The Noncanonical Binding Site of the MED-1 GATA Factor Defines Differentially Regulated Target Genes in the *C. elegans* Mesendoderm

Gina Broitman-Maduro,[1,2,3] Morris F. Maduro,[1,2,3] and Joel H. Rothman[1,*]
[1]Department of Biology and Neuroscience
    Research Institute
University of California, Santa Barbara
Santa Barbara, California 93106

Summary

Mesoderm and endoderm in *C. elegans* arise from sister cells called MS and E, respectively. The identities of both of these mesendodermal progenitors are controlled by MED-1 and -2, members of the GATA factor family. In the E lineage, these factors activate a sequential cascade of GATA factors, beginning with their immediate targets, the endoderm-specifying *end* genes. We report that MED-1 binds invariant noncanonical sites in the *end* genes, revealing that the MEDs are atypical members of the GATA factor family that do not recognize GATA sequences. By searching the genome for clusters of these MED sites, we have identified 19 candidate MED targets. Based on their expression patterns, these define three distinct classes of MED-regulated genes: MS-specific, E-specific, and E plus MS-specific. Some MED targets encode transcription factors related to those that regulate mesendoderm development in other phyla, supporting the existence of an ancient metazoan mesendoderm gene regulatory network.

Introduction

Cell fates are differentially apportioned during animal development by the deployment of specific gene regulatory networks involving cross-regulating transcription factors and their target genes (Bolouri and Davidson, 2003; Davidson et al., 2003). In many cases, two distinct networks are activated in the daughters of a single cell. While mechanisms that control such developmentally asymmetric cell divisions have been characterized, there is no example in which unique gene regulatory networks activated in each of the daughters is comprehensively known.

The pathway that specifies endoderm development in *C. elegans* is among the best understood transcriptional regulatory network in this rapidly developing animal (Maduro and Rothman, 2002). This innermost germ layer arises from a single cell, E, that is born by an asymmetric division of the mesendoderm precursor, EMS, the ventral-most cell in the four-cell embryo. The sister of E, the MS cell, engenders most of the mesoderm. Remarkably, despite their radically different fates, E and MS are initially specified by the same tran-

*Correspondence: rothman@lifesci.ucsb.edu
[2]These authors contributed equally to this work.
[3]Present address: Department of Biology, University of California, Riverside, Riverside, California 92521.

scriptional cascade: the maternal bZIP/homeodomain protein SKN-1 directly activates transcription of the *med-1* and *-2* genes, which encode redundant members of the GATA factor family (Figure 1) (Bowerman et al., 1993; Maduro et al., 2001). The differential action of MED-1,2 is regulated by the Tcf/Lef-1-like factor POP-1 (Lin et al., 1998; Lin et al., 1995; Maduro et al., 2002). In MS, POP-1 represses the endoderm-specifying *end-1,3* genes (Maduro and Rothman, 2002; Zhu et al., 1997), allowing MED-1,2 to activate expression of MS-specific target genes. Convergent Wnt, MAP kinase, and *src* signaling pathways activated in the E lineage as a result of a polarizing inductive interaction in EMS convert POP-1 from a repressor to an activator of the endoderm-specifying *end-1,3* genes, thereby allowing both MED-1,2 and POP-1 to promote endoderm development in the E lineage (Rocheleau et al., 1997; Thorpe et al., 1997; Rocheleau et al., 1999; Shin et al., 1999; Bei et al., 2002; Maduro et al., 2001; Maduro et al., 2002).

Results and Discussion

MED-1 binds in vivo to the *end-1,3* promoters present on extrachromosomal transgenes in both the E and MS cells (Maduro et al., 2002), raising the possibility that MS-specific target genes may bind MED-1 in both E and MS through the same recognition sequence as the E-specific *end-1* and *end-3* genes. To investigate this possibility and to initiate analysis of the gene regulatory networks that direct E and MS development, we sought to define the MED-1 sites in *end-1* and *end-3*. We performed electrophoretic mobility shift assays with full-length MED-1 protein and DNA fragments that accurately drive reporter expression in the early E lineage (Figure 2; data not shown). At low concentrations of MED-1, we detected a single MED-1::DNA complex; additional bands appeared as the concentration was increased, consistent with the formation of higher order complexes (Figure 2C). We conclude that MED-1 can bind to the *end-1* and *end-3* promoters in vitro, validating our in vivo results (Maduro et al., 2002).

