Transcription factor redundancy and tissue-specific regulation: Evidence from functional and physical network connectivity

Steven G. Kuntz, Brian A. Williams, Paul W. Sternberg, et al.

*Genome Res.* 2012 22: 1907-1919 originally published online June 22, 2012
Access the most recent version at doi:10.1101/gr.133306.111

Supplemental Material

http://genome.cshlp.org/content/suppl/2012/07/24/gr.133306.111.DC1.html

References

This article cites 64 articles, 34 of which can be accessed free at:
http://genome.cshlp.org/content/22/10/1907.full.html#ref-list-1

Creative Commons License

This article is distributed exclusively by Cold Spring Harbor Laboratory Press for the first six months after the full-issue publication date (see http://genome.cshlp.org/site/misc/terms.xhtml). After six months, it is available under a Creative Commons License (Attribution-NonCommercial 3.0 Unported License), as described at http://creativecommons.org/licenses/by-nc/3.0/.

Email alerting service

Receive free email alerts when new articles cite this article - sign up in the box at the top right corner of the article or [click here](http://genome.cshlp.org/subscriptions)
Transcription factor redundancy and tissue-specific regulation: Evidence from functional and physical network connectivity

Steven G. Kuntz,1 Brian A. Williams,1 Paul W. Sternberg,1,2,3 and Barbara J. Wold1,3

1Division of Biology, 2Howard Hughes Medical Institute, California Institute of Technology, Pasadena, California 91125, USA

Two major transcriptional regulators of Caenorhabditis elegans bodywall muscle (BWM) differentiation, hlh-1 and unc-120, are expressed in muscle where they are known to bind and regulate several well-studied muscle-specific genes. Simultaneously mutating both factors profoundly inhibits formation of contractile BWM. These observations were consistent with a simple network model in which the muscle regulatory factors drive tissue-specific transcription by binding selectively near muscle-specific targets to activate them. We tested this model by measuring the number, identity, and tissue-specificity of functional regulatory targets for each factor. Some joint regulatory targets (218) are BWM-specific and enriched for nearby HLH-I binding. However, contrary to the simple model, the majority of genes regulated by one or both muscle factors are also expressed significantly in non-BWM tissues. We also mapped global factor occupancy by HLH-I, and created a genetic interaction map that identifies hlh-1 collaborating transcription factors. HLH-I binding did not predict proximate regulatory action overall, despite enrichment for binding among BWM-specific positive regulatory targets of hlh-1. We conclude that these tissue-specific factors contribute much more broadly to the transcriptional output of muscle tissue than previously thought, offering a partial explanation for widespread HLH-I occupancy. We also identify a novel regulatory connection between the BWM-specific hlh-1 network and the hlh-8/twist nonstriated muscle network. Finally, our results suggest a molecular basis for synthetic lethality in which hlh-1 and unc-120 mutant phenotypes are mutually buffered by joint additive regulation of essential target genes, with additional buffering suggested via newly identified hlh-1 interacting factors.

[Supplemental material is available for this article.]
to muscle, based on myosin reporter assays, and hlh-1 is the most efficient, with all five up-regulating endogenous hlh-1 and unc-120 (Fukushige and Krause 2005; Fukushige et al. 2006; Broitman-Maduro et al. 2009). unc-120 is the most critical hlh-1 collaborator, based on its synthetic lethality with hlh-1 (Baugh et al. 2005b) and its expression throughout BWM (Baugh et al. 2005a; Fukushige et al. 2006). In contrast, the other factors are confined to developmentally early times of specification or very early differentiation and are restricted to subsets of BWM or are not specific to muscle (Baugh et al. 2003; Amin et al. 2007; Yanai et al. 2008; Broitman-Maduro et al. 2009). For these reasons, we consider hlh-1 and unc-120 to be the core regulators for the differentiated BWM network.

Detailed knowledge of connectivity between hlh-1 or unc-120 and their downstream targets comes from studies of specific target genes, myo-3, unc-54, and pat-3, where binding of the factor was observed at a specific (mutable and essential) cis-regulatory module (CRM) (Francis and Waterston 1985; Fukushige et al. 2006; Lei et al. 2009). These targets and their CRMs serve as internal standards for genomic assays in this work. Whether specific binding and action of HLH-1 and/or UNC-120 proteins regulates the hundreds of additional BWM genes has been untested, and the extent of individual versus shared connectivity is unknown. Recent studies that mapped HLH-1 protein occupancy across the entire genome (Gerstein et al. 2010; Lei et al. 2010) showed widespread binding with proximity to both muscle-specific and nonmuscle genes, but those studies did not directly address the relationship between binding and genome-wide regulatory dependence on the transcription factors.

Nonstriated muscles comprise a minor fraction of C. elegans (four embryonic, 16 post-embryonic muscles, and 10 contractile somatic gonad sheath cells) (Sulston and Horvitz 1977; Sulston et al. 1983). NSM uses hlh-8 as its major transcriptional regulator (Harfe et al. 1998b; Corsi et al. 2000; Liu and Fire 2000) along with unc-120 (Fukushige et al. 2006) and, in a subset of the NSM, mls-1 (Kostas and Fire 2002; Reece-Hoyes et al. 2007). Ectopic hlh-8 produces NSM phenotypes in other cell types (Harfe et al. 1998b; Wang et al. 2006; Zhao et al. 2007). hlh-8 and hlh-1 are transiently coexpressed in the post-embryonic M cell, whose progeny ultimately include 14 BWM cells expressing only HLH-1, 16 NSM cells expressing only hlh-8, and two nonmuscle coelomocytes (Sulston and Horvitz 1977). The molecular and regulatory relationship of BWM and NSM networks is a second focus of this work, based on our findings of crosstalk between them.

Here, we construct a C. elegans BWM genomic resource consisting of RNA-seq transcriptomes of the wild-type, hlh-1 mutant, and unc-120 mutant BWM plus ChIP-seq HLH-1 factor occupancy. Cell type classifications (BWM, general, or non-BWM) are assigned to regulatory target genes by comparing transcriptome measurements from BWM-enriched embryos and normal embryos. We then dissect unique and shared regulatory contributions from each factor by comparing transcriptomes of wild-type embryos with those of hlh-1 mutant and unc-120 mutant embryos. This reveals the regulatory influence of each factor on muscle-specific versus broadly expressed genes. We also provide a genetic resource of previously unknown hlh-1 interacting factors identified via a synthetic RNAi screen. Finally, we measure the number, location, and DNA sequence motif composition of in vivo HLH-1-bound regions to evaluate how biochemical factor binding is related to regulatory impact (Fig. 1B).

Prior genetic studies showed that no single factor in the core BWM network is essential for muscle differentiation (Baugh et al. 2005b; Fukushige et al. 2006; Broitman-Maduro et al. 2009), suggesting there is partial “redundancy” between factors, although no specific molecular explanation was suggested. Among genes affected by hlh-1 mutation in our study, one coherent set of transcription factors includes hlh-8/twist, which is known to positively regulate NSM differentiation. We discuss how this finding, plus other properties of the hlh-1/unc-120 network, contributes to the tolerance of worm BWM myogenesis to hlh-1 and unc-120 mutation.

