a cell culture, our hope was to not activate accessory factors involved in stress response. Nevertheless, there are always consequences to enriching for muscle. At the more extreme end, because the animal is not normal, it will become necrotic sooner than a wild-type embryo. We avoided this side effect by using a relatively early developmental time point prior to hatching. Muscle-enriched embryos might also display unwanted consequences of excess muscle, such as cytokine and signalling imbalances. Despite these muscle isolating caveats, we believe the nuclear enrichment to both be necessary and useful given the notably increased signal.

Harvesting worms. Gravid adults were washed off plates and bleached to obtain young embryos. The embryos were then shaken in S-complete medium at a density of 5 embryos/µL for 400 minutes. The embryos were spun down and prepared for the desired set of observations.

Motif finding. To determine what HLH-1 may be binding and its associated factors, we looked at which motifs are enriched near the ChIP-identified binding sites. We utilized multiple motif-finding algorithms on sequences within various radii of the binding site. Both a greedy motif-finding algorithm and MEME found similar motifs to be overrepresented in the sequences, depending on the size of the radius utilized. This is not unexpected, given the statistical impact that varying the sequence volume has on motif-finders. The different motifs represent both the actual binding motif for the HLH-1 transcription factor and possible associated binding sites important for accessory transcription factor binding. The prevalence of the motifs outside of the HLH-1 targeted regions varies between the different motifs. Several of the motifs, such as the E-boxes, are targeted by multiple transcription factors with non-overlapping expression patterns (Krause et al. 1997; Grove et al. 2009). Given that these motifs depend on as little as 6 bases, it is possible that across 100 million random base pairs (the length of the *C. elegans* genome) that

they could appear over 24,000 times by chance. Starting with a PSFM (position-specific frequency matrix) as the reference motif, a 95% match guarantees an essentially perfect match (100% matches are generally impossible given the variation within the reference motif). An 85% match, depending on the transcription factor, may or may not be a real motif. A scan through the genome for matches to each of these motifs at 85% and 95% identity provides the baseline frequency. The number of motifs identified within the anti-HLH-1 ChIP identified regions was also determined at the different thresholds.

SUPPLEMENTARY FIGURE LEGENDS

Figure S1: The impact of mutation on gene expression levels

(A) The expression of a gene, *rpb-6*, enriched in wild-type animals (top 4 lanes) across RNAi conditions but with decreased expression in the *hlh-1* mutant (bottom 3 lanes) (B) The expression of the nematode Twist, *hlh-8*, is shown. Its expression is increased primarily in the muscle enriched mutant animals, with the muscle-enriched wild-type animals having less expression than the muscle-normal wild-type. (C) The impact of mutation on *hlh-1* regulated genes is shown in relation to the expression levels in wild type animals. Genes whose expression levels do not significantly change are not shown. (D) The expression levels of annotated genes from different tissues across the nematode are shown. BWM has clearly the strongest enrichment. The expression level of each gene is normalized to the no RNAi condition and then each RNAi condition is mean centered to correct for expression differences between samples. Genes included in each group are representative of the group, based on Wormbase annotations, but are not exclusively expressed in those tissues. The proportion of hypodermis and intestine appear to rise with triple RNAi in part due to gene overlap and in part because as other tissues disappear, the proportional representation of the tissues that remain increases. It is also possible that some tissues are enriched specifically in the triple RNAi treatment, which is why it was combined with the *mex-3* RNAi treatment for our analyses.

Figure S2: HLH-1 binding results are reproducible and supported by independent data

The correlation of HLH-1 occupancy signal between different ChIP-seq analyses using two muscle-enrichment conditions is shown in blue (triple RNAi knock-down) and red (*mex-3* RNAi). (A) The binding sites for two muscle troponins (*pat-10*, and *tnt-3*) are shown across the

two independent ChIP experiments. (B) The correlation of our observed ChIP binding with previously published regulatory regions (shown as red boxes) is shown for three well-studied genes (*myo-3*, *unc-54*, and *hlh-1*).

Figure S3: Association of motif locations with other regulatory data

(A) The proximity of various motifs to ChIP-defined peaks is shown across four conditions. (B) The relative enrichment of the motifs in each category is also shown. (C) The conservation of motifs in relation to background and their corresponding ChIP-seq peaks is shown.

Figure S4: Relative motif enrichment between genes in different regulatory classes

The enrichment of HLH-1 region-associated motifs across different regulatory conditions is shown. Yellow represents an enrichment of motifs from the x-axis condition over the y-axis condition. Blue represents a depletion of the motif from the x-axis condition over the y-axis condition. No enrichment or depletion of the GTGTACTCCTCTCGG motif was found across the tested conditions. The CnnTnGAGCGCnnTT motif is enriched near *unc-120* negatively regulated genes. The CGnnGCGAGACCC motif is enriched near genes up-regulated by *hlh-1* and down-regulated by *unc-120*.

SUPPLEMENTARY TABLES

Table S1: Transcription factors regulated by *hlh-1* and *unc-120* and their expression domains

