The homeodomain protein PAL-1 specifies a lineage-specific regulatory network in the *C. elegans* embryo

L. Ryan Baugh[1,*], Andrew A. Hill[2], Julia M. Claggett[1], Kate Hill-Harfe[1,2,†], Joanne C. Wen[1], Donna K. Slonim[2,‡], Eugene L. Brown[2] and Craig P. Hunter[1,§]

[1]Department of Molecular and Cellular Biology, Harvard University, Cambridge, MA 02138, USA
[2]Department of Genomics, Wyeth Research, Cambridge, MA 02140, USA
*Present address: Biology Division, California Institute of Technology, Pasadena, CA 91125, USA
[†]Division of Genetics, Department of Pediatrics, University of Florida College of Medicine, Gainesville, FL 32610, USA
[‡]Department of Computer Science, Tufts University, Medford, MA 02155, USA
[§]Author for correspondence (e-mail: hunter@mcb.harvard.edu)

Summary

Maternal and zygotic activities of the homeodomain protein PAL-1 specify the identity and maintain the development of the multipotent C blastomere lineage in the *C. elegans* embryo. To identify PAL-1 regulatory target genes, we used microarrays to compare transcript abundance in wild-type embryos with mutant embryos lacking a C blastomere and to mutant embryos with extra C blastomeres. *pal-1*-dependent C-lineage expression was verified for select candidate target genes by reporter gene analysis, though many of the target genes are expressed in additional lineages as well. The set of validated target genes includes 12 transcription factors, an uncharacterized wingless ligand and five uncharacterized genes. Phenotypic analysis demonstrates that the identified PAL-1 target genes affect specification, differentiation and morphogenesis of C-lineage cells. In particular, we show that cell fate-specific genes (or tissue identity genes) and a posterior HOX gene are activated in lineage-specific fashion. Transcription of targets is initiated in four temporal phases, which together with their spatial expression patterns leads to a model of the regulatory network specified by PAL-1.

Key words: *C. elegans*, Regulatory network, Patterning, Embryonic, Cell fate specification

Introduction

The *C. elegans* embryo develops rapidly with an invariant, fully described lineage (Sulston et al., 1983), allowing determination of defined gene expression states in each cell of the developing embryo. As a first step toward this goal, we previously measured temporal profiles of transcript abundance during development with small cohorts of staged whole embryos (Baugh et al., 2003). Because whole embryos are used, expression in different lineages is integrated in the measurement. However, the phenotypes of previously described mutants offer a genetic approach with which to dissect lineage-specific patterns of gene expression from whole-embryo data.

Each of the six founder blastomeres gives rise to different cell types by characteristic patterns of cell division (lineages) (Sulston et al., 1983). Founder blastomere fates are specified by a variety of spatially and temporally restricted maternal gene activities. In addition, the embryo contains a global anteroposterior (AP) patterning system that differentially specifies the daughters of blastomeres dividing on the AP axis (Kaletta et al., 1997; Lin et al., 1998). When gastrulation commences at the 26-cell stage, all of the founder blastomeres have been born and tissue identity begins to be specified, as indicated by the initial expression of cell fate-specific genes, whose functions are required for the development of specific cell types, and increasing resistance of the embryo to cell fate transformations induced by ectopic expression of such genes (Gilleard, 2001; Zhu, 1998). It has been noted that although such fate-specific genes are expressed in multiple lineages (most tissues being polyclonal), the cells expressing them are born at about the same time and in regional domains as if in creation of a tissue or organ primordium (Labouesse and Mango, 1999). From work on the *C. elegans* pharynx it is clear that organ identity genes control autonomous organ-specific genetic networks (Gaudet and Mango, 2002). However, it remains to be determined how the lineage-based mechanisms in the early embryo work together with the global AP patterning system to pattern tissue and organ identity gene expression, thus causally relating the maternal genetic network that patterns the early embryo with the zygotic networks that pattern later developmental structures.

The fate of the C and D founder blastomeres, the somatic descendants of P2, is specified by the Caudal-like homeobox gene *pal-1*. Maternal PAL-1 activity is temporally and spatially targeted to the C and D founder blastomeres by first restricting translation of maternal *pal-1* mRNA to the descendants of the posterior blastomere P1 (EMS and P2) and then by restricting the activity of the translated protein to the somatic descendants of P2 (C and D) (Hunter and Kenyon, 1996). The KH domain protein MEX-3 is required to restrict translation of maternal

pal-1 mRNA to the posterior blastomeres at the four-cell stage (EMS and P2) (Draper et al., 1996; Huang et al., 2002; Hunter and Kenyon, 1996), while the bZIP transcription factor SKN-1 blocks PAL-1 function in EMS, and the zinc-finger protein PIE-1 maintains the germline blastomeres P2 and P3 in a transcriptionally quiescent state so that PAL-1 activity is restricted to their somatic descendants, C and D (Bowerman et al., 1993; Hunter and Kenyon, 1996; Mello et al., 1996; Seydoux et al., 1996).

The C lineage gives rise primarily to muscle and epidermis but also two neuronal cells and a cell death (Sulston et al., 1983). In the absence of maternal PAL-1 activity, the C and D blastomeres fail to develop in any discernible way, while ectopic PAL-1 activity causes other blastomeres to produce muscle, epidermal and neuronal cells by a C-like lineage (Draper et al., 1996; Hunter and Kenyon, 1996). Although other somatic lineages also give rise to muscle and epidermal cells, the lack of discernable C cell fates in the absence of maternal PAL-1 function (Hunter and Kenyon, 1996), indicates that PAL-1 activates cell fate specification factors (tissue identity genes) in the C lineage.