To identify the MED-1 binding sites in *end-1* and *end-3*, we performed DNase I footprinting analysis. These experiments revealed two MED-1 binding sites in *end-1* and four in *end-3* (Figure 2A). Somewhat unexpectedly, none of these six sites contain the rather flexible HGATAR core sequence typical of GATA factors (Lowry and Atchley, 2000), but instead define the more specific consensus 5′-RRRAGTATAC-3′, henceforth denoted a MED site, which contains an invariant, partially palindromic, core sequence (underlined). To investigate whether the MED sites were the only sites of interaction, we constructed versions of the *end-1* promoter fragment in which one or both MED sites were mutated. While mutation of either site alone did not abrogate the gel shift, loss of both eliminated it entirely, indicating that a single MED site is sufficient for binding (Figure 2C). In contrast, we found that both MED sites were required for in vivo expression from the same promoter
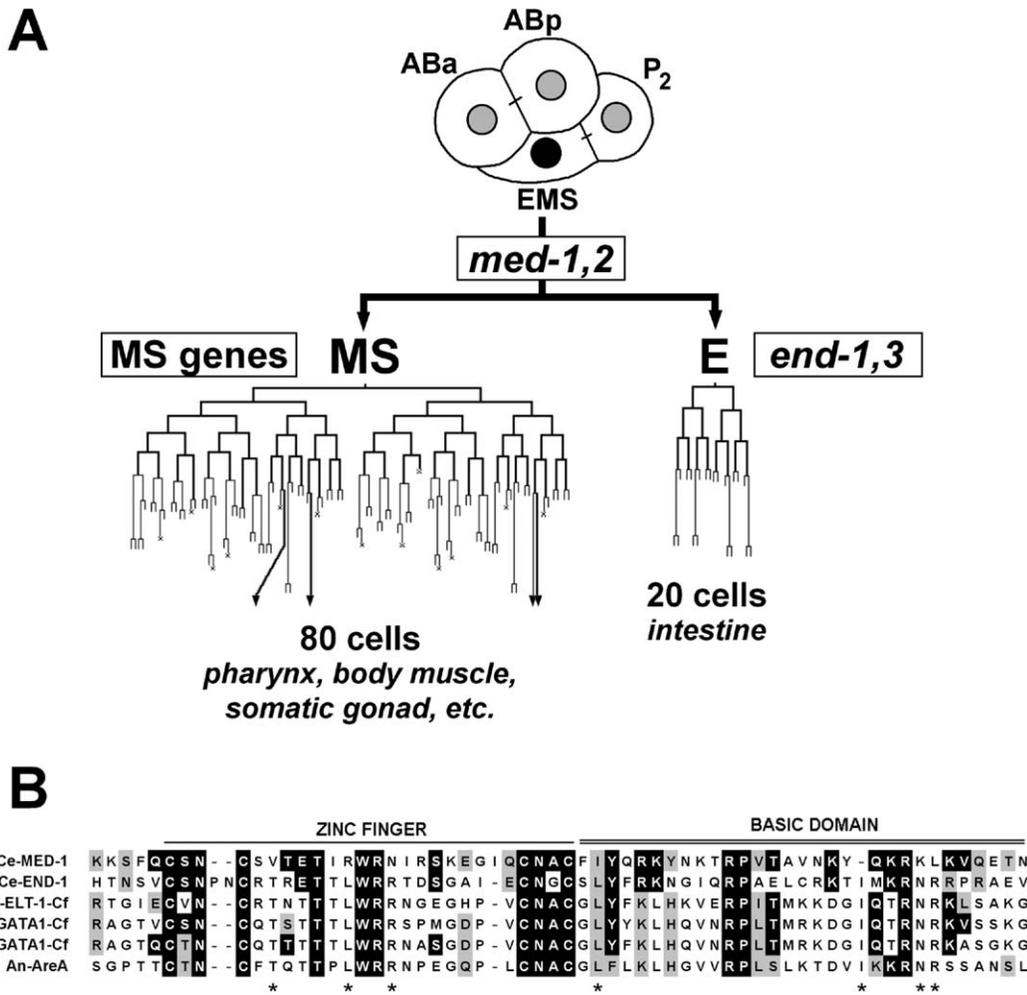
Figure 1. Origin of the MS and E Blastomeres and Alignment of the GATA-Like Zinc Finger of MED-1

(A) The *C. elegans* mesendodermal precursor (EMS), shown at the four-cell stage, generates the MS (mesoderm) and E (endoderm) progenitors, which are specified by the *med-1,2* genes. Although MS and E are sister cells, they give rise to completely different lineage patterns (Sulston et al., 1983). The E fate is specified by the direct activation of *end-1,3* by MED-1,2 (Maduro et al., 2001; Maduro and Rothman, 2002). (B) Alignment of the MED-1 zinc finger and basic domain with those of other GATA factors. Amino acids in black boxes represent identities with MED-1, while gray boxes indicate conservative changes defined by BLOSUM62 (Henikoff and Henikoff, 1992). Residues marked with an asterisk (*) are conserved in all GATA factors shown except for MED-1, possibly accounting for its unusual binding specificity. Ce-MED-1, Ce-END-1, Ce-ELT-1 are from *C. elegans* (Accession numbers CAA92204, T37244, and CAA92494, respectively); cGATA1 is from chicken (*G. gallus*, A32993); mGATA1 is from mouse (*M. musculus*; P17679); The fungal An-AreA is from *A. niger* (X99940). "Cf" indicates the carboxyl finger of a two-fingered protein.

fragment (Figure 2D). While the *end-1* and *end-3* promoter fragments used for the binding studies contain a consensus GATA site, removal of both MED sites (but with the GATA site intact) eliminated MED-1 binding to the *end-1* fragment. Moreover, neither WGATAR site in *end-1* or *end-3* showed any evidence of DNase I protection by MED-1 (data not shown). The finding that MED-1 does not recognize a canonical binding site was not predicted from its primary amino acid sequence, as its single C_4 zinc finger most strongly resembles those of the GATA class of transcription factors (Lowry and Atchley, 2000; Maduro et al., 2001). However, its finger is different at some of the key residues known to make base-specific contacts within the GATA core, and the finger overall is the most divergent of the *C. elegans*

GATA factors (Figure 1B) (Lowry and Atchley, 2000; Omichinski et al., 1993; Starich et al., 1998).