Gene	Positively Regulated by:		Negatively Regulated by:		Regional Expression		
	hlh-1	*unc-120*	*hlh-1*	*unc-120*	Enriched in BWM	Broadly Expressed	Absent in BWM
arx-6		x				x	
atf-5		x				x	
atf-6		x				x	
C25H3.6		x				x	
C49C3.5	x	x				x	
cbp-1		x			x		
ceh-13		x				x	
ceh-14			x		x		
ceh-21				x	x		
ceh-22		x				x	
ceh-26		x				x	
ceh-31		x				x	
ceh-32		x				x	
ceh-34		x				x	
daf-19			x		x		
daf-3		x			x		
daf-8		x				x	
dpy-22		x			x		
dsc-1			x				x
dve-1		x				x	
egl-27		x			x		
egl-38		x				x	
egl-43		x	x			x	
egl-5		x				x	
elt-6			x			x	
F17A9.3		x				x	
F21A10.2		x			x		
F33E11.2	x	x			x		
F41H10.4		x				x	
F47G9.4	x	x				x	
F52C12.4		x			x		
F53F8.1		x				x	
F54D5.11		x				x	
fos-1		x			x		

gfl-1		**x**				**x**	
hbl-1		**x**			**x**		
hif-1		**x**			**x**		
hlh-10			**x**				**x**
hlh-33			**x**			**x**	
hlh-8		**x**	**x**			**x**	
hnd-1		**x**				**x**	
icd-1				**x**		**x**	
K01H12.1				**x**		**x**	
K04C1.2		**x**				**x**	
K11D2.4				**x**	**x**		
K11H12.8				**x**	**x**		
lag-1		**x**				**x**	
lin-26		**x**				**x**	
lin-36		**x**				**x**	
lin-49		**x**			**x**		
lin-59			**x**		**x**		
ltd-1		**x**				**x**	
mdl-1		**x**				**x**	
mex-3			**x**	**x**		**x**	
mex-5			**x**		**x**		
mex-6		**x**				**x**	
mls-1			**x**				**x**
mxl-1		**x**				**x**	
nfya-1		**x**			**x**		
nhr-104		**x**				**x**	
nhr-115				**x**		**x**	
nhr-116				**x**		**x**	
nhr-123			**x**				**x**
nhr-137		**x**				**x**	
nhr-143		**x**				**x**	
nhr-153		**x**				**x**	
nhr-163		**x**				**x**	
nhr-17		**x**				**x**	
nhr-22		**x**				**x**	
nhr-23		**x**				**x**	
nhr-25		**x**	**x**			**x**	
nhr-3			**x**			**x**	
nhr-4		**x**				**x**	
nhr-44		**x**				**x**	
nhr-46		**x**				**x**	
nhr-49		**x**				**x**	
nhr-55			**x**			**x**	
nhr-56		**x**				**x**	
nhr-59		**x**				**x**	

nhr-63			**x**			**x**	
nhr-64		**x**				**x**	
nhr-66		**x**				**x**	
nhr-68		**x**				**x**	
nhr-70				**x**	**x**		
nhr-71				**x**		**x**	
nhr-90		**x**				**x**	
nhr-91		**x**				**x**	
nhr-92		**x**				**x**	
nhr-95		**x**				**x**	
oma-1		**x**				**x**	
oma-2		**x**			**x**		
osm-1			**x**			**x**	
peb-1		**x**				**x**	
pha-1		**x**				**x**	
pos-1		**x**				**x**	
pqn-45		**x**			**x**		
pqn-47		**x**				**x**	
prkl-1		**x**				**x**	
psa-4		**x**				**x**	
R04A9.5		**x**				**x**	
ref-1		**x**			**x**		
rpm-1			**x**	**x**		**x**	
sbp-1		**x**			**x**		
sdc-1		**x**			**x**		
sdc-3		**x**			**x**		
set-17		**x**				**x**	
sox-2		**x**				**x**	
spr-3		**x**			**x**		
sur-2		**x**			**x**		
T26A5.8	**x**	**x**				**x**	
taf-10	**x**					**x**	
taf-12		**x**				**x**	
taf-3		**x**			**x**		
taf-9	**x**	**x**				**x**	
tbp-1	**x**	**x**				**x**	
tra-4			**x**		**x**		
unc-115		**x**	**x**			**x**	
unc-120				**x**		**x**	
unc-130	**x**	**x**				**x**	
unc-30		**x**				**x**	
unc-42		**x**				**x**	
unc-86		**x**				**x**	
xbp-1				**x**		**x**	
Y37E11B.2	**x**					**x**	

Y38E10A.6		**x**				**x**	
Y38F2AR.13		**x**			**x**		
Y57G11A.3		**x**			**x**		
Y75B8A.29	**x**					**x**	
Y75B8A.6		**x**				**x**	
zag-1		**x**	**x**			**x**	
zip-5		**x**			**x**		
ZK1128.4	**x**					**x**	
ZK1240.3	**x**	**x**			**x**		
ZK418.9		**x**			**x**		
zyx-1		**x**			**x**		

Table S2: Sample of genes up-regulated in muscle-enriched animals and classical muscle genes that are not enriched in muscle

2058 genes are significantly enriched in muscle-rich animals. Included here is a subset of muscle-enriched genes ($p<0.05$). However, we noted that many genes currently annotated in Wormbase as BWM are also expressed in other tissues and therefore are not expected to be enriched specifically in BWM over other tissues. A subset of these genes, including some examples of previously documented muscle structural genes (Fox et al. 2007), is given. Genes that are known muscle genes but are not upregulated in muscle-enriched animals (and therefore not muscle preferred) are denoted with a double asterisk (**), while those only upregulated by one standard deviation are denoted by a single asterisk (*).

Gene	Gene Description
act-2	**actin**
act-4	**actin**
deb-1	**vinculin (dense bodies)**
alp-1*	**alpha-actin associated/Enigma (BWM & PhM)**
dhp-2	**dihydropyrimidinase**
dim-1	**immunoglobulin-repeat (myofilament anchoring)**
egl-15	**FGF-like receptor tyrosine kinase**
egl-19	**alpha subunit of mammalian L-type calcium ion channel**
egl-20	**WNT**
emb-9	**basement membrane collagen**
epi-1	**laminin alpha chain**