To learn how maternal PAL-1 activity leads to the patterned specification of multiple cell fates within a single blastomere lineage, we aim to identify the genes directly and indirectly activated by PAL-1 and determine their loss-of-function phenotypes and regulatory interactions. To identify genes expressed in the C lineage by microarray we have used *pie-1* and *mex-3* mutations, as well as *skn-1* RNAi, to produce embryos that either lack a C blastomere or that contain almost exclusively C-like blastomeres. Our results are verified by reporter gene analysis and complemented by phenotypic analysis of PAL-1 targets, and a model for the regulatory network specified by PAL is presented.

Materials and methods

Microarray sample preparation

Wild-type samples were prepared and described previously (Baugh et al., 2003) but re-hybridized here to a different microarray. Strains JJ532 [*pie-1(zu154) unc-25(e156)/qC1[dpy-19(e1259) glp-1(q239)]; III*] and JJ518 [*mex-3(zu155) dpy-5(e61)/hT1; I*] were grown at 15°C either on *E. coli* OP50 or on HT115 (for RNAi), and Unc and Dpy adults were respectively picked and cut for embryo collection. Embryo collection and staging was as described (Baugh et al., 2003) except that embryos were washed in 10 mM NaCl as opposed to water and aged in the lids of 0.6 ml tubes rather than on microscope slides. RNA was extracted and amplified as described (Baugh et al., 2001); for protocol see http://mcb.harvard.edu/hunter/Protocols/protocols. htm.

Microarray hybridization and data reduction

Biotinylated, amplified RNA (1 μg) was hybridized to the Affymetrix *C. elegans* microarray as described (Baugh et al., 2003). Array data were quantile normalized and reduced by the robust multi-chip average algorithm (RMA) (Irizarry et al., 2003), using the Bioconductor Affy package (version 1.0, www.bioconductor.org) for the R statistics software (version 1.5.0, www.r-project.org). All expression levels reported here were back-transformed to the linear scale, i.e. reported values are $2^{(RMA)}$. All raw data have been submitted to the Gene Expression Omnibus database, Accession Number GSE2180, and averaged data and analysis are available in the supplementary material.

Clustering of gene expression profiles

Clusters were generated by a modified version of the QT clustering algorithm (Heyer et al., 1999). This algorithm assembles a series of clusters ordered by size with a defined limit on the largest pair-wise distance allowed between any two profiles in a cluster. Distance between profiles is measured as 1-R, where R is the Pearson correlation coefficient. Although we limited this distance to 0.3, some genes are included in clusters simply by chance. To reduce the spurious inclusion of these genes in the final clusters, we systematically re-sampled our data (100 times) with two forms of synthetic noise added at each reiteration to generate an R_{avg}. Noise was added to $\log_2$ scale RMA expression data, and was generated by a two-component model consisting of an additive Gaussian background with standard deviation 0.2, and a multiplicative Gaussian sampling error with a standard deviation of 0.05. Simulated data were floored at 1 RMA unit. Graphs plotting average expression for each cluster, and the cluster to which each gene belongs can be found in Fig. S1 and Data S1, S2 in supplementary material.

ANOVA

Analysis of variance (ANOVA) was performed by using a randomization test to assess differences in expression among genotypes at each time point. Tests were performed to assess for each gene, overall variation among all three genotypes at each time point, and variation between each of the three possible pairs of genotypes at each time point. All statistical tests were performed on the $\log_2$ scale data. Data for the *pie-1(zu154)* and *pie-1(zu154); pal-1(RNAi)* genotypes were pooled to form a single group, denoted 'pm'. Sample labels were randomly shuffled 100 times, and for each shuffling, at each time point, a null distribution of F-statistics were computed among all three genotypes, and between each of the three pair-wise combinations of genotypes at each time point [N2 versus *mex-3(zu155); skn-1(RNAi)*, N2 versus pm, and *mex-3(zu155); skn-1(RNAi)* versus pm]. *P*-values for differential expression among groups were determined by referring the F-statistics from the observed data to the null distribution arising from the random permutations. *P*-values are not adjusted for multiple testing. The total number of statistical tests per gene was 40 (10 time points with three pair-wise comparisons and 1 overall comparison).

Target scoring

Clustering, correlation to known target genes and ANOVA were used to score the PAL-1 target potential of each gene. Only genes with maximum expression over time and genotype greater than the median of all genes over time and genotype (transcript abundance of thirteen RMA units) were considered as potential targets. Clusters 24, 50, 60, 140, 187, 141, 168, 85, 105, 130, 131, 144, 177, 195 and 88 were selected as potential target clusters. Because our selection of *pal-1* target clusters was subjective, each gene belonging to one of these clusters was given a target score of only 1. We also leveraged our limited prior knowledge to give genes a target score of 1 if they were either one of the ten best correlated with *vab-7* over time and genotype or one of the 100 best correlated with *cwn-1*. *vab-7* and *cwn-1* had been validated as target genes and are both expressed specifically in the C lineage. The decision to include the 10 and 100 best correlated genes for each was based on inspection of expression patterns.

The most rigorous approach was model-based and quantitative ANOVA, which was used to assign a target score of 1-5, depending on the *P*-value assigned to each gene at each time point for the observed differences in expression between the three genotypes. For each gene at each time point it was also determined if it was higher in *mex-3(zu155); skn-1(RNAi)* than wild type and lower in *pie-1(zu154)* than wild type. Genes with a *P*-value less than a given cut-off and appropriate differences between genotypes