Figure 1. Experimental flow and muscle differentiation network. (A) The three types of nematode muscle and their associated transcription factors. (B) The experiments performed across three RNAi conditions (mex-3; skn-1; elt-1 RNAi, mex-3 RNAi, and no RNAi) in N2, hlh-1(cc561), and unc-120(st364). (C) The embryonic development lineages with BWM cell number in normal embryos (red) are highlighted to show which lineages are enhanced under the RNAi conditions (mex-3 RNAi and mex-3; skn-1; elt-1 RNAi).

1908 Genome Research www.genome.org
Results

Increasing muscle by respecification reduces nonmuscle background

Our overall study design for genomic measurements is shown in Figure 1B. We increased BWM in embryos by knocking down known specification genes for non-BWM lineages that act prior to hlh-1 and unc-120 (Fig. 1C). Only one-sixth of C. elegans normally becomes bodywall muscle (Sulston et al. 1983). This presents signal-to-noise problems for ChIP and transcriptome experiments by diluting signal and obscuring any signal's cell type source. Increasing the proportion of BWM can ameliorate these problems, but prior methods (Bowerman et al. 1992; Draper et al. 1996; Page et al. 1997; Baugh et al. 2005a) had specific disadvantages for our purposes (Methods). We increased muscle content without directly augmenting the muscle network itself by respecifying nonmuscle fates to a muscle fate. RNAi knockdown of mex-3 can double muscle (Draper et al. 1996), while joint knockdown of mex-3, skn-1, and elt-1 is expected to convert over 80% of cells to BWM. mex-3 acts three cell divisions before HLH-1 expression (Draper et al. 1996; Hunter and Kenyon 1996), skn-1 acts two or three cell divisions beforehand (Bowerman et al. 1992; Blackwell et al. 1994), and elt-1 acts around the time hlh-1 will be activated (Spieth et al. 1991; Page et al. 1997; Michaux et al. 2001) but still permits hlh-1 expression.

We assayed three conditions: no RNAi (empty vector); mex-3 RNAi only; and elt-1, mex-3, and skn-1 triple RNAi. Since knocking down multiple genes via RNAi can significantly reduce the efficiency of each individual knockdown (Gonczy et al. 2000; Gouda et al. 2010), we concatenated RNAi coding sequences to produce a single transcript. As expected, muscle-specific transcripts such as tnt-2 and tnt-3 were enriched by the RNAi strategy across two biological replicates (Fig. 2A). Known nonmuscle genes, such as tnc-2, were reduced with RNAi (Fig. 2B). Genes broadly expressed in both BWM and non-BWM, such as pat-10, were not significantly affected by RNAi (Fig. 2C). Muscle from the triple RNAi sample, unlike mex-3 alone, should be dominated by the C and D lineages, at the expense of the MS lineage. The mex-3 RNAi condition, which doubles the BWM contribution compared with wild type, is included to retain MS-derived muscle for observation (Figs. 1C, 2F).

Figure 2. Mutation and RNAi-based muscle enrichment impact gene expression levels. (A) Muscle troponin T tnt-3 exemplifies genes enriched in muscle-enhanced embryos (red). Muscle-normal expression (black) is nonzero since these animals retain significant muscle. (B) Non-BWM troponin C tnc-2 exemplifies genes depleted in muscle-enriched embryos. BWM-enriched animals (red) have reduced non-BWM tissue, so little expression is seen. (C) Troponin pat-10, which is expressed in both BWM and non-BWM, exhibits a negligible net change. (D) dhp-2 (dihydropyrimidinase) and (E) lbp-3 (lipid-binding) are both muscle-enriched. dhp-2 is affected by hlh-1 loss of function, and lbp-3 is affected by unc-120 loss-of-function, suggesting positive regulation by the BWM transcription factors. (F) Group averages of the ratiometric change for the muscle-enriched genes (those significantly up-regulated in muscle-enriched animals), nonmuscle-enriched (those significantly down-regulated in muscle-enriched animals) genes, and annotated BWM genes (which significantly overlap with NSM and PhM genes and are, therefore, more broadly expressed) are plotted against the RNAi feeding conditions. (G) Overlap of hlh-1 and unc-120 regulated genes, both positively and negatively regulated, including the 441 genes positively regulated by both unc-120 and hlh-1. The number of genes in the other intersects are listed. (H) BWM transcription factors regulate more broadly expressed RNAs than BWM-specific ones. A total of 1139 genes whose RNA levels are significantly regulated by hlh-1 and/or unc-120 are expressed exclusively in BWM; hlh-1 and unc-120 significantly regulated 1068 and 2694 broadly expressed genes. (RPKM) Reads per kilobase of gene structure model per million reads.
RNA-seq reveals hlh-1 and unc-120 regulatory targets, many of which are shared

We quantified transcriptomes from total polyA+ RNA using RNA-seq (Mortazavi et al. 2008). The ~400-min developmental time point (twofold through threefold stage at 25°C in wild-type animals) was used to ensure that BWM cells had already been specified, thus capturing embryos during differentiation. Due to concerns over variations in timing across mutants and RNAi conditions, we verified that both ntl-1 and ceh-51 had shut down in all animals (RPKM < 2), indicating that our samples across replicates and conditions represent middle to late differentiation. To identify regulatory targets of hlh-1 and unc-120, we compared wild type and temperature-sensitive hlh-1(cc561) and unc-120(st364) mutants cultured at the nonpermissive temperature, both with and without RNAi feeding. When the hlh-1(cc561) parent generations are elevated to the nonpermissive temperature prior to egg fertilization at the L4 stage, no detectable HLH-1 remains in the resulting embryos, and there is no maternal or zygotic effect (Chen et al. 1994), as confirmed here by immunoprecipitation (Methods).

We defined hlh-1 and unc-120 regulatory targets as genes whose expression differs significantly (average expression ± one standard deviation) between wild-type embryos and the mutant and the mutant hlh-1(cc561) or unc-120(st364) under RNAi-treated conditions (Mex-3 and Mex-3; elb-1; skn-1 triple RNAi) (Fig. 1B). Among 13,216 genes expressed above background, 1445 are hlh-1 regulatory targets by these relatively conservative criteria, and we expect them to include both direct and indirect mechanisms (Table 1). Of these, 837 decreased significantly (one standard deviation) in the mutant, consistent with a semi-conservative criteria, and we expect them to include both direct and indirect mechanisms (Table 1). Of these, 837 decreased significantly (one standard deviation) in the mutant, consistent with a simple positive mechanism of action for HLH-1 (Lei et al. 2009; see Discussion). Conversely, 608 genes were up-regulated by loss of HLH-1. These are explained most simply by indirect negative mechanisms, although the mammalian hlf-1 ortholog, MyoD, can act as a direct negative factor at a few target genes (Berkes et al. 2004; Dilworth et al. 2004; Penn et al. 2004). Mutation of unc-120 significantly reduced RNA levels from 2718 genes in muscle-enhanced embryos (Table 1), while 956 genes had significantly higher transcript levels.