The rigid consensus and apparent requirement of at least two MED-1 sites for in vivo expression of *end-1* suggested that it might be possible to identify new mesendoderm genes by searching for regulatory regions that contain clusters of MED sites. Using WormEnhancer (http://www.wormenhancer.org), we identified 12 sequences in the *C. elegans* genome containing at least two nonoverlapping MED sites within 100 bp of each other. Based on preliminary characterization of genes associated with these clusters, we widened the search consensus to 5′-RAGTATAC-3′ and identified 50 clusters. We narrowed the list to 21 candidate genes based on the proximity of the clusters to a nearby coding re-
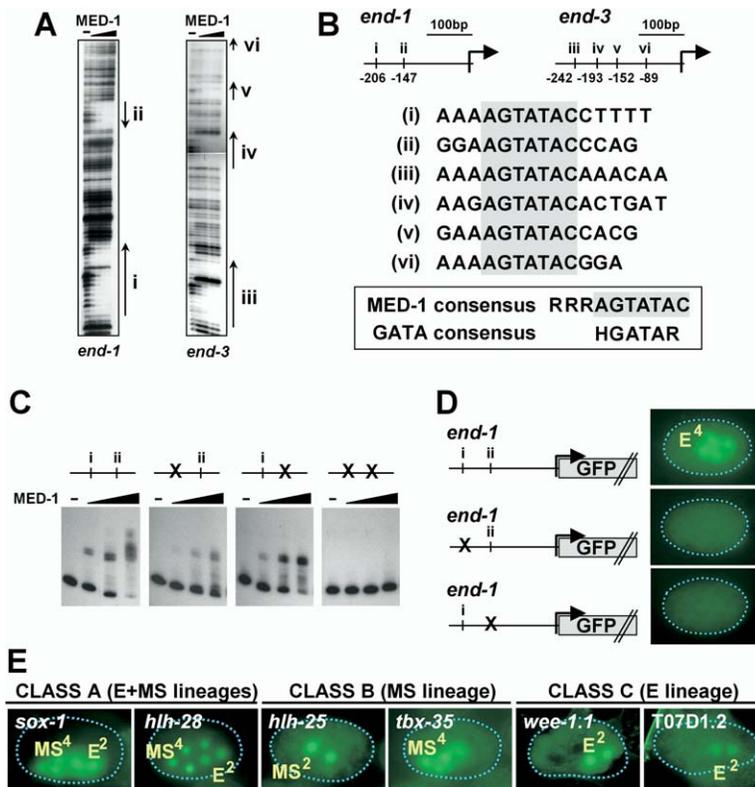
Figure 2. MED-1 Binding Sites and Expression of Putative Target Genes

(A) DNase I footprint analysis of *end-1* and *end-3* promoter fragments. The fragments used contained either 236 bp upstream of the *end-1* start codon or 284 bp upstream of the *end-3* start codon; in both cases, these fragments are sufficient to drive faithful E-specific expression of a reporter construct. Two sites (i-ii) are protected in *end-1*, while four sites (iii-vi) are protected in *end-3*.

(B) Location of MED-1 sites in *end-1* and *end-3*, shown as distance upstream of the start codon, and alignment of DNase I-protected sites to form a consensus with an invariant core sequence (gray shading). The MED-1 binding site does not conform to the canonical GATA binding site.

(C) Gel shift with wild-type and point-mutated MED sites. Substitution of each MED-1 site in the *end-1* promoter (shown as "X" in the diagrams) permits MED-1 binding at the other site, but mutation of both sites abolishes the MED-1 shift. Site (i) was mutated from 5′-aagtatacc-3′ to 5′-aactagtcc-3′ and site (ii) was mutated from 5′-aagtat acc-3′ to 5′-aagtctaga-3′.

(D) Fluorescence micrographs of ~100-cell embryos showing expression of minimal *end-1* promoter::GFP fusions. Mutation of either MED-1 site abolishes expression.

(E) Reporter expression of putative MED-1 target genes falls into three classes. The Class A genes *sox-1* and *hlh-28* are expressed in both the MS and E lineages, the genes *hlh-25* and *tbx-35* are expressed in the MS lineage (Class B), and *wee-1.1* and T07D1.2 are expressed in the early E lineage (Class C). In embryo images, anterior is to the left, and dorsal is up; the eggshell is indicated with a dotted line.

gion (i.e., within 2 kb upstream, or within a first intron; Table 1). The majority of such clusters (15/21) contain two MED sites, 5/21 contain three sites, and a single gene (*end-3*) contains four sites. Approximately half (10) of the putative MED target genes are predicted to be transcription factors. Two genes (F49B2.4 and F35H8.7/

Table 1. Putative MED-1 Target Genes Identified by MED Binding Site Clusters

Cluster	Number of Sites	Nearest Gene (Distance to ATG)	Gene Family	Reporter Expression[a]	Class[b]
1	3	C32E12.5 (1st intron), *sox-1*	Sry-like HMG box (Sox)	early MS and E lineages	A
2	2	ZK849.2 (300bp)	PDZ domain	none detected	
3	2	F49B2.4 (1.7kbp)	cyclin-like F-box	n.d.	
4	3	ZK177.10 (150bp), *tbx-35*	T-box	early MS lineage	B
5	2	C17C3.7 (1st intron), *hlh-25*	bHLH	early MS lineage	B
6	2	C17C3.10 (1st intron), *hlh-27*	bHLH	early MS lineage, others	B
7	2	F35H8.7 (250bp), *wee-1.1*	wee1 kinase	early E lineage, others	C
8	2	K10F12.6 (1.4kbp)	uncharacterized	n.d.	
9	2	F31E3.2 (intron), *ceh-20*	homeobox	none detected	
10	3	B0303.8 (200bp)	uncharacterized	none detected	
11	2	Y66A7A.8 (1st intron), *tbx-33*	T-box	n.d.	
12	2	F13E9.6 (150bp)	uncharacterized	n.d.	
13	2	K02H11.7 (50bp), *str-243*	chemoreceptor	n.d.	
14	2	F58G4.4 (200bp)	LAG-2-like	early E lineage	C
15	2	T11A5.5 (100bp)	uncharacterized	early E lineage	C
16	2	F58E10.2 (220bp), *end-1*	GATA factor	early E lineage	C
17	4	F58E10.5 (250bp), *end-3*	GATA factor	early E lineage	C
18	2	T07D1.2 (250bp)	uncharacterized	early E lineage	C
19	2	B0563.1(600bp)	SRP-RNP	n.d.	
20	3	F31A3.4 (200bp), *hlh-29*	bHLH	early MS, E lineage	A
21	3	F31A3.2 (200bp), *hlh-28*	bHLH	early MS, E lineage	A

n.d., not determined. All genes are defined by the gene models described in Wormbase (http://www.wormbase.org) release WS110.
[a] "E" or "MS" lineage indicates that expression was visible in the early lineage of E or MS in embryos younger than the E[8]/MS[16] stage.
[b] Class A, expression in both MS and E lineages; B, MS lineage only; C, E lineage only.

wee-1.1) are predicted to have roles in the cell cycle. *wee-1.1* has been previously shown to be expressed in the E cell and repressed in MS in a POP-1-dependent manner (Wilson et al., 1999), suggesting that *wee-1.1*, and at least some of the other candidate genes, are bona fide MED-1 targets. Finally, we note that six of the MED clusters appear to represent recently duplicated genes (*hlh-25/hlh-27*, *end-1/3* and *hlh-28/29*), suggesting that genetic redundancy by gene duplication is a feature of both the mesoderm and endoderm networks in *C. elegans* (Maduro and Rothman, 2002).