frm-5	**FERM domain (cell adhesion)**
let-2	**alpha-2 type IV collagen**
lev-11	**tropomyosin**
lin-1	**ETS transcription factor**
lin-2	**membrane associated guanylate kinase**
lin-39	**sex combs reduced/Hox5**
mlc-1*	**myosin light chain (BWM, NSM, & PhM)**
mlc-2*	**myosin light chain (BWM, NSM, & PhM)**
mlc-3*	**myosin light chain (BWM, NSM, & PhM)**
mup-2	**troponin T**
myo-3	**myosin heavy chain A**
pat-3	**beta-integrin subunit**
pat-4*	**integrin-linked kinase (BWM & PhM)**
pat-6*	**alpha-parvin**
pat-10*	**troponin C (BWM & NSM)**
tmd-2	**tropomodulin**
tni-1	**troponin I**
tni-3*	**troponin I (BWM, NSM, & PhM)**
tnt-2*	**troponin T (BWM & NSM)**
tnt-3	**troponin T**
unc-112	**Mitogen inducible gene- (dense bodies and M lines)**
unc-116	**kinesin-1 heavy chain**
unc-15	**paramyosin**
unc-22*	**twitchin**

unc-23	**chaperone**
unc-27*	**troponin I (BWM, NSM, & PhM)**
unc-44	**ankyrin-like protein**
unc-45	**chaperone**
unc-52	**perlecan**
unc-53	**NAV1/2/3**
unc-54	**myosin class II heavy chain**
unc-60*	**actin depolymerizing factor/cofilin (BWM & intestine)**
unc-68	**ryanodine receptor**
unc-70	**beta-spectrin**
unc-73	**guanine nucleotide exchange factor**
unc-78*	**actin-interacting protein (BWM, NSM, & PhM)**
unc-83	**transmembrane protein**
unc-87*	**calponin-like (BWM, NSM, & PhM)**
unc-89	**protein kinase (A bands)**
unc-94	**unknown (thin filaments)**
unc-95	**paxillin-related (thick and thin filaments)**
unc-96	**unknown (thick filaments)**
unc-97*	**PINCH (BWM & NSM)**
unc-98*	**zinc-finger (BWM & NSM)**
unc-120*	**MADS-box transcription factor (NSM & BWM)**

Table S3: Location of HLH-1 ChIP peak relative to nearest gene structure

HLH-1 ChIP-Seq peaks are variably located with respect to the nearest gene's anatomy. The Gene Body category below refers to both exons and introns; the Exon category includes the CDS plus the 5' and 3' UTRs; the CDS category refers to the protein coding sequence (i.e. translated exons only). The greatest enrichment in binding is in the 500 bp proximal promoter region, which we expect to be rich in proximal regulatory sites; followed by the 5'UTR; and finally by the upstream intergenic region. Other regions of the gene bodies are relatively depleted for binding.

Region	Number of peaks per kb of feature	Fold enrichment over expected
5000 bp Upstream	**0.137**	**1.5**
500 bp Upstream	**1.10**	**12**
Entire Gene Body	**0.111**	**1.2**
Exons	**0.159**	**1.7**
Introns	**0.017**	**0.2**
CDS	**0.108**	**1.2**
5' UTR	**0.065**	**0.7**
3' UTR	**0.014**	**0.2**

Table S4: Genes with nearby HLH-1 binding

Many genes showed HLH-1 binding in the gene body or in the 5' 5000 base pairs. Both *hlh-1* positively regulated genes and muscle-enriched genes are more likely to have detectable HLH-1 binding than the rest of the genome. Genes that are negatively regulated by *hlh-1* and observed to be up-regulated only in the mutant BWM are actually less likely to have HLH-1 binding than the background level.

	Number of genes with binding	Total number of genes	Percent of genes with binding
HLH-1 positively regulated genes	**473**	**837**	**56.5%**
HLH-1 positively regulated genes	**117**	**212**	**55.2%**

enriched in BWM			
UNC-120 positively regulated genes	**1437**	**2718**	**52.9%**
UNC-120 positively regulated genes enriched in BWM	**483**	**827**	**58.4%**
UNC-120 and HLH-1 positively regulated genes	**238**	**441**	**54.0%**
Genes enriched in BWM	**1169**	**2175**	**53.7%**
Genes enriched in BWM not positively regulated by HLH-1	**1015**	**1846**	**55.0%**
Genes enriched in BWM not positively regulated by UNC-120	**649**	**1231**	**52.7%**
Genes enriched in BWM not positively regulated by HLH-1 or UNC-120	**602**	**1145**	**52.5%**
Genes enriched in *hlh-1* mutants	**265**	**557**	**47.6%**
Genes enriched only in *hlh-1* mutant muscle	**119**	**308**	**38.6%**
All genes with no expression change	**6501**	**13477**	**48.2%**
All genes not enriched in BWM	**8350**	**18173**	**45.9%**
All genes depleted in BWM	**1849**	**4696**	**39.4%**
All genes not positively regulated by *hlh-1*	**8935**	**19278**	**46.3%**
All genes negatively regulated by *hlh-1*	**2434**	**5801**	**42.0%**
BWM genes (Wormbase)	**562**	**941**	**59.7%**
NSM genes (Wormbase)	**502**	**876**	**57.3%**

Table S5: Synthetic PAT Scoring

Transcription factors were screen for synthetic paralysis at the two-fold stage (PAT) using RNAi feeding in *hlh-1(cc561)* mutant animals. Several genes gave significant increases in the phenotype in the mutant background, including the genes *lin-39*, *grh-1*, and *ceh-20*. Shown are the percentage of PAT phenotype seen in the screen and the significance.