Among 592 transcription factor genes detectably expressed in our samples, 22 were up-regulated and 12 down-regulated in hlh-1 mutant embryos, while 13 were up-regulated and 103 down-regulated in unc-120 mutants (Supplemental Table S1). tbp-1 and nhr-63, regulated by hlh-1, also interact with hlh-1 in an RNAI synthetic lethal screen (see below), making them especially strong candidate members of a more complete BWM transcription network (Discussion). hlf-1 has a proportionally more negative effect than unc-120 on transcription factors, which may reflect their differing roles in BWM and NSM. A specific and unexpected example of hlf-1 negative regulation was hlf-8. Because hlf-8 is the major positive transcriptional regulator of NSM (Cori et al. 2000, 2002; Liu and Fire 2000), this observation suggests a previously unknown regulatory connection between the BWM and NSM networks. In the unc-120 mutants, the transcript level of unc-120 itself was up-regulated, suggesting that there is a negative autoregulatory feedback loop. unc-120 was not, however, significantly affected by hlf-1 mutation. This result differs from a prior report of unc-120 regulation by hlf-1 (Yanai et al. 2008), although that study was performed at earlier developmental time-points during specification and used a different significance criteria and measurement technology (RT-PCR).

Seven hundred and sixty genes were significantly regulated by both unc-120 and hlf-1 (P < 0.001 for joint regulation), thus offering an explanation for these genes’ continued though diminished expression when either factor is mutated. Four hundred and forty-one of these genes were positively regulated by both factors and 144 were jointly negatively regulated (Table 1; Fig. 2G). An additional 175 genes were divergently regulated between the two factors. The negative and mixed groups include genes best explained by differential regulation between BWM and NSM regulatory networks (Discussion).

**hlf-1 and unc-120 regulate many BWM and non-BWM muscle genes**

We next asked how hlf-1 and unc-120 regulatory target status is parsed among genes expressed preferentially in BWM, genes expressed in both BWM and other tissues, and genes expressed exclusively elsewhere. The last group serves as a measure of noncell autonomous effects and background. Bodywall muscle-enhanced genes were defined by comparing RNA-seq of N2 control (no RNAi) embryos with RNAi-treated muscle-enhanced animals. Overall, 2058 genes had expression levels significantly higher (one standard deviation) in BWM-enhanced worms than in control, including many classical markers of BWM (Supplemental Table S2). From a previously described set of muscle structural genes (Fox et al. 2008), our set included 20 of their 38 genes, with the discordant 18 being

Table 1. Impact of hlf-1 mutation on expression levels in BWM-enhanced worms

<table>
<thead>
<tr>
<th>Regulation</th>
<th>Muscle-specific</th>
<th>Widespread</th>
<th>Low in muscle</th>
<th>Absent in muscle</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td>hlf-1 and unc-120 positively regulated</td>
<td>126</td>
<td>195</td>
<td>120</td>
<td>0</td>
<td>441</td>
</tr>
<tr>
<td>unc-120 only positively regulated</td>
<td>667</td>
<td>1157</td>
<td>358</td>
<td>0</td>
<td>2182</td>
</tr>
<tr>
<td>hlf-1 only positively regulated</td>
<td>66</td>
<td>151</td>
<td>99</td>
<td>0</td>
<td>316</td>
</tr>
<tr>
<td>Expressed but unchanged</td>
<td>919</td>
<td>3839</td>
<td>814</td>
<td>1157</td>
<td>6729</td>
</tr>
<tr>
<td>hlf-1 only negatively regulated</td>
<td>93</td>
<td>183</td>
<td>43</td>
<td>50</td>
<td>369</td>
</tr>
<tr>
<td>unc-120 only negatively regulated</td>
<td>95</td>
<td>470</td>
<td>130</td>
<td>37</td>
<td>732</td>
</tr>
<tr>
<td>hlf-1 and unc-120 negatively regulated</td>
<td>38</td>
<td>71</td>
<td>28</td>
<td>7</td>
<td>144</td>
</tr>
<tr>
<td>hlf-1 positively regulated and unc-120 negatively regulated</td>
<td>20</td>
<td>25</td>
<td>35</td>
<td>0</td>
<td>80</td>
</tr>
<tr>
<td>unc-120 positively regulated and hlf-1 negatively regulated</td>
<td>34</td>
<td>38</td>
<td>23</td>
<td>0</td>
<td>95</td>
</tr>
<tr>
<td>Total</td>
<td>2058</td>
<td>6129</td>
<td>1650</td>
<td>1251</td>
<td>11,088</td>
</tr>
</tbody>
</table>

% positively regulated by hlf-1 | 10% | 6.1% | 15% | 0% | 7.5% |
% positively regulated by unc-120 | 40% | 23% | 30% | 0% | 25% |

*Genes whose expression is significantly higher in muscle-enriched animals.
Genes with similar expression levels with and without RNAi.
Genes expressed less, though still present, in muscle-enriched animals.
Genes not expressed in muscle-enriched animals.
explained by their known expression in pharynx or other tissues. Among 2901 genes expressed preferentially in nonmuscle tissues (Table 1), 1251 appeared entirely restricted to nonmuscle. The true number of genes with this “nonmuscle” pattern is almost certainly higher because our assay is not sensitive to genes expressed at low levels in only a few cells, and there are many such genes in C. elegans.

As expected, some classic BWM differentiation genes were among the most strongly down-regulated by hlh-1 loss (Fig. 2D) or unc-120 loss (Fig. 2E). RNA levels for these genes decreased significantly (-one standard deviation), but none lost all detectable RNA (see Discussion). At the other extreme, a different subgroup of BWM genes, including tri-1 and major actins and myosins, were unaffected by either hlh-1 or unc-120 mutation (Supplemental Fig. S1; Supplemental Table S2). Both unc-120 and hlh-1 contribute widely to BWM and non-BWM exclusive genes (Fig. 2H). Overall, known BWM genes displayed a broad range of quantitative responses to hlh-1 and unc-120 mutation, in both fractional and absolute change. This suggests different regulatory strength contributions from them and, implicitly, from additional transcription factors interacting with subsets of target genes (Supplemental Fig. S1). Of 2058 genes preferentially expressed in muscle, 10% (212) depended significantly on hlh-1 for expression and 40% (827) depended significantly on unc-120 (Table 1). Of all hlh-1 positive regulatory targets, 212 are muscle-preferred, 371 widespread, and 254 depleted but present in BWM. Likewise, unc-120 positive regulatory targets are distributed with 827 being muscle-preferred, 1390 widely expressed, and 501 being depleted but present in BWM.

**hlh-1, but not unc-120, negatively regulates some NSM genes**

An unexpected result was that NSM annotated genes were prominent among the group of 307 genes up-regulated specifically in the BWM-enriched hlh-1 mutant animals and not in wild-type animals (Table 2). Prominent in this group was hlh-8, the central regulator in the NSM differentiation network (Corsi et al. 2000), and mls-1, another transcription factor in the NSM network (Kostas and Fire 2002). These findings were unexpected since triple-RNAi treatment in wild-type embryos abolished, or at least reduced, the entire NSM, including enteric muscles and the M cell lineage. This suggests that a subnetwork of BWM genes behaves differently than the rest of the tissue and that genes in this group are candidate regulatory targets of hlh-8 and/or mls-1. Overall, 9.5% (195 genes) of the BWM preferred expression group were annotated in WormBase as also expressed in normal NSM, and, of these, 26 were up-regulated along with hlh-8 and mls-1. BWM/NSM shared genes showed significant overlap with hlh-1 positively regulated target genes (hypergeometric P < 0.001). Ninety-five percent of NSM genes, including hlh-8 and mls-1, were not detectably elevated in unc-120 mutants, meaning it is likely that repression of NSM circuitry is specific to hlh-1. Though unaffected in BWM-enhanced wild-type embryos, hlh-8 was positively regulated by unc-120 in muscle-normal animals, presumably by acting in the NSM (which is absent in the BWM-enhanced condition).