We constructed GFP reporters for several of the putative MED-1 target genes to assess their expression patterns. Of 15 genes tested, 12 were expressed in the early EMS descendants, suggesting that they are directly activated by the MEDs. These genes fell into three classes (Table 1; Figure 2E). Class A, represented by three genes, *sox-1*, *hlh-28*, and *hlh-29*, are expressed in all cells of the early MS and E lineages; thus, these genes are apparently activated wherever MED-1,2 are expressed. Class B includes three putative transcription factors, *tbx-35*, *hlh-25*, and *hlh-27*, expressed exclusively in the early MS lineage. The six Class C genes, *end-1*, *end-3*, *wee-1.1*, a gene encoding a Notch-like receptor (F58G4.4), and two uncharacterized genes, T11A5.5 and T07D1.2, show E lineage-specific expression.

To assess whether expression of these genes depends on MED activity, we examined the expression of the MS-specific *tbx-35* and *hlh-25* genes in greater detail. Expression of both genes was found to be altered in genetic backgrounds in which MED-1 is absent or ectopically expressed (Figures 3B-3D): in all cases, expression of these genes coincided with that of *med-1* and specification of an MS fate. Finally, we used the "nuclear spot assay" (Fukushige et al., 1999) to test for interaction of GFP::MED-1 with the *tbx-35* and *hlh-25* promoters in vivo. When an *end-1::lacZ* reporter is introduced into a *med-1*::GFP::MED-1 strain, the GFP becomes localized to discrete spots within nuclei of EMS descendants, representing the intranuclear location of the extrachromosomal transgene (Maduro et al., 2002). When a *tbx-35::lacZ* or *hlh-25*::lacZ reporter is used, similar spots form, consistent with in vivo interaction of GFP::MED-1 with their promoters in both the MS and E lineages (Figures 3E-F; data not shown). Collectively, these findings demonstrate that *tbx-35* and *hlh-25* are bona fide MED-1 target genes. As these genes are activated only in the MS lineage, a mechanism must exist to prevent their MED-dependent activation in the E lineage (Figure 3G), analogous to POP-1 repression of *end-1,3* in the MS lineage (Maduro et al., 2002).

The failure of MS to be specified in *skn-1(–)* or *med-1, 2(–)* mutant embryos results in its transformation into a C-like mesectodermal precursor (Bowerman et al., 1992; Maduro et al., 2001). We performed RNAi experiments in an effort to examine the functional requirement for the MS lineage-specific *tbx-35, hlh-25/27*, or *hlh-28/29* genes in specification of the early MS lineage. We were unable to observe any phenotype consistent with failure of MS specification in *tbx-35(RNAi), hlh-25(RNAi), hlh-25(RNAi); hlh-28(RNAi)* or *tbx-35(RNAi); hlh-25(RNAi)* using the potent delivery method of gonadal dsRNA injection (Fire et al., 1998). One inter-

pretation for this finding is that these genes function redundantly with other regulators. Evidence for the function of genetically redundant genes may be obtained by assessing whether their expression is sufficient to transform the fates of cells in which they are normally not expressed. For example, although apparent null mutations of *end-1* show no phenotype owing to genetic redundancy with *end-3* (M.M. and J.H.R., unpublished data), we previously found that overexpression of the END-1 GATA factor is sufficient to specify an endodermal fate in all early non-endodermal somatic cells (Zhu et al., 1998). We found that overexpression of *hlh-25* driven by a heat-shock promoter:: *hlh-25* fusion transgene during embryogenesis results in arrested embryos, consistent with possible respecification of cell fates. To test whether ectopic *hlh-25* activity is sufficient to specify mesoderm tissues normally generated in the MS lineage, we examined the effect of broadly expressing hs-*hlh-25* in mutant backgrounds that greatly reduce the amount of pharynx or body wall muscle. In *skn-1* mutant embryos, no pharynx tissue is made, and ectopic C lineage-type muscle is instead produced by the descendants of MS and E (Bowerman et al., 1992). The maternal Caudal-like homeodomain protein PAL-1 is required for specification of C fate, including C-derived muscle; thus, *skn-1; pal-1* double mutant embryos produce very few muscle cells (Hunter and Kenyon, 1996). We compared heat-shocked *skn-1(RNAi)* and *skn-1(RNAi)*; hs-*hlh-25* embryos and observed no difference in the amount of pharynx muscle made (data not shown). In contrast, a fraction of heat-shocked *skn-1(RNAi); pal-1(RNAi);* hs-*hlh-25* embryos contained extra muscle cells compared to heat-shocked *skn-1(RNAi); pal-1(RNAi)* embryos (Figures 3H–3J). Although this effect was not as dramatic as that observed with overexpressed *end-1* (Zhu et al., 1998), these data suggest that *hlh-25* is sufficient to direct a program of body muscle development in at least some cells. Elucidating the requirements for *hlh-25*, *tbx-35*, and other regulators for MS fate will likely require eliminating their functions in many combinations and isolating chromosomal mutations in them.