RNAi feeding	**Wild type (N2)**	***hlh-1(cc561)***
No RNAi (HT115)	**0% (0/411)**	**5% (4/88)**

***ceh-20* RNAi**	**0% (0/161)**	**31% (57/121)**
***lin-39* RNAi**	**1% (1/112)**	**27% (27/101)**
***grh-1* RNAi**	**0% (0/250)**	**25% (23/93)**
***nhr-63* RNAi**	**1% (2/141)**	**27% (27/99)**
***hnd-1* RNAi**	**2% (5/208)**	**18% (9/49)**
***ceh-51* RNAi**	**8% (14/160)**	**21% (39/186)**
***unc-120* RNAi**	**2% (12/589)**	**36% (70/194)**
***nhr-4* RNAi**	**3% (3/115)**	**39% (37/95)**
***nhr-116* RNAi**	**4% (5/119)**	**27% (24/90)**
***tbp-1* RNAi**	**8% (14/168)**	**70% (237/340)**
***nhr-46* RNAi**	**3% (1/30)**	**24% (18/75)**
***hmg-1.2* RNAi**	**1% (2/294)**	**23% (62/266)**
F52C12.4 RNAi	**0% (0/17)**	**30% (15/49)**
D1046.2	**5% (1/20)**	**20% (13/65)**
Y62E10A.17	**5% (1/22)**	**63% (216/346)**
ztf-25	**4% (2/51)**	**28% (16/57)**
nhr-134	**3% (5/160)**	**16% (34/216)**
sknr-1	**0% (0/109)**	**32% (17/53)**
hlh-19	**0% (0/60)**	**15% (16/106)**
sex-1	**0% (0/136)**	**25% (16/65)**
sdc-2	**0% (0/108)**	**34% (34/99)**
fkh-3	**0% (1/225)**	**19% (38/199)**
nhr-71	**0% (0/167)**	**20% (35/178)**

Table S6: Troponin family gene expression and HLH-1 binding

The association of HLH-1 binding and expression of known troponin genes is shown. Genes expressed in BWM generally have HLH-1 binding, but are not dependent on HLH-1 for expression. The only case in which troponin expression is dependent on HLH-1 expression is when it is bound by HLH-1.

Gene	Type of Troponin	BWM expression	NSM expression	Pharynx expression	HLH-1 Binding	TF Dependent
tni-1	I	Yes	Yes		Bound	*unc-120*
unc-27	I	Yes	Yes	Yes	Bound	*hlh-1*
tni-3	I	Yes	Yes	Yes		*unc-120*
tni-4	I			Yes		
pat-10	C	Yes	Yes		Bound	*hlh-1*
tnc-2	C			Yes		
mup-2	T	Yes	Yes		Bound	
tnt-2	T	Yes	Yes			
tnt-3	T	Yes		Yes	Bound	*hlh-1*
tnt-4	T		Yes			

Table S7: Archetypal List of Muscle Factors

By using four overlapping criteria, a set of 78 genes have been selected as gold standard muscle genes. They are preferentially expressed in BWM, dependent on *hlh-1* for expression, and have HLH-1 occupancy at an octa E-box.

Gene Public Name	Gene Description (Concise)
alh-8	methylmalonate-semialdehyde dehydrogenase
B0310.3	
C06H2.2	
C11G10.2	
C37E2.1	
C39D10.3	
C48E7.6	
CC8.2	
cor-1	actin-binding protein coronin
dhp-2	dihydropyrimidinase (DPYS)
dhs-11	short-chain dehydrogenase (mitochondrial)
dim-1	three immunoglobulin-like repeats in the carboxy terminus
dpy-18	alpha subunit of prolyl-4-hydroxylase
	(procollagen modifying enzyme)
E02H9.6	
etr-1	ELAV-type RNA-binding protein
F13B12.6	
F29B9.8	
F30A10.9	
F40F4.7	
F52E1.5	
F53F10.1	
fbxa-197	
inx-1	innexin family
K10C9.3	
let-2	alpha-2 type IV collagen
let-756	fibroblast growth factor (FGF)-like ligand
lev-11	tropomyosin
nas-39	
pcp-5	
pqn-42	glutamine/asparagine (Q/N)-rich ('prion') domain

pup-3	**poly(U) polymerase**
R04D3.4	
R11A5.6	
rnt-1	**RUNX family of transcription factor**
rpl-2	**large ribosomal subunit L8**
rpl-4	**large ribosomal subunit L4**
srp-1	**ovalbumin-like serpin (ov-serpin)**
str-262	
sup-12	**RNA-binding protein with RNA recognition motif (RRM)**
syg-1	**immunoglobulin**
syg-2	**immunoglobulin**
T01B4.3	
T05C1.4	**calmodulin-binding transcriptional activators**
T06D8.5	
T10G3.1	
T12D8.9	
T22E5.7	
T23G7.3	
T27F6.4	
tag-165	**methionine synthase reductase (MTRR)**
tag-264	
tag-278	
tnt-3	**troponin T**
tsp-11	
ttr-35	
ttr-36	
twk-31	**two pore-domain potassium channel family**
unc-45	**N-terminal TRP (tetratricopeptide repeat) and a C-terminal UCS (UNC-45-CRO1-She4p) domain; essential for proper thick filament formation and sarcomere organization**
W07E11.1	
W09C5.1	
W09G3.7	
Y119D3B.14	**mitochondrial translation elongation factor G2 (EFG2), a GTP-binding protein**
Y37D8A.2	
Y39E4B.7	
Y45F3A.9	
Y49F6B.9	
Y55D5A.3	
Y56A3A.36	
Y69A2AR.31	
Y73B3A.20	
zmp-1	**zinc matrix metalloproteinase**

Table S8: Sequencing read numbers

Sample Name	Conditions	Raw Reads	No Match	QC failed	Unique (0-2 mismatch)	Repeats (0-2 mismatch)
RNA-seq						
600CJAAXXs6	N2 Triple RNAi	17,518,614	2,135,996	546,864	14,582,024	3,544,477
42YHDAAXXs1	N2 *mex-3* RNAi	18,554,432	4,877,289	496,743	12,699,024	2,889,803
42YL6AAXXs2	N2 no RNAi	25,367,724	5,007,930	197,471	19,235,995	4,261,811