**Table 2. Genes up-regulated in the hlh-1 mutant muscle-enhanced worms known to be expressed in wild-type NSM cells**

<table>
<thead>
<tr>
<th>Gene</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>B0336.3</td>
<td>RNA recognition</td>
</tr>
<tr>
<td>apg-3</td>
<td>G protein signaling</td>
</tr>
<tr>
<td>arr-1</td>
<td>beta-arrestin</td>
</tr>
<tr>
<td>CO3H5.2</td>
<td>UDP transporter</td>
</tr>
<tr>
<td>ced-1</td>
<td>Lipoprotein receptor</td>
</tr>
<tr>
<td>cts-1</td>
<td>Citrate synthase</td>
</tr>
<tr>
<td>dpy-23</td>
<td>Adaplin</td>
</tr>
<tr>
<td>disc-1</td>
<td>Defecation suppressor</td>
</tr>
<tr>
<td>egl-20</td>
<td>WNT, signaling protein</td>
</tr>
<tr>
<td>exp-1</td>
<td>GABA receptor</td>
</tr>
<tr>
<td>F47B7.2</td>
<td>Sulphydryl oxidase</td>
</tr>
<tr>
<td>H2B01.6</td>
<td>ATP synthase</td>
</tr>
<tr>
<td>hlb-8</td>
<td>TWIST, transcription factor</td>
</tr>
<tr>
<td>mls-1</td>
<td>TBX1, transcription factor</td>
</tr>
<tr>
<td>mrp-2</td>
<td>Multidrug resistance protein</td>
</tr>
<tr>
<td>mua-6</td>
<td>Intermediate filament</td>
</tr>
<tr>
<td>mph-4</td>
<td>Muscle junctions</td>
</tr>
<tr>
<td>nlp-13</td>
<td>Neuropeptide</td>
</tr>
<tr>
<td>nmy-1</td>
<td>Nonmuscle myosin</td>
</tr>
<tr>
<td>ppk-3</td>
<td>PIP kinase</td>
</tr>
<tr>
<td>rom-1</td>
<td>Rhomboid related</td>
</tr>
<tr>
<td>shc-1</td>
<td>Signaling (arc, jnk, insulin)</td>
</tr>
<tr>
<td>snb-1</td>
<td>Synthetic vesicle</td>
</tr>
<tr>
<td>trs-1</td>
<td>tRNA synthetase</td>
</tr>
<tr>
<td>uvr-3</td>
<td>Pantothenate kinase</td>
</tr>
<tr>
<td>ZK1123</td>
<td>Unknown</td>
</tr>
</tbody>
</table>

**Synthetic PAT screen for coregulators of hlh-1 and mediators of hlh-8/hlh-1 crosstalk**

As shown above, unc-120 partly explains the robustness of worm myogenesis to hlh-1 mutation, but other factors might perform a similar function for additional hlh-1 targets. To find other regulators that collaborate with hlh-1, we performed a feeding RNAi synthetic paralysis-at-twofold (Pat; WBP0000053) phenotype analysis in the hlh-1(cc561) mutant background, using a library of 512 genes that encode known and suspected transcription factors. In nematodes, elongation of the embryo depends on muscle contractions (Williams and Waterston 1994). The Pat phenotype, therefore, serves as a readily scored surrogate for major BWM failure. As expected, unc-120 scored strongly in this assay. Other strong interactors included ceh-20, grh-1, tbp-1, lin-26, pos-1, oma-2, nhr-4, nhr-46, nhr-63, nhr-116, tnm-1, hnd-1, and ceh-51 (Supplemental Table S5). The Pat phenotype suggests that each of these contributes to expression of one or more genes needed for differentiation of muscle in the absence of hlh-1.

A majority of these hlh-1 genetic interacting factors are themselves regulated by hlh-1 and/or unc-120. TATA-binding protein (tbp-1) is part of the transcription initiation complex and is positively regulated by both hlh-1 and unc-120, suggesting elevated demand for it by some muscle differentiation genes. nhr-63 is negatively regulated by hlh-1, while nhr-116 is negatively regulated by unc-120. lin-26, pos-1, oma-2, nhr-4, and nhr-46 are positively regulated by unc-120, suggesting feed-forward loops that are familiar structures in developmental circuits. nhr-63, grh-1, and ceh-20 are normally expressed in NSM, so they are candidates for genes that could interact with both hlh-1 and the hlh-8/mls-1 circuitry.

**Genes positively regulated by hlh-1 and unc-120 are enriched for HLH-4 occupancy, which is widespread**

To map sites of HLH-1 occupancy in vivo, we performed chromatin immunoprecipitation from RNAi fed wild-type embryos (both mce-3 and triple RNAi) with an anti-HLH-1 antibody, followed by DNA sequencing (ChIP-seq) (Johnson et al. 2007; Zhong et al. 2010). The most prominent signals were consistent in all conditions (Supplemental Fig. S2), but BWM-enrichment was important for detecting the majority of HLH-1 ChIP signals. We evaluated ChIP-seq signal intensities and locations relative to background
Eighty-nine percent of the stringent set of HLH-1-occupied regions were within 5 kb 5'ward of an annotated gene start (including regions that also fell within an upstream gene), with 36% of those concentrated in the proximal 500 bp. Sixteen percent of regions were in introns, 6.6% in exons, and only 1.2% in 3'UTRs (Supplemental Table S3). The 5-kb 5'-ward, 500-bp proximal, and exon sequences were enriched genome-wide for HLH-1 ChIP peaks (P < 0.01), while other regions were depleted or not enriched (peaks per kb).

Because hlh-1 is a highly cell-type-specific activating transcription factor, an initial expectation was that most BWM-specific genes would have one or more adjacent HLH-1 ChIP regions. For specific CRMs and promoters previously shown to drive BWM expression (Okkema et al. 1993; Krause et al. 1994), this was true, with ChIP signals at expected locations (Supplemental Fig. S2). Genome-wide, 59.7% of the 941 annotated BWM genes had HLH-1 occupancy (broad set) within the gene body or 5 kb upstream (Supplemental Table S4), while 54% of our BWM-enriched expression gene set did. Sixty-seven percent of genes near a stringent HLH-1 peak (i.e., within 5 kb of the start site) are expressed at a significant level (RPKM > 3) in BWM (P < 0.001). However, the vast majority (80% in the stringent set and 87% in the broad set) were not muscle-preferred in their expression pattern (Fig. 3C; Discussion). Rather, the majority are expressed widely in muscle and nonmuscle tissue.

Genes whose expression depended positively on hlh-1 (Table 1; dhp-2 in Fig. 3D) were significantly, but not strikingly, enriched for HLH-1 occupancy within 5 kb upstream or in the gene body, compared with other genes in the genome (57% vs. 49%, P < 0.01) (Supplemental Table S4), while negatively regulated targets were not enriched (48% vs. 49%; hlh-8 in Fig. 3D). The overlap was on par with that of Lei et al. (2010), with our high-confidence peaks and their ChIP-chip analysis both yielding 5% of the occupied genes depending on hlh-1, while their ChIP-seq analysis yielded 9% of occupied genes depending on hlh-1. Genes depending on unc-120 (Table 1) were also more likely to have HLH-1 occupancy than the rest of the genome (53% vs. 49%, P < 0.001) (lin-25 in Fig. 3D). Genes jointly up-regulated by both hlh-1 and unc-120 function were similarly enriched for HLH-1 occupancy (54% vs. 49%, P < 0.002), though it is not required (skr-2 in Fig. 3D).

hlh-8, mls-1, and grl-26 were among 608 genes under negative regulation by hlh-1 but not by direct binding, according to the ChIP data (Fig. 3D). hlh-8, in particular, had no HLH-1 binding in its gene body nor within 20 kb upstream of the TSS or 10 kb downstream.