It is notable that we were able to identify a set of apparent MED-1 targets reliably by searching for regulatory regions containing MED site clusters. Of the many possible sites that a transcription factor can recognize in the genome, typically only a small fraction are actually bound, or function, in vivo, presumably owing to differences in chromosomal context or requirements for coregulators that bind cooperatively (Iyer et al., 2001; Ren et al., 2000; Wyrick and Young, 2002). We might therefore have expected that only a small fraction, rather than 80%, of the putative MED-1 targets tested would have shown EMS lineage-specific expression. Genome-wide expression analysis has been used to study genes that are coregulated under differing genetic, developmental, or environmental conditions (e.g., Wyrick and Young, 2002). This strategy was used to identify candidate target genes of the *C. elegans* organ identity gene PHA-4, an approach that was successful because PHA-4 is required continuously for expression of pharynx genes (Gaudet and Mango, 2002). To identify genes regulated by a particular transcription factor, genome-wide location analysis can be used to identify
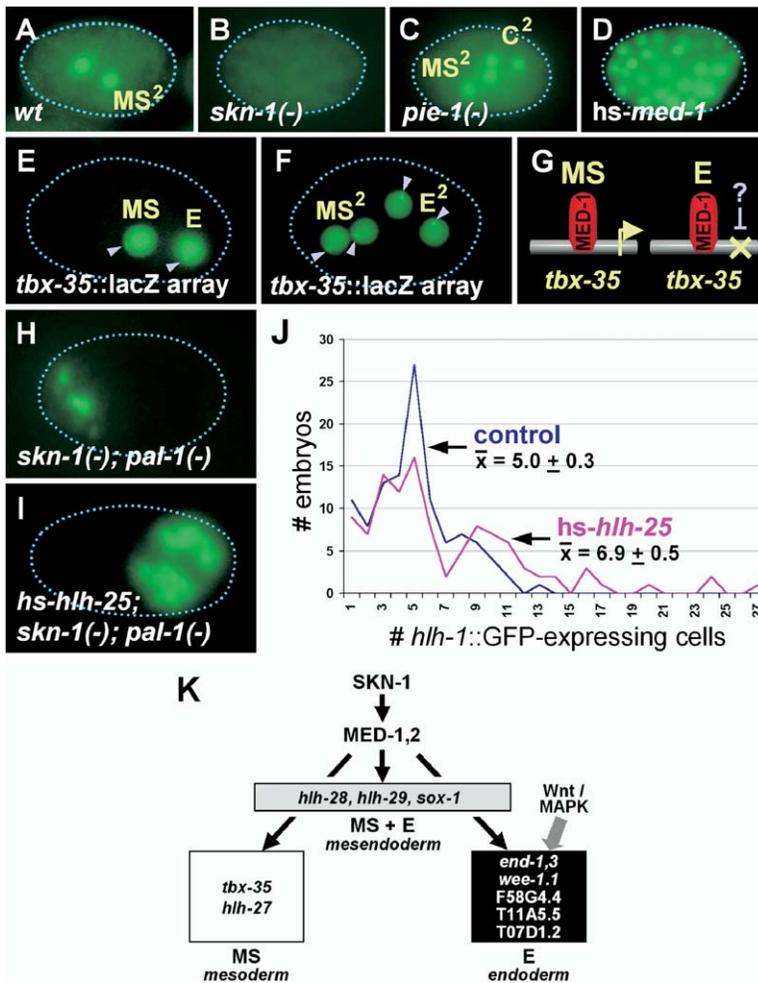
Figure 3. Evidence That MED-1 Regulates Putative Targets In Vivo and That These May Function in Specification of MS-Derived Fates

(A–D) *hlh-25*::GFP responds appropriately to perturbations in *med-1* activity. (A) Wild-type *hlh-25*::GFP in MS daughters. (B) *skn-1(RNAi)* embryos, which do not express *med-1,2* and fail to specify MS (Bowerman et al., 1992; Maduro et al., 2001), also fail to express *hlh-25*. (C) In *pie-1* mutants, the C blastomere usually adopts the fate of MS (Mello et al., 1992), and *hlh-25* is ectopically expressed in the C daughters. (D) Ectopic MED-1 produced by a heat-shock construct is sufficient to drive widespread *hlh-25* expression.

(E–G) Nuclear spot assay validates MED-1 interaction with *tbx-35* in living embryos. GFP::MED-1 forms subnuclear spots in the presence of a *tbx-35*::lacZ transgene array in the nuclei of MS and E nuclei (E) and their daughters (F), even though *tbx-35* is expressed only in the MS lineage. The mechanism that prevents activation of *tbx-35* in the E cell is unknown (G).

(H–J) Overexpression of *hlh-25* by heat shock (hs) directs muscle cell specification. (H) Mutants lacking *skn-1* and *pal-1* activities make very little muscle, as shown by the muscle marker *hlh-1*::GFP (Hunter and Kenyon, 1996; Krause et al., 1990). (I) Overexpression of *hlh-25* from a heat shock construct early in development restores muscle differentiation to a small proportion of these embryos. (J) Distribution of embryos displaying muscle cells following heat shock of control *skn-1(RNAi); pal-1(RNAi)* embryos and those carrying an *hs-hlh-25* transgene. The increase in numbers of muscle cells in *hs-hlh-25* animals is significant (p<0.001).

(K) Model of transcriptional regulatory hierarchies in *C. elegans* mesendoderm specification. Maternal SKN-1 directly activates *med-1* and *-2* (Maduro et al., 2001), which in turn directly control expression of target genes specific for mesoderm, endoderm, or mesendoderm. The Wnt/MAPK pathway input dictates specificity of target gene activation for those targets whose expression is restricted to MS or E.

the promoter regions bound by the regulator in vivo (Wyrick and Young, 2002). Our successful in silico identification of genome-wide MED-1 targets is likely a result of the imposed constraints of the search (i.e., the close occurrence of >2 MED sites in proximity to the start of a coding region) combined with the rigid sequence requirements of the MED consensus binding site. Further, MED-1,2 function during a very short time in early embryonic development, apparently by initiating, but not maintaining, the expression of downstream targets (Maduro and Rothman, 2002).