600CJAAXXs7	**N2 no RNAi**	**20,452,854**	**2,633,060**	**996,592**	**16,348,196**	**3,994,947**
42JU1AAXXs2	***hlh-1* Triple RNAi**	**22,907,213**	**4,372,389**	**209,962**	**17,842,900**	**4,586,323**
434ACAAXXs2	***hlh-1 mex-3* RNAi**	**21,861,515**	**4,329,400**	**362,429**	**16,066,515**	**4,597,324**
42JTNAAXXs2	***hlh-1* no RNAi**	**19,470,982**	**3,341,531**	**311,262**	**14,303,117**	**5,145,196**
62TVAAXXs2	***unc-120* Triple RNAi**	**39,992,495**	**7,216,453**	**297,589**	**30,712,030**	**13,278,254**
62TVAAXXs4	***unc-120 mex-3* RNAi**	**32,854,137**	**4,896,838**	**253,076**	**25,965,891**	**13,807,406**
62TVAAXXs3	***unc-120* no RNAi**	**42,119,911**	**7,816,120**	**310,577**	**32,158,573**	**12,442,631**
600CJAAXXs5	***unc-120* no RNAi**	**21,026,815**	**3,349,977**	**970,572**	**16,226,598**	**4,938,302**
ChIP-seq						
42R96AAXXs2	**antiHLH-1 triple RNAi**	**16,375,373**	**3,754,846**	**396,986**	**11,127,618**	**2,929,227**
42UKBAAXXs5	**antiHLH-1 *mex-3* RNAi**	**19,159,689**	**4,726,502**	**597,369**	**12,536,205**	**2,955,402**
42VNDAAXXs2	**input DNA triple RNAi**	**17,125,987**	**2,057,200**	**391,368**	**13,391,899**	**3,619,064**
42VNDAAXXs4	**input DNA *mex-3* RNAi**	**14,767,546**	**2,228,720**	**488,179**	**10,922,547**	**2,922,232**

SUPPLEMENTARY REFERENCES

Blackwell TK, Bowerman B, Priess JR, Weintraub H. 1994. Formation of a monomeric DNA binding domain by Skn-1 bZIP and homeodomain elements. *Science* **266**: 621-628.

Bowerman B, Eaton BA, Priess JR. 1992. skn-1, a maternally expressed gene required to specify the fate of ventral blastomeres in the early C. elegans embryo. *Cell* **68**: 1061-1075.

Draper BW, Mello CC, Bowerman B, Hardin J, Priess JR. 1996. MEX-3 is a KH domain protein that regulates blastomere identity in early C. elegans embryos. *Cell* **87**: 205-216.

Fox JE, Burow ME, McLachlan JA, Miller CA, 3rd. 2008. Detecting ligands and dissecting nuclear receptor-signaling pathways using recombinant strains of the yeast Saccharomyces cerevisiae. *Nat Protoc* **3**: 637-645.

Fox RM, Von Stetina SE, Barlow SJ, Shaffer C, Olszewski KL, Moore JH, Dupuy D, Vidal M, Miller DM, 3rd. 2005. A gene expression fingerprint of C. elegans embryonic motor neurons. *BMC Genomics* **6**: 42.

Fox RM, Watson JD, Von Stetina SE, McDermott J, Brodigan TM, Fukushige T, Krause M, Miller DM, 3rd. 2007. The embryonic muscle transcriptome of Caenorhabditis elegans. *Genome Biol* **8**: R188.

Fukushige T, Brodigan TM, Schriefer LA, Waterston RH, Krause M. 2006. Defining the transcriptional redundancy of early bodywall muscle development in C. elegans: evidence for a unified theory of animal muscle development. *Genes Dev* **20**: 3395-3406.

Gerstein MB Lu ZJ Van Nostrand EL Cheng C Arshinoff BI Liu T Yip KY Robilotto R Rechtsteiner A Ikegami K et al. 2010. Integrative analysis of the Caenorhabditis elegans genome by the modENCODE project. *Science* **330**: 1775-1787.

Grove CA, De Masi F, Barrasa MI, Newburger DE, Alkema MJ, Bulyk ML, Walhout AJ. 2009. A multiparameter network reveals extensive divergence between C. elegans bHLH transcription factors. *Cell* **138**: 314-327.

Hunter CP, Kenyon C. 1996. Spatial and temporal controls target pal-1 blastomere-specification activity to a single blastomere lineage in C. elegans embryos. *Cell* **87**: 217-226.