Figure 3. HLH-1 ChIP-seq binding is associated with, but not predictive of, regulation. HLH-1 binds to the genes (A) dhp-2 and (B) lin-25 (arrows). (C) Venn diagram shows four criteria for rating interactions of BWM preferred expression (red circle, 2175 genes), hlh-1 regulation of expression (blue circle, 757 genes), HLH-1 ChIP-seq binding (green circle, 9519 genes), and the presence of a local HLH-1 binding motif (yellow circle, 3469 genes). The intersect of 78 genes is highlighted as the "Archetypal Muscle List." (D) Gene expression (RPKM levels from RNA-seq) levels illustrate different regulatory dependency patterns for hlh-1 and unc-120, with (upper) or without (lower) detectable nearby HLH-1 occupancy. dhp-2 (gold from panel C) represents archetypal muscle genes, positively regulated by hlh-1 with significant HLH-1 occupancy. lin-25 (green, from C) is positively regulated by unc-120 but not hlh-1, even though it has HLH-1 occupancy. skr-2 (blue, from C) is positively regulated by both hlh-1 and unc-120. hlh-8 (black) represents a class up-regulated only in hlh-1 mutant BWM, suggesting indirect negative regulation.

Kuntz et al. (Pepke et al. 2009) to produce a high confidence set of 1047 peaks appearing in both RNAi conditions and a more inclusive union set of 9415 peaks appearing in either RNAi condition (Mex-3 RNAi yielded 7021 peaks and triple RNAi yielded 3441 peaks; examples in Fig. 3A,B). The peak yield was similar to that of Lei et al. (2010), who recovered 20,143 peaks in their ChIP-seq experiment that they narrowed to 4016 high-confidence peaks (Lei et al. 2010). Their use of a different antibody but a different enrichment and detection technique (Lei et al. 2010), leading to a similar sample size as our high-confidence set of peaks and a statistically significant overlap (P < 0.001). The muscle enrichment and regulatory dependence of genes near peaks from both broad and stringent sets were comparable, but the stringent group was more strongly enriched. Thus, >50% and 20% of our hlh-1 positively regulated gene list was captured in the broad and stringent HLH-1-bound sets, respectively.
ChIP-regions contain the canonical HLH-1 binding motif and novel associated motifs

HLH-1 occupied regions were used to derive overrepresented sequence motifs. We expected to identify motifs responsible for direct HLH-1 binding, together with possibly collaborating motifs, since the latter are often present in functional cis-regulatory modules (Davidson 2007). Two substantially different motif discovery algorithms found similar motifs (Methods). The primary motif was AACACTGT (Fig. 4A, first motif), which is an E-box family motif (CANNTG is the known general motif for the bHLH family). The core hexamer matches previous HLH-1 motif determinations from yeast one-hybrid assays (Grove et al. 2009), in vivo ChIP-chip and ChIP-seq (Lei et al. 2010), and mammalian MyoD (Cao et al. 2010), while the adjacent AA produces a more specific site which is analogous, but not identical, to the most highly preferred myogenin binding octa-E-box in mouse (CAGCTGRR) (Cao et al. 2010; A. Kirilusha and B. Wold, in prep.). Six other motifs (Fig. 4A, Supplemental Fig. S3) plus GA- and CT-simple repeat-rich regions (Guhathakurta et al. 2002; GuhaThakurta et al. 2004) were found. Expanding the search radius from 50 bp to 100 bp found a second E-box: CAACCTG (web logo not shown), reported previously as a secondary site for HLH-1 binding (Grove et al. 2009; Lei et al. 2009).

Analyzed across all HLH-1 ChIP regions with a 250-bp radius, the two E-boxes and the GAGaCCtGAGa motif (Fig. 4A, second motif), for which there is no known factor, were strongly centered near ChIP-seq summits, with the most statistically significant central concentration being in muscle-specific and unc-120 positively regulated genes (P < 0.05; Fig. 4E; Supplemental Fig. S4). The centered position argues that a motif is partly or solely responsible for the observed ChIP signal. Other motifs were more evenly distributed in the tested regions (Fig. 4B; Supplemental Fig. S3), consistent with accessory or independent roles. As expected, the HLH-1 octa-box motif was most highly overrepresented among genes expressed preferentially in muscle (Fig. 4E).

HLH-1 ChIP-seq peaks near specific functional subsets of genes were analyzed for motif discovery, position, and frequency. Gene groups tested were those (1) strongly positively dependent on hlh-1 for expression, (2) strongly negatively regulated by hlh-1, (3) absent in bodywall muscle, (4) dependent on unc-120 for expression, (5) less stringently dependent on hlh-1 for expression, and (6) dependent on both hlh-1 and unc-120. Motifs identified above were rediscovered within some subsets, in addition to two novel candidates: AAAANNNAAA and GCCGaTTTGCG (Fig. 4A, third motif; Supplemental Fig. S3, sixth motif). The GCCGaTTTGCG motif was specifically associated with genes that do not positively depend on hlh-1 and with genes that do depend on unc-120. In fact, this motif was selectively depleted from the positively regulated HLH-1 gene set (P < 0.01).

HLH-bound regions are preferentially conserved

If HLH-1 occupied E-box motif instances located near hlh-1 regulated genes are functionally significant, we expect them to be preferentially conserved in evolution. Moreover, we expect functional HLH-1 binding sites to be embedded in larger domains of conservation that typify cis-regulatory modules. This was the case around our set of HLH-1 occupied sites, with preferential conservation among sequenced nematodes of ±200 bp (Fig. 4D). This conservation was not restricted to sites near HLH-1 regulated genes; rather, the larger set of HLH-1 ChIP regions located near genes that were not regulated by hlh-1 or were not BWM-specific displayed similar preferential conservation (Supplemental Fig. S3B). This suggests that HLH-1 ChIP signals overall identify functionally important sequences, but that these need not be adjacent to muscle-specific or HLH-1-dependent genes. Among the new candidate motifs, three others show preferential conservation, while GCCGaTTTGCG did not (Supplemental Fig. S3).