It is possible that the genome is generally receptive to binding and activation through pairs of MED sites during early phases of development and, therefore, that there is strong selection against these sequences near genes whose mesendodermal expression would be detrimental. This would predict that expression of MED-1 is alone sufficient to activate a mesendodermal program of development; we have observed that ectopic expression of *med-1*, which occurs in certain mutant backgrounds, is necessary and sufficient to specify ectopic mesendoderm fates (Maduro et al., 2001). There is precedent for GATA factors possessing the ability to enhance the accessibility of DNA to the transcription machinery, in addition to functioning as activators: the fungal GATA factor AreA has been shown to be involved in chromatin remodeling (Muro-Pastor et al., 1999), and vertebrate GATA4 can bind to its site in compacted chromatin and establish an open chromatin state (Cirillo et al., 2002). Together, these observations suggest that the primary determinant for MED-1,2 activation of a gene may be the presence of MED binding sites in its promoter.

The apparent participation of *sox-1*, *tbx-35*, and *hlh-25* in the pathway for EMS development raises the possibility that we have identified elements of an ancient conserved mesendoderm gene regulatory network (e.g., Rodaway and Patient, 2001) that is used by both protostomes (e.g., *C. elegans*) and deuterostomes (e.g., vertebrates and echinoderms). SOX-1 belongs to the Group C *Sox* genes (Bowles et al., 2000; Wegner, 1999), of which murine *Sox-4* is a member (Schilham et al., 1996). Mice lacking *Sox-4* display defects in the development of cardiac tissue and blood, two mesodermal tissues (Schilham et al., 1997; Schilham et al., 1996). TBX-35 is a member of the T-box family, defined by the

vertebrate *Brachyury* gene, which is required for mesoderm formation in a variety of animals (Cunliffe and Smith, 1992; Herrmann and Kispert, 1994; Stennard et al., 1996; Willison, 1990). HLH-25 is a member of the bHLH family, of which MyoD, myogenin, Myf5, and MRF4 function in a cascade controlling muscle development (Molkentin and Olson, 1996). While *hlh-1* is the proposed MyoD homolog (Krause et al., 1990), *hlh-25*, which is expressed upstream of *hlh-1*, may be analogous to another member of the vertebrate cascade that functions upstream of MyoD. The endomesoderm gene regulatory network in the sea urchin contains SoxB1 (a member of a different class of Sox gene), a *Brachyury* homolog, and the GATA factors GataE and GataC (Davidson et al., 2002), although the relationships between some of these regulators appears to diverge from what we have observed here. It is clear that further deconstruction of the mesendoderm gene regulatory network in nematodes and other animals is required to assess the degree to which the architecture of these networks has been maintained over >600 million years.

The genes we have identified define three classes of responsiveness to the MED transcription factors positioning them centrally in a hierarchical network that initiates mesendoderm development in *C. elegans* (Figure 3K). Repression of E-specific genes by POP-1 is insufficient to explain the differential action of the MEDs on the two nonequivalent daughter cells of EMS, as revealed by the existence of class B target (MS-specific) genes. It will be of particular interest to assess how the MED transcription factors (which are nearly identical in sequence and expression) differentially activate E-specific and MS-specific gene expression. For example, does POP-1 repress class B genes in the E lineage or activate them in the MS lineage? The identification of these three classes of MED-regulated genes will make it possible to dissect the regulatory code that establishes differential activation of gene regulatory networks in daughter cells produced by asymmetric cell divisions.

Experimental Procedures

Cloning and Strains
Reporter GFP constructs were generated by using PCR to amplify from a region several hundred base pairs upstream of the MED-1 sites to a similar distance downstream of the predicted ATG. Products were cloned into an NLS::GFP and/or NLS::*lacZ* expression vector (from A. Fire). Oligonucleotide sequences and cloning details are available on request. Transgenic *unc-119(ed4)* animals were generated by coinjection of a reporter fusion plasmid and an *unc-119(+)* plasmid. Spot assay strains were generated by introducing *lacZ* reporter arrays into a GFP:MED-1 strain as described in Maduro et al. (2002). dsRNA was synthesized from T7-tagged PCR products and injected into the gonad of hermaphrodites as described (Maduro et al., 2001). Strain *ccIs7963 V* (*hlh-1*::GFP) was a gift from J. Hsieh and A. Fire.

For the heat shock experiments, the predicted *hlh-25* coding region was cloned into the heat shock vectors pPD49.78 and pPD49.83 (gifts from A. Fire). Control animals carrying an *hlh-1* reporter (*ccIs7963*) and an hs-*hlh-25* transgene marked by rescue of *unc-119* were injected with *skn-1* and *pal-1* dsRNA and allowed to recover for 18 hr at 20°C. Gravid hermaphrodites were heat shocked for 30 min at 33°C, allowed to lay eggs for 5 hr at 20°C, and the eggs scored for the number of *hlh-1*::GFP-expressing nuclei after a further 10 hr.

Purification of Recombinant MED-1
The MED-1 coding region was cloned as a PCR product into the 6× His fusion plasmid pET15b. The fusion protein was expressed in *E. coli* codon plus RIL cells (Novagen) by growth at 37°C to an OD of 0.3 followed by induction with 0.1 mM of isopropyl-β-D-thiogalactopyranoside overnight at 30°C. Cells were pelleted by centrifugation and resuspended in BugBuster HT (Novagen) according to the manufacturer's instructions. The recombinant protein was found to be completely insoluble and was therefore purified as follows: inclusion bodies were washed 2× with 10% Bugbuster and 2× with water, then denatured in 8 M urea, 50 mM Tris (pH 7.5), and 25mM DTT on ice. Protein was renatured in a 10× volume of 1M NDSB-256 (Novagen), 50 mM Tris (pH 7.5), 500 mM NaCl, 0.5 mM DTT, 0.5mM zinc sulfate, and 10% glycerol (Vuillard et al., 1998). The refolded extract was then purified on a Ni^{2+} affinity column (Amersham). The His tag was removed by thrombin cleavage and the protein stored at −20°C in 50% glycerol, 1 mM DTT, 10 mM Tris (pH 7.5).