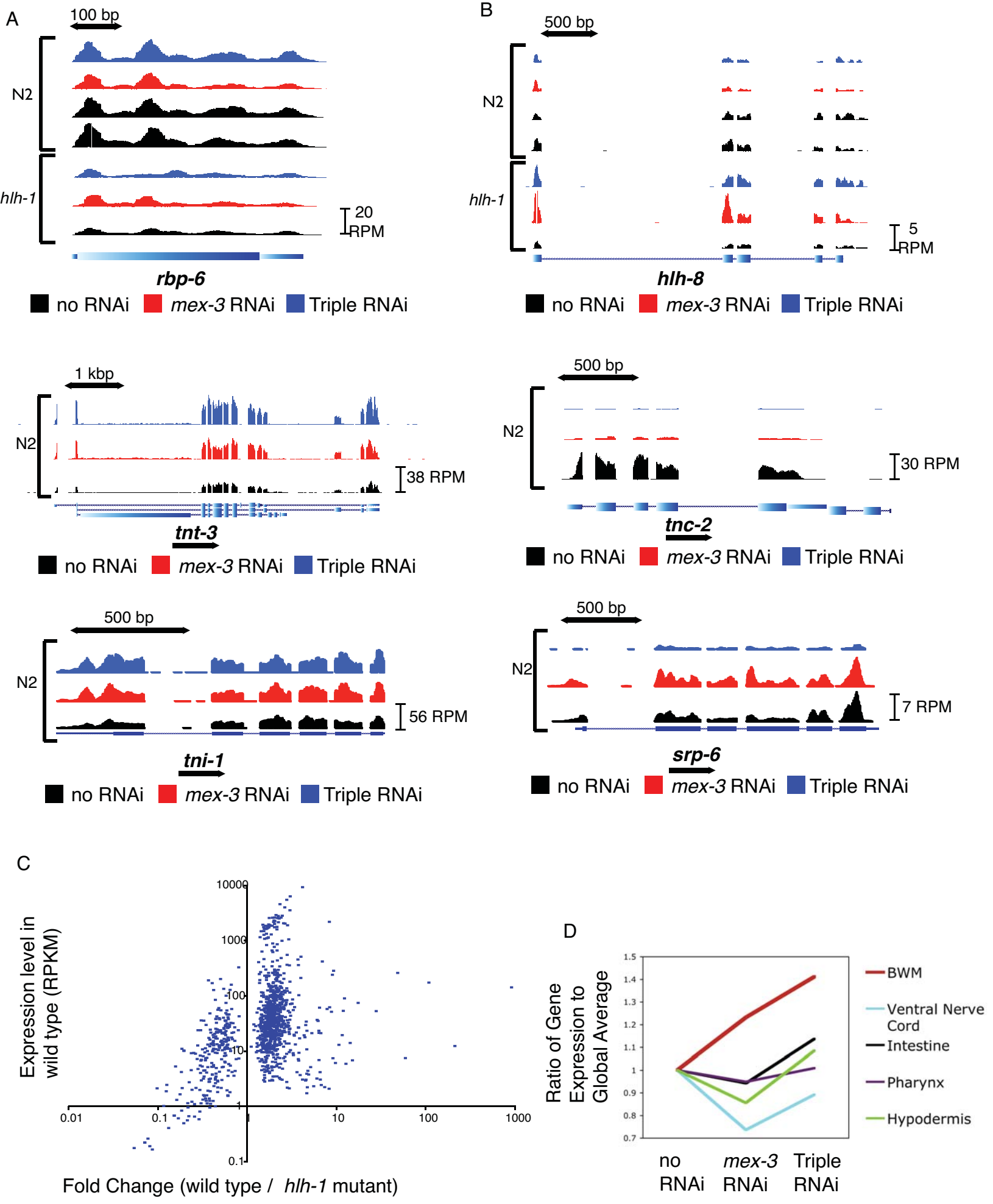
Krause M, Park M, Zhang JM, Yuan J, Harfe B, Xu SQ, Greenwald I, Cole M, Paterson B, Fire A. 1997. A C. elegans E/Daughterless bHLH protein marks neuronal but not striated muscle development. *Development* **124**: 2179-2189.

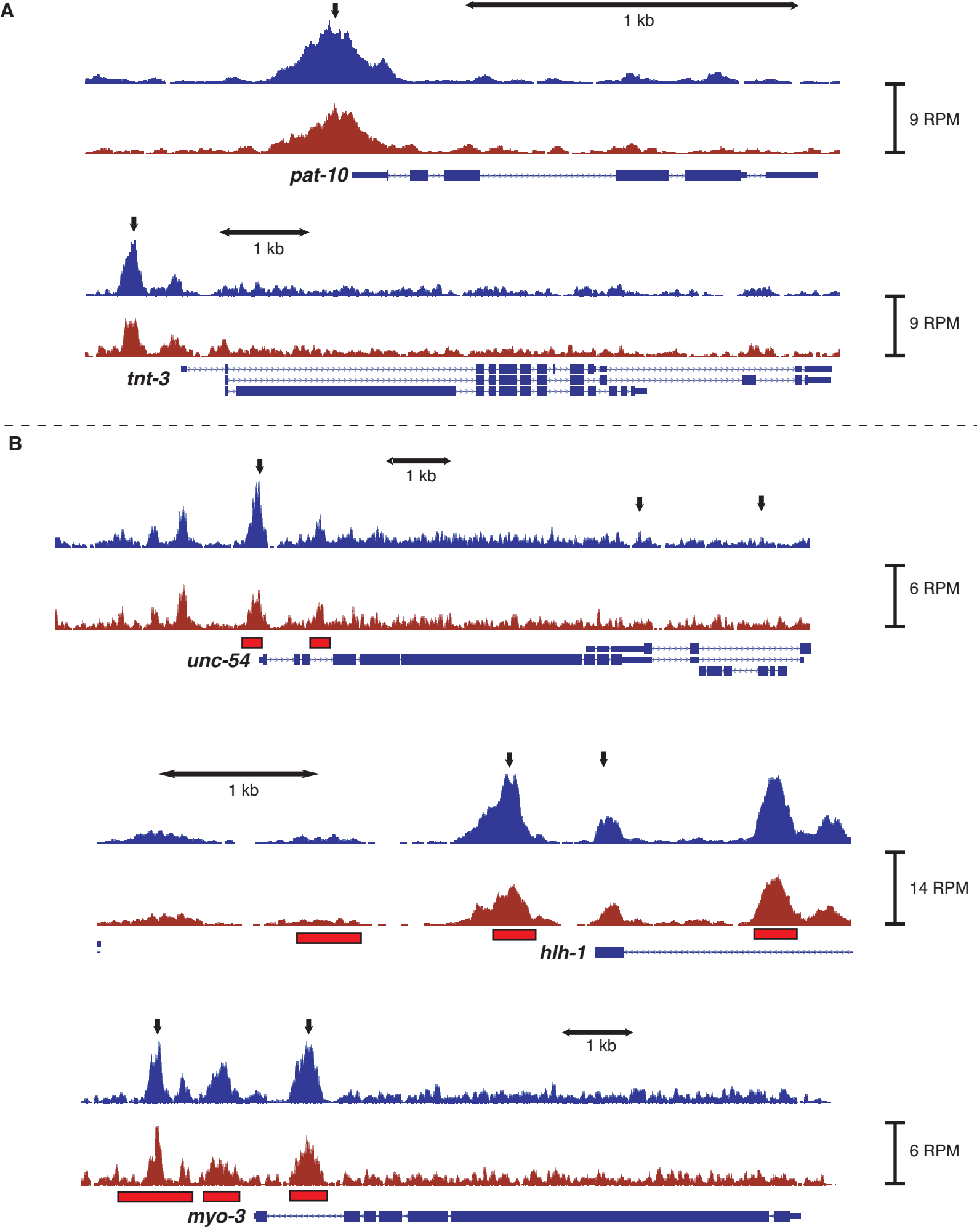
Lei H, Fukushige T, Niu W, Sarov M, Reinke V, Krause M. 2010. A widespread distribution of genomic CeMyoD binding sites revealed and cross validated by ChIP-Chip and ChIP-Seq techniques. *PLoS ONE* **5**: e15898.