Discussion

Through analysis of the BWM differentiation network, we uncovered a significant overlap in transcription factor function, helping to explain the redundancy of the core factors, and a surprising lack of muscle factor target specificity to BWM. This work expanded loss-of-function analysis for BWM regulation to cover the entire transcriptome, finding that 21% (4359) of C. elegans genes are significantly affected by mutation of either hlh-1 or unc-120. It was found that 3.7% (760) were affected by both regulators (Table 1), and these were highly enriched for BWM annotation. Their pattern of regulatory dependence helps to explain how and why hlh-1 and unc-120 act as a synthetic lethal pair in the embryo. However, an equally strong result was that 71% of jointly regulated genes and 79% of HLH-1 or 69% of UNC-120 single targets are not tissue-specific—rather, they are substantially regulated by the muscle-specific factors in BWM and, presumably, by other unknown regulators in nonmuscle tissues. By integrating our mapping of regulatory connectivity with in vivo physical HLH-1 occupancy, we were able to define a set of “archetypal” direct transcriptional targets for hlh-1 (Fig. 3C and below); identify biologically pertinent indirect regulatory relationships, including the major NSM-specific regulator, hlh-8 (Fig. 5A,B,C; below); and define a set of HLH-1 occupancy sites located near broadly expressed genes. Since a significant number of broadly expressed HLH-1 occupied loci were functionally affected by hlh-1 mutation (220 genes), we conclude that hlh-1 either originated as a highly muscle-specific factor that has been drafted over time to help regulate widely expressed target genes in the specific context of muscle, or that hlh-1 was originally a more general factor whose role was narrowed to muscle tissue early in animal evolution. New hlh-1 collaborating factors expand the BWM differentiation regulatory network and network orthology based on an RNAi screen for muscle failure.

unc-120/HLH-1 compensation is based on overlapping roles in regulatory target control

In nematodes, myogenesis is robust to mutation of either unc-120 or hlh-1 (Baugh and Hunter 2006; Fukushige et al. 2006); by contrast, the hlh-1 bHLH ortholog, myogenin, is absolutely required for mammalian differentiation, and, in Drosophila, the unc-120 MADs family ortholog, Mei2 (also known as D-MEF2), is absolutely required (Black and Olson 1998). Nevertheless, the common theme is that bodywall or skeletal muscle differentiation in all three phyla uses both bHLH and MADs regulators. Our results help to explain the C. elegans network’s unique behavior in three ways. First, 760 genes are jointly regulated (Table 1), and the regulatory contributions from the two factors are roughly additive rather than highly synergistic. Significant residual expression (>50%) was observed in each mutant strain for the vast majority of shared positively regulated targets (89%), and only one gene (tag-10) lost more than 90% of its expression (Supplemental Fig. S3C). Though all these numbers are sensitive to thresholds, the qualitative results remained unchanged even when the stringency was significantly increased. The troponin gene family, several of whose members have been studied individually, nicely illustrates the varied regulatory specificity.
Figure 4. HLH-1-associated motifs correlate with directional expression control. (A) The web logo position-specific frequency matrix (PSFM) diagrams for three representative motifs and the accompanying number of sites identified near HLH-1 occupancy (250-bp radius). (B) The relative locations of three motifs compared to their experimentally identified binding sites (analyzed per Ozdemir et al. 2011). The AACACTGTG motif is centered on the called ChIP-seq peak (50% within ±25 bp of the peak for hlh-1 positively regulated or muscle-enriched genes). The GAGACGCAGA motif (second panel) is less central (within 75 bp). The GCCGTT GCCG motif (third panel) shows no significant centrality. The gray line represents a uniform distribution. (C) The occurrence of each motif within ±250 bp of the HLH-1 occupancy peak near genes (ChIP regions within 5 kb of a gene TSS) belonging to expression groups is shown. The E-box shows the greatest enrichment for genes characterized as hlh-1 positively regulated (first panel). The GAGACGCAGA motif is more closely associated with unc-120 positively regulated genes (second panel), whereas the GCCGTT GCCG motif is enriched near genes absent in BWM (third panel). (D) The conservation across sequenced nematodes (elegans, briggsae, remanei, and brenneri) of ChIP-seq identified regions with the three motifs is shown. Conservation around the in vivo binding (blue) and around the motif (red) is shown compared to background (light blue and pink) (Ozdemir et al. 2011), with higher values representing a higher level of conservation. The E-box and GAGACGCAGA motifs, along with their surrounding sequences, are strongly conserved, while the GCCGTT GCCG motif is not at all conserved. (E) Heat maps show the level of motif enrichment (yellow) or depletion (blue) for the CAGCTGTG, GAGACGCAGA, and GCCGTT GCCG motifs near broadly expressed genes that are similarly regulated (y-axis). The E-box is enriched near genes positively regulated by hlh-1 and unc-120. The GAGACGCAGA motif is enriched near genes negatively regulated by hlh-1 and positively regulated by unc-120. The GCCGTT GCCG motif is depleted near genes positively regulated by either factor. (F) There are four classes of E-boxes observed: Class I contains muscle E-boxes that are bound by HLH-1, and it is predicted that mutation of these sites will lead to changes in expression, as the nearby genes are both specific to BWM and regulated (positively or negatively, in contrast to the Archetypal Genes, which are exclusively positively regulated) by hlh-1; Class II contains E-boxes that are similarly functional but are near genes not exclusively expressed in BWM; Class III contains E-boxes that are not required for expression but likely make contributions to nearby genes that are expressed exclusively in BWM; and Class IV contains seemingly nonfunctional E-boxes that are not required for expression or associated with BWM expression.
hlh-1 and unc-120 network in worm myogenesis

across multiple muscle types as well as in nonmuscle tissue (Supplemental Table S6). unc-120’s large pool of positive targets may be partly explained by a broader activating role in both BWM and NSM, rather than just one muscle type, meaning that genes expressed in both may be regulated primarily by unc-120. hlh-1’s smaller pool of regulatory targets and proportionally larger role in repression may reflect its narrow and specific function in BWM alone. This appears to include the capacity to silence the NSM network and potentially other nonmuscle gene networks. We cannot say yet how joint regulation is partly explained by a broader activating role in both BWM and NSM, rather than just one muscle type, meaning that genes expressed normally in NSM and could function as part of this intersecting circuit. Target genes that are divergently regulated by unc-120 and hlh-1 might be explained as additional genes normally necessary in NSM (a domain of unc-120 regulation) but not in BWM.

Crosstalk between NSM and BWM regulators without wholesale tissue conversion

The finding that hlh-8 and mls-1 are strongly up-regulated by hlh-1 mutation raises several questions. At the tissue level, does up-regulation of hlh-8 and mls-1 produce a wholesale transformation of BWM into NSM? It appears not, since many BWM-specific genes were readily detected in hlh-1 mutant RNA. Of 104 genes annotated as expressed in wild-type BWM but not in NSM, all continued to be expressed significantly in hlh-1 mutant embryos, with or without RNAi feeding. At the network level, are muscle genes normally expressed in both NSM and BWM similarly expressed in hlh-1 mutants, as would be expected if they are primarily positively regulated by unc-120, hlh-8, and/or mls-1? Indeed, the vast majority (552 of 596) of genes annotated for both BWM and NSM expression in wild-type worms were similarly expressed in hlh-1 mutants. Only four BWM/NSM shared genes were reduced by more than two standard deviations from their wild-type level in the mutant. However, hlh-8 up-regulation is not sufficient to explain muscle differentiation in the absence of hlh-1, because it is not synthetic muscle-lethal with hlh-1.

At the level of circuit structure and molecular mechanism, why and how is hlh-8 switched on in the absence of hlh-1, and what does this imply about their relationship in normal muscle development? Negative regulation of hlh-8 by hlh-1 is likely to be by an indirect mechanism, partly because hlh-1 is known as a positive regulator of its direct targets. In addition, we detected no HLH-1 occupancy near hlh-8 or mls-1 via ChIP-seq, even at the most relaxed peak calling stringency. Our RNA data show that known positive regulators of hlh-8, such as unc-62, ceh-20, and mab-5, are all present in both wild-type and hlh-1 mutant animals,
making activation of \textit{hlh-8} highly plausible when negative regulation mediated indirectly by \textit{hlh-1} is relieved (Harfe et al. 1998a).