Gel Shift and DNaseI Footprinting Experiments
Probes for EMSA were gel-purified PCR products generated with a ^{32}P end-labeled primer and an unlabeled primer. The regions amplified were from −233 bp to +27 bp relative to the *end-1* ATG, and −302 bp to −9 bp relative to the *end-3* ATG. Gel shift and DNase-I footprinting reactions were performed at 23°C in 10% glycerol, 10 mM Tris (pH 7.5), 50 mM KCl, 0.5 mM DTT and 25 pM ^{32}P-end-labeled probe with 12, 24, 48 nM MED-1. Following a 20 min incubation, protein-DNA complexes were separated by EMSA on a 9% acrylamide gel in 0.5% TGE buffer (12.5 mM Tris, 95 mM glycine, 5 mM EDTA [pH 8.3]) containing 2.5% glycerol at 8 V/cm at 22°C. For footprinting, the complexes were treated with 2 U of DNaseI (Ambion) for 1 min and then organically extracted with phenol. The aqueous phase was desalted using a P6 column (BioRad), lyophilized to 2 μL, resuspended in formamide gel loading buffer, and analyzed on a denaturing 7% sequencing gel according to standard protocols.

References

Bei, Y., Hogan, J., Berkowitz, L.A., Soto, M., Rocheleau, C.E., Pang, K.M., Collins, J., and Mello, C.C. (2002). SRC-1 and Wnt signaling act together to specify endoderm and to control cleavage orientation in early C. elegans embryos. Dev. Cell *3*, 113–125.

Bolouri, H., and Davidson, E.H. (2003). Transcriptional regulatory cascades in development: initial rates, not steady state, determine network kinetics. Proc. Natl. Acad. Sci. USA *100*, 9371–9376.

Bowerman, B., Eaton, B.A., and Priess, J.R. (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.

Bowerman, B., Draper, B.W., Mello, C.C., and Priess, J.R. (1993). The maternal gene skn-1 encodes a protein that is distributed unequally in early C. elegans embryos. Cell *74*, 443–452.

Bowles, J., Schepers, G., and Koopman, P. (2000). Phylogeny of the SOX family of developmental transcription factors based on sequence and structural indicators. Dev. Biol. *227*, 239–255.

Cirillo, L.A., Lin, F.R., Cuesta, I., Friedman, D., Jarnik, M., and Zaret, K.S. (2002). Opening of compacted chromatin by early develop-

mental transcription factors HNF3 (FoxA) and GATA-4. Mol. Cell *9*, 279–289.

Cunliffe, V., and Smith, J.C. (1992). Ectopic mesoderm formation in Xenopus embryos caused by widespread expression of a Brachyury homologue. Nature *358*, 427–430.

Davidson, E.H., Rast, J.P., Oliveri, P., Ransick, A., Calestani, C., Yuh, C.H., Minokawa, T., Amore, G., Hinman, V., Arenas-Mena, C., et al. (2002). A provisional regulatory gene network for specification of endomesoderm in the sea urchin embryo. Dev. Biol. *246*, 162–190.

Davidson, E.H., McClay, D.R., and Hood, L. (2003). Regulatory gene networks and the properties of the developmental process. Proc. Natl. Acad. Sci. USA *100*, 1475–1480.

Fire, A., Xu, S., Montgomery, M.K., Kostas, S.A., Driver, S.E., and Mello, C.C. (1998). Potent and specific genetic interference by double-stranded RNA in Caenorhabditis elegans. Nature *391*, 806–811.

Fukushige, T., Hendzel, M.J., Bazett-Jones, D.P., and McGhee, J.D. (1999). Direct visualization of the elt-2 gut-specific GATA factor binding to a target promoter inside the living Caenorhabditis elegans embryo. Proc. Natl. Acad. Sci. USA *96*, 11883–11888.

Gaudet, J., and Mango, S.E. (2002). Regulation of organogenesis by the Caenorhabditis elegans FoxA protein PHA-4. Science *295*, 821–825.

Henikoff, S., and Henikoff, J.G. (1992). Amino acid substitution matrices from protein blocks. Proc. Natl. Acad. Sci. USA *89*, 10915–10919.

Herrmann, B.G., and Kispert, A. (1994). The T genes in embryogenesis. Trends Genet. *10*, 280–286.

Hunter, C.P., and 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.

Iyer, V.R., Horak, C.E., Scafe, C.S., Botstein, D., Snyder, M., and Brown, P.O. (2001). Genomic binding sites of the yeast cell-cycle transcription factors SBF and MBF. Nature *409*, 533–538.

Krause, M., Fire, A., Harrison, S.W., Priess, J., and Weintraub, H. (1990). CeMyoD accumulation defines the body wall muscle cell fate during C. elegans embryogenesis. Cell *63*, 907–919.

Lin, R., Thompson, S., and Priess, J.R. (1995). pop-1 encodes an HMG box protein required for the specification of a mesoderm precursor in early C. elegans embryos. Cell *83*, 599–609.

Lin, R., Hill, R.J., and Priess, J.R. (1998). POP-1 and anterior-posterior fate decisions in C. elegans embryos. Cell *92*, 229–239.

Lowry, J.A., and Atchley, W.R. (2000). Molecular evolution of the GATA family of transcription factors: conservation within the DNA-binding domain. J. Mol. Evol. *50*, 103–115.

Maduro, M.F., Meneghini, M.D., Bowerman, B., Broitman-Maduro, G., and Rothman, J.H. (2001). Restriction of mesendoderm to a single blastomere by the combined action of SKN-1 and a GSK-3beta homolog is mediated by MED-1 and -2 in C. elegans. Mol. Cell *7*, 475–485.