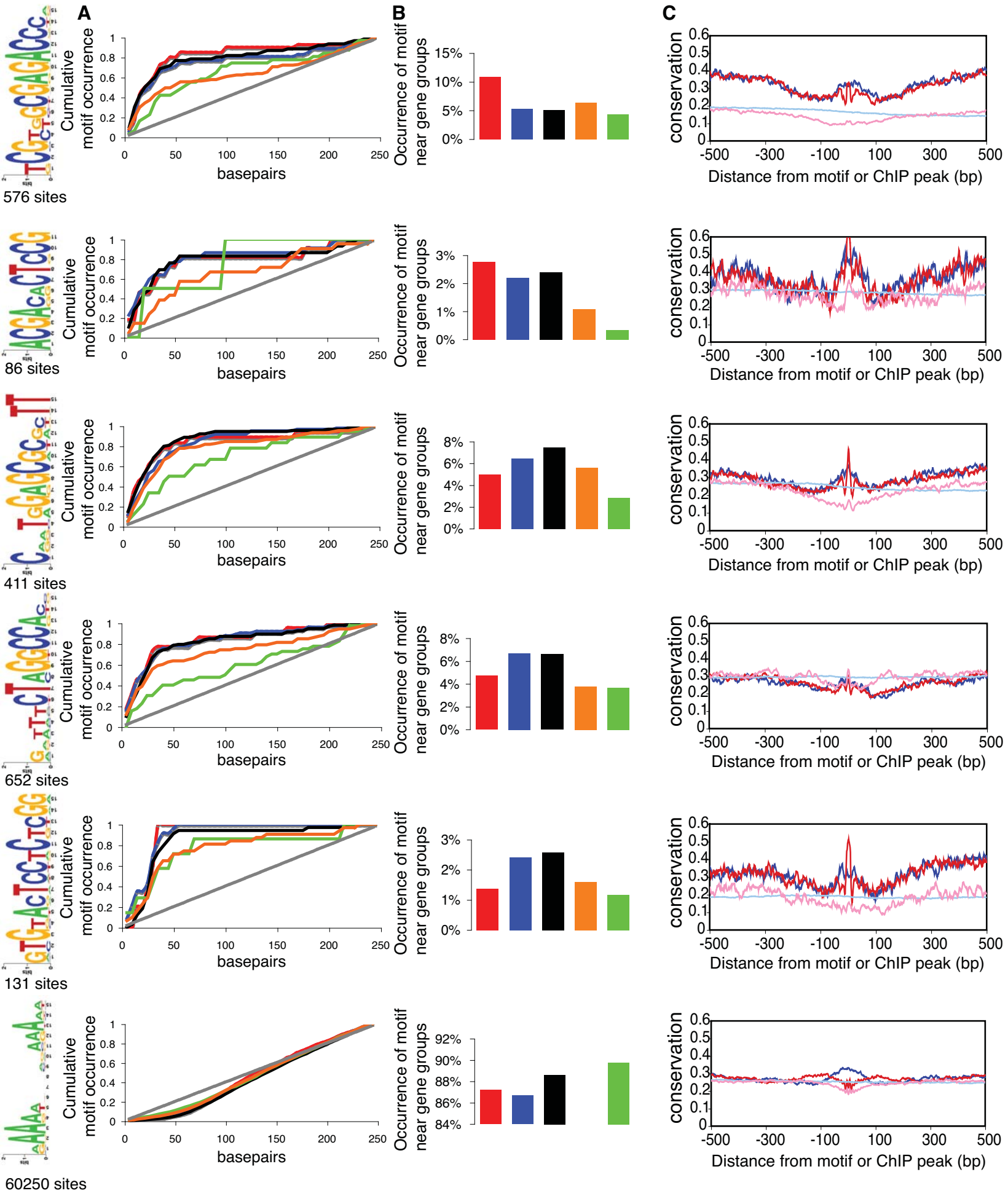
Michaux G, Legouis R, Labouesse M. 2001. Epithelial biology: lessons from Caenorhabditis elegans. *Gene* **277**: 83-100.