Drawing on our data and additional studies, we propose a specific model for regulation of \textit{hlh-8} by \textit{hlh-1} (Fig. 5). From the synthetic PAT phenotype analysis, \textit{celb-20}, \textit{lin-39}, and \textit{grh-1} were identified as strong genetic \textit{hlh-1} interactors. By independent criteria, each of these is also a candidate to help activate \textit{hlh-8}, \textit{grh-1} positively regulates \textit{mab-5} (Venkatesan et al. 2003), and \textit{mab-5\textbackslash{}unc-62/celb-20} positively regulates \textit{hlh-8} in \textit{NSM} (Liu and Fire 2000). Furthermore, we found that \textit{egl-15a} and \textit{grh-1} are up-regulated in our \textit{hlh-1} mutants (Fig. 5A), and there is additional genetic precedent in \textit{Drosophila} for interaction between \textit{mab-5} and \textit{egl-15} (Zhong and Sternberg 2006). \textit{EGL-15\textbackslash{}FGFR} is necessary for proper sex myoblast (NSM) migration (Stern and Horvitz 1991). The primary splicing variant in \textit{NSM}, \textit{EGL-15a}, is down-regulated by \textit{SUP-12}, which destroys \textit{EGL-15a} but not \textit{EGL-15b}, the primary splicing variant in \textit{BWM} (Kuroyanagi et al. 2007). We found that the \textit{EGL-15a} RNA splice isoform is up-regulated in \textit{hlh-1} mutants. \textit{SUP-12} expression depends on \textit{HLH-1} activity, and \textit{SUP-12} has a high confidence \textit{HLH-1} occupancy domain (Fig. 5B). We therefore suggest that in normal \textit{BWM}, \textit{HLH-1} drives \textit{SUP-12} to down-regulate \textit{EGL-15a} and \textit{grh-1}, while in \textit{hlh-1} mutant muscle, \textit{SUP-12} is not expressed, and \textit{EGL-15a} and \textit{grh-1} increase, thus activating \textit{mab-5\textbackslash{}unc-62/celb-20} and leading to up-regulation of \textit{hlh-8} and some of its \textit{NSM} target genes (Fig. 5C). This should especially favor target genes normally expressed in both \textit{NSM} and \textit{BWM}, since collaborating factors from \textit{BWM} are present.

There is also evidence for reciprocal repression of \textit{BWM} by \textit{NSM}, since \textit{hlh-8} mutants have an unstable and sometimes higher number of \textit{BWM} cells and their sex-specific muscles disappear (Corsi et al. 2000, 2002). Up-regulation of \textit{hlh-8} in \textit{BWM} in the absence of \textit{hlh-1} is reminiscent of the connection reported in the post-embryonic M lineage (Harfe et al. 1998a), and regulation of normal M lineage development, which generates both \textit{NSM} and \textit{BWM}, might account for the crosstalk we see in \textit{BWM} upon \textit{hlh-1} mutation.

**Defining “archetypal” \textit{hlh-1} target genes and their candidate CRMs**

We distilled a set of 78 genes and associated candidate \textit{cis}-acting regulatory modules that meet four criteria for being “archetypal” regulatory and molecular \textit{hlh-1} targets: (1) They are expressed preferentially in \textit{BWM}; (2) they display significantly reduced RNA levels in \textit{hlh-1} mutants; (3) they have \textit{HLH-1} occupancy at one or more sites in our ChIP-seq data; and (4) the \textit{HLH-1} occupancy region contains one or more instances of the extended myogenic “octa-Ebox” (AAACGTCG) (Supplemental Table S7; Fig. 3C). An additional 154 genes satisfy criteria 2, 3, and 4 but are also strongly expressed in tissues other than \textit{BWM}. These genes apparently depend on \textit{HLH-1} in the context of muscle and on other factors elsewhere.

Membership in our list of candidate \textit{hlh-1} \textit{BWM} CRMs did not use DNA sequence conservation as an initial criterion, since recently evolved active instances may exist and will be pertinent to the network. This allowed us to ask if the candidate canonical muscle CRM group is preferentially conserved, and it was. Preferential conservation encompassed a region of \(\pm 150\) bp relative to the \textit{HLH-1} ChIP peak (Fig. 4D), showing that this group of candidate CRMs has been under pressure to function. The muscle octa-Ebox is even more highly conserved than the surrounding domain, suggesting that it drives binding of functional consequence.

The archetypal muscle \textit{HLH-1} targets were defined without using \textit{unc-120} data, yet they are highly enriched in \textit{unc-120} regulatory targets. Thus, 56% of the \textit{hlh-1} archetypal loci are positively regulated by \textit{unc-120}, while only 18% of all genes are regulated by \textit{unc-120}. Similarly, 52% of genes satisfying criteria 2, 3, and 4 were positive regulatory targets of \textit{unc-120}. Finally, of the genes in the archetypal group that are also \textit{unc-120}-regulated, 20% contain an 85% match instance to GAGACGCGAGA within their \textit{HLH-1}-occupied region.

The archetypal \textit{HLH-1} target genes and candidate CRMs have been defined by intentionally stringent multiple-measurement intersection to help learn the defining and shared characteristics of \textit{BWM} regulation. It is, therefore, an underestimate of the \textit{BWM} group and highlights the important role \textit{hlh-1} plays in regulating genes not specific to the \textit{BWM} group.

**HLH-1 occupancy versus \textit{hlh-1} regulatory impact**

Expression of \textit{hlh-1} is specific to the bodywall muscle system, the phenotype of \textit{hlh-1} null mutants is myogenic, and \textit{hlh-1} orthologs across metazoan phyla regulate muscle development and differentiation. This relative simplicity made it possible to address some questions about factor occupancy that have been difficult in systems with larger genomes and more complex organization. First, factor occupancy alone, as measured by \textit{HLH-1} ChIP-seq, is a permissive condition for regulating gene expression, but it is not powerfully predictive of regulatory activity. In isolation, \textit{HLH-1} occupancy had low specificity for \textit{hlh-1}-dependent RNA expression at nearby genes or TSS. \textit{HLH-1} occupancy in an independent study (Lei et al. 2010) had almost identical specificity (9% in Lei et al. [2010] versus 10% in this analysis). The Lei et al. (2010) study also concluded that many sites were upstream of nonmuscle genes. Their conclusion that binding was not predictive of enhancer activity mirrors our conclusion that it is not predictive of regulatory activity (Lei et al. 2010). Similarly, the majority of mouse MyoD occupancy sites are located closest to nonmuscle-specific genes (Cao et al. 2010). Substantial technical issues surrounding assay sensitivity, combined with consequences from assigning binding regions to genes by an overly simple proximity algorithm, are probably responsible for some lack of predictive power. Nevertheless, the data are most consistent with a majority of detected \textit{HLH-1} occupancy binding events having no regulatory effect or an effect too small to measure. Where \textit{hlh-1} regulation is detected, the majority of it is associated with genes that are not specific for \textit{BWM} alone.