Maduro, M.F., and Rothman, J.H. (2002). Making worm guts: the gene regulatory network of the Caenorhabditis elegans endoderm. Dev. Biol. *246*, 68–85.

Maduro, M.F., Lin, R., and Rothman, J.H. (2002). Dynamics of a developmental switch: recursive intracellular and intranuclear redistribution of Caenorhabditis elegans POP-1 parallels Wnt-inhibited transcriptional repression. Dev. Biol. *248*, 128–142.

Mello, C.C., Draper, B.W., Krause, M., Weintraub, H., and Priess, J.R. (1992). The pie-1 and mex-1 genes and maternal control of blastomere identity in early C. elegans embryos. Cell *70*, 163–176.

Molkentin, J.D., and Olson, E.N. (1996). Defining the regulatory networks for muscle development. Curr. Opin. Genet. Dev. *6*, 445–453.

Muro-Pastor, M.I., Gonzalez, R., Strauss, J., Narendja, F., and Scazzocchio, C. (1999). The GATA factor AreA is essential for chromatin remodelling in a eukaryotic bidirectional promoter. EMBO J. *18*, 1584–1597.

Omichinski, J.G., Clore, G.M., Schaad, O., Felsenfeld, G., Trainor, C., Appella, E., Stahl, S.J., and Gronenborn, A.M. (1993). NMR

structure of a specific DNA complex of Zn-containing DNA binding domain of GATA-1. Science *261*, 438–446.

Ren, B., Robert, F., Wyrick, J.J., Aparicio, O., Jennings, E.G., Simon, I., Zeitlinger, J., Schreiber, J., Hannett, N., Kanin, E., et al. (2000). Genome-wide location and function of DNA binding proteins. Science *290*, 2306–2309.

Rocheleau, C.E., Downs, W.D., Lin, R., Wittmann, C., Bei, Y., Cha, Y.H., Ali, M., Priess, J.R., and Mello, C.C. (1997). Wnt signaling and an APC-related gene specify endoderm in early C. elegans embryos. Cell *90*, 707–716.

Rocheleau, C.E., Yasuda, J., Shin, T.H., Lin, R., Sawa, H., Okano, H., Priess, J.R., Davis, R.J., and Mello, C.C. (1999). WRM-1 activates the LIT-1 protein kinase to transduce anterior/posterior polarity signals in C. elegans. Cell *97*, 717–726.

Rodaway, A., and Patient, R. (2001). Mesendoderm. an ancient germ layer? Cell *105*, 169–172.

Schilham, M.W., Moerer, P., Cumano, A., and Clevers, H.C. (1997). Sox-4 facilitates thymocyte differentiation. Eur. J. Immunol. *27*, 1292–1295.

Schilham, M.W., Oosterwegel, M.A., Moerer, P., Ya, J., de Boer, P.A., van de Wetering, M., Verbeek, S., Lamers, W.H., Kruisbeek, A.M., Cumano, A., and Clevers, H. (1996). Defects in cardiac outflow tract formation and pro-B-lymphocyte expansion in mice lacking Sox-4. Nature *380*, 711–714.

Shin, T.H., Yasuda, J., Rocheleau, C.E., Lin, R., Soto, M., Bei, Y., Davis, R.J., and Mello, C.C. (1999). MOM-4, a MAP kinase kinase kinase-related protein, activates WRM-1/LIT-1 kinase to transduce anterior/posterior polarity signals in C. elegans. Mol. Cell *4*, 275–280.

Starich, M.R., Wikstrom, M., Arst, H.N., Jr., Clore, G.M., and Gronenborn, A.M. (1998). The solution structure of a fungal AREA protein-DNA complex: an alternative binding mode for the basic carboxyl tail of GATA factors. J. Mol. Biol. *277*, 605–620.

Stennard, F., Carnac, G., and Gurdon, J.B. (1996). The Xenopus T-box gene, Antipodean, encodes a vegetally localised maternal mRNA and can trigger mesoderm formation. Development *122*, 4179–4188.

Sulston, J.E., Schierenberg, E., White, J.G., and Thomson, J.N. (1983). The embryonic cell lineage of the nematode Caenorhabditis elegans. Dev. Biol. *100*, 64–119.

Thorpe, C.J., Schlesinger, A., Carter, J.C., and Bowerman, B. (1997). Wnt signaling polarizes an early C. elegans blastomere to distinguish endoderm from mesoderm. Cell *90*, 695–705.

Vuillard, L., Rabilloud, T., and Goldberg, M.E. (1998). Interactions of non-detergent sulfobetaines with early folding intermediates facilitate in vitro protein renaturation. Eur. J. Biochem. *256*, 128–135.

Wegner, M. (1999). From head to toes: the multiple facets of Sox proteins. Nucleic Acids Res. *27*, 1409–1420.

Willison, K. (1990). The mouse Brachyury gene and mesoderm formation. Trends Genet. *6*, 104–105.

Wilson, M.A., Hoch, R.V., Ashcroft, N.R., Kosinski, M.E., and Golden, A. (1999). A Caenorhabditis elegans wee1 homolog is expressed in a temporally and spatially restricted pattern during embryonic development. Biochim. Biophys. Acta *1445*, 99–109.

Wyrick, J.J., and Young, R.A. (2002). Deciphering gene expression regulatory networks. Curr. Opin. Genet. Dev. *12*, 130–136.

Zhu, J., Hill, R.J., Heid, P.J., Fukuyama, M., Sugimoto, A., Priess, J.R., and Rothman, J.H. (1997). end-1 encodes an apparent GATA factor that specifies the endoderm precursor in Caenorhabditis elegans embryos. Genes Dev. *11*, 2883–2896.

Zhu, J., Fukushige, T., McGhee, J.D., and Rothman, J.H. (1998). Reprogramming of early embryonic blastomeres into endodermal progenitors by a Caenorhabditis elegans GATA factor. Genes Dev. *12*, 3809–3814.