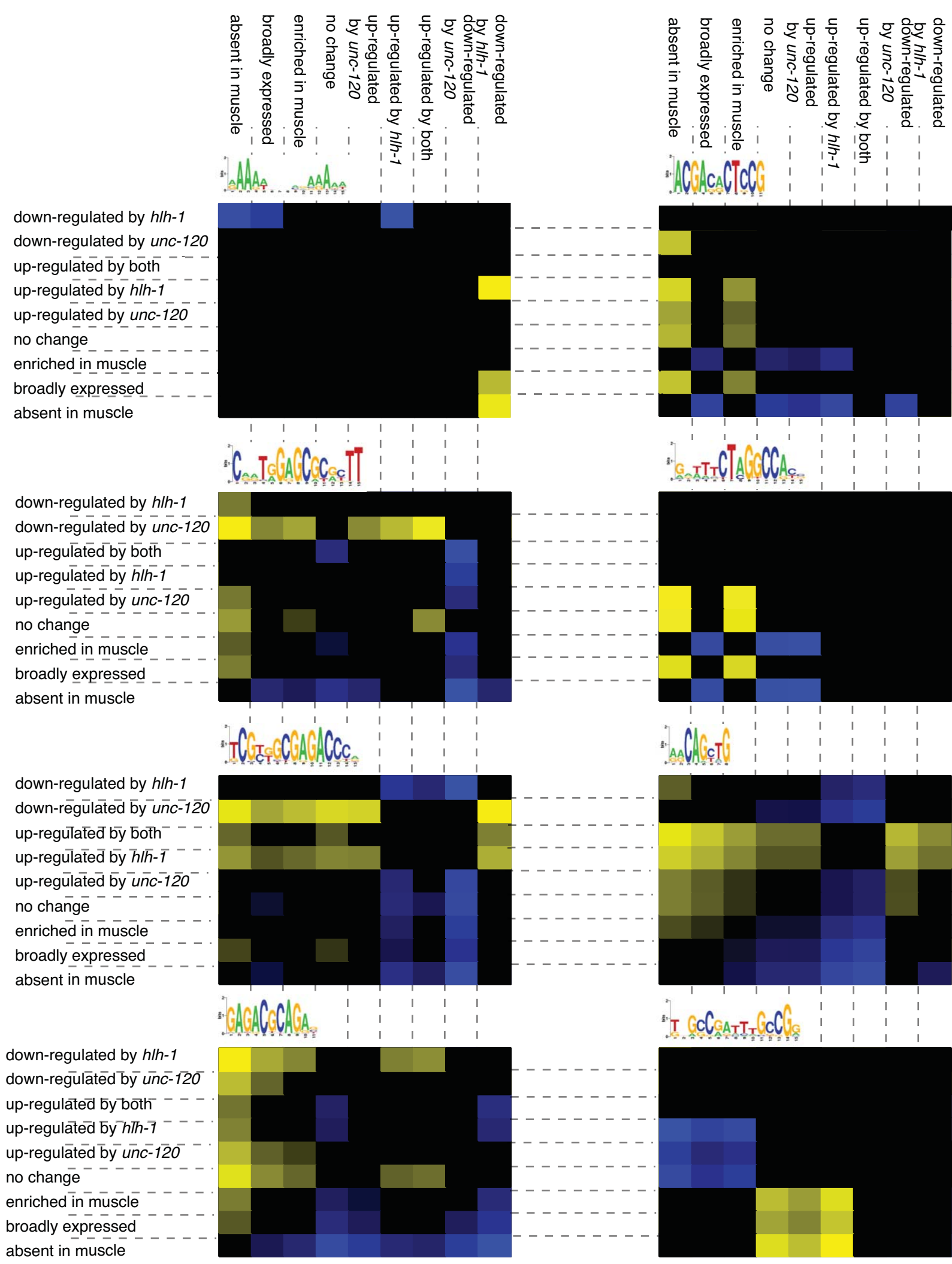
Sulston JE, Schierenberg E, White JG, Thomson JN. 1983. The embryonic cell lineage of the nematode Caenorhabditis elegans. *Dev Biol* **100**: 64-119.

Von Stetina SE, Watson JD, Fox RM, Olszewski KL, Spencer WC, Roy PJ, Miller DM, 3rd. 2007. Cell-specific microarray profiling experiments reveal a comprehensive picture of gene expression in the C. elegans nervous system. *Genome Biol* **8**: R135.


A
100 bp
N2
hlh-1
20 RPM
rbp-6
no RNAi
mex-3 RNAi
Triple RNAi
B
500 bp
N2
hlh-1
5 RPM
hlh-8
no RNAi
mex-3 RNAi
Triple RNAi
1 kbp
N2
38 RPM
tnt-3
no RNAi
mex-3 RNAi
Triple RNAi
500 bp
N2
30 RPM
tnc-2
no RNAi
mex-3 RNAi
Triple RNAi
500 bp
N2
56 RPM
tni-1
no RNAi
mex-3 RNAi
Triple RNAi
500 bp
N2
7 RPM
srp-6
no RNAi
mex-3 RNAi
Triple RNAi
C
Expression level in wild type (RPKM)
10000
1000
100
10
1
0.1
0.01
0.1
1
10
100
1000
Fold Change (wild type / hlh-1 mutant)
D
Ratio of Gene Expression to Global Average
1.5
1.4
1.3
1.2
1.1
1
0.9
0.8
0.7
BWM
Ventral Nerve Cord
Intestine
Pharynx
Hypodermis
no RNAi
mex-3 RNAi
Triple RNAi


A
1 kb
9 RPM
pat-10
1 kb
9 RPM
tnt-3
B
1 kb
6 RPM
unc-54
1 kb
14 RPM
hlh-1
1 kb
6 RPM
myo-3


A
B
C
Cumulative motif occurrence
basepairs
Occurrence of motif near gene groups
conservation
Distance from motif or ChIP peak (bp)
576 sites
86 sites
411 sites
652 sites
131 sites
60250 sites
hlh-1 positively-regulated
unc-120 positively-regulated
Not regulated
BWM
non-BWM
ChIP-seq centered conservation
Motif centered conservation
Background sequence conservation
Background motif conservation


down-regulated by hlh-1
down-regulated by unc-120
up-regulated by both
up-regulated by hlh-1
up-regulated by unc-120
no change
enriched in muscle
broadly expressed
absent in muscle
down-regulated by hlh-1 down-regulated by unc-120
up-regulated by both
up-regulated by hlh-1
up-regulated by unc-120
no change
enriched in muscle
broadly expressed
absent in muscle