From \textit{HLH-1} ChIP-seq regions, we refined the \textit{HLH-1} binding motif preferentially affiliated with \textit{HLH-1}-regulated loci (AAACGTCG) and showed it is the dominant, centrally located driver for \textit{HLH-1} genome occupancy. The distribution of additional novel motifs from \textit{HLH-1}-occupied regions, among the functionally distinct target gene groups, provided insights into \textit{BWM/NSM cis}-regulatory logic. The GAGACGCGAGA has a high rate of co-occurrence with the octa-Ebox and presence in \textit{HLH-1} ChIP regions. It is a plausible candidate to bind \textit{UNC-120} or an intimate \textit{UNC-120} collaborator, as the motif is preferentially concentrated near genes positively regulated by \textit{unc-120}, significantly enriched in genes annotated for \textit{NSM} and \textit{BWM} (Fig. 4E; Supplemental Fig. S4), and present in more than half of archetypal \textit{BWM} candidate CRMs. Several other discovered motifs also colocalize with \textit{HLH-1} binding, but more weakly, and these are reasonable candidates to collaborate with \textit{HLH-1} in other target gene subgroups (Supplemental Fig. S3). For example, CGnnCGAGCA CCC is enriched near genes...
positively regulated by hlh-1 but negatively regulated by unc-120. This pattern is expected for BWM genes that must be turned off in NSM, a group with which it has significant overlap. In contrast, GCCGATTTGCGT is selectively depleted near genes positively regulated by either hlh-1 or unc-120, suggesting that this motif mediates a function that is orthogonal to muscle differentiation (Fig. 4E). None of the newly discovered motifs closely resemble known transcription factor binding sites, although some bear resemblance to previously reported muscle-associated motifs (Guhathakurta et al. 2002; GuhaThakurta et al. 2004), and they are all candidates to bind hlh-1 interacting factors from the RNAi screen.

**Methods**

Additional details may be found in the Supplemental Material.

**General methods and strains**

We obtained *C. elegans* strains N2, PD4605 (hll-1(cc561)), and RW364 (unc-120(st364)) from the Caenorhabditis Genetics Center (CGC) and cultured them using standard methods (Brenner 1974). To increase the proportion of muscle, we chose to knock down early specification genes (mex-3, elt-1, and skn-1) to permit muscle specification (Fig. 1C; Supplemental Material).

**RNAi feeding**

Bacteria (Ahringer Lab RNAi Library) were used for control (HT115) and mex-3 RNAI feeding, in addition to RNAi knockdown for the synthetic PAT screen. The elt-1, mex-3 and skn-1 inserts were fused in a single vector for triple RNAi feeding (Gouda et al. 2010). Knocking down multiple RNA transcripts can suffer from poor efficiency (Gonczy et al. 2000), but our concatenation technique maintained high penetrance. Effectiveness was measured in aliquots from each biological replicate, with 100% of animals being affected by RNAI (100% embryonic-lethal, with 0% making it to the comma stage morphologically). Containing with phallolidin and DAPI revealed significant enrichment of myosin in N2 worms, though not every cell was converted to muscle. Synchronized worms were grown on seeded NGM special plates with IPTG and carboxy-penicillin at 15°C until the L4 stage, and then at 25°C until gravid adults began egg-laying. Gravid adults were bleached, and the eggs were shaken at 25°C in S-complete media (five embryos/µL) for 400 min to ensure muscle differentiation but avoid tissue necrosis (fig. 1B).

**ChIP-seq**

Immunoprecipitation with an existing anti-HLH-1 polyclonal antibody (Lei et al. 2009) was performed in N2 and hll-1(cc561) animals with a modified protocol (Weinmann and Farnham 2002). While not a null, the mutation effectively destroys HLH-1 function (though not every cell was converted to muscle). Synchronized and DAPI revealed significant enrichment of myosin in N2 worms, the comma stage morphologically). Costaining with phalloidin affected by RNAi (100% embryonic-lethal, with 0% making it to a single vector for triple RNAi feeding (Gouda et al. 2010).

**RNA-seq**

For RNA-based sequencing, embryos were flash-frozen in TRIzol (Sigma) and freeze-cracked on dry ice five times. The embryos were passed through a 21 G needle and a 25 G needle (10× each) to shear the eggshell. The RNA was precipitated, treated with Turbo DNase (Ambion), and dT-purified (Invitrogen Dynabeads Oligo-dT). External quantification standards were spiked in to the mRNA, which was then fragmented to an average length of 200 nt by heating at 94°C in the presence of Mg++ for 90 sec. The fragmented mRNA was then random-primed with hexamers for reverse transcription first strand synthesis, followed by nick translation second strand synthesis using a double-stranded cDNA synthesis kit (Invitrogen). Comparisons were made between two biological replicates that were independently sequenced for each condition, except for untreated hlh-1 mutants, for which only a single pooled sample was sequenced. For RNA sequencing, the average number of reads was 26 ± 9 million reads, the average number of unique reads was 20 ± 7 million reads, and the average number of repeat reads was 7 ± 4 million reads (Supplemental Table S8). Quality control failed on an average of 0.5 ± 0.3 million reads per sequenced lane.

**RNAi feeding for synthetic lethal screening**

Bacteria from the OpenBioSystems RNAi library and the Ahringer RNAi library were used for RNAi feeding of L4 animals for 36 h at 25°C. Adults were then transferred to fresh plates for egg-laying for 4 h at 25°C. Adults were removed, and embryos were allowed to develop for 18–24 h prior to scoring. Embryos were scored for developmental progression using a dissecting microscope. The stage of developmental arrest in embryonic lethal worms was noted as during the twofold stage (Pat) or otherwise.

**Data analysis**

WormBase release WS190 was used for all analysis. Read mapping and read processing were performed with Bowtie and ERANGE (Pepke et al. 2009). Python was used to perform calculations described in the text. Genes associated with stress response (e.g., heat shock genes) were monitored for signs of damage or stress. We looked for enriched motifs near the ChIP-identified binding sites using MEME on sequences within various radii of the binding site. A greedy algorithm-based motif finder reproducibly identified the major nonrepeat motifs found with MEME. Enrichments were determined by χ² and hypergeometric statistical analysis.

**Data access**

All sequencing data have been submitted to the NCBI Gene Expression Omnibus (GEO) (http://www.ncbi.nlm.nih.gov/geo/) for...
“HLH-1 binding in muscle-enriched embryos and RNA expression in muscle-enriched embryos across different mutations” (accession number GSE28563, including GSM707199–GSM707213), “Genome-wide maps of HLH-1 binding in muscle-enriched embryos” (accession number GSE28561, including GSM707199–GSM707202), and “Genome-wide RNA expression in muscle-enriched embryos across different mutations” (accession number GSE28562, including GSM707203–GSM707213).

Acknowledgments

Some nematode strains used in this work were provided by the the Caenorhabditis Genetics Center, funded by the NIH National Center for Research Resources (NCRR). We thank M. Krause for the anti-HLH-1 antibody, E. Fisher for her assistance in generating density graphs, S. Pepe for her assistance in generating conservation graphs, A. Kirilusha for his motif counter, H. Amrhein for assistance in managing genomic databases, M. Anaya and J. Vielmetter for purification of the antibody, I. Antoshechkin and L. Schaeffer for library making and sequencing, J. DeModena for the gator, the National Human Genome Research Institute (NHGRI), the Beckman Institute at Caltech, and the NCRR.

References


Received November 11, 2011; accepted in revised form May 24, 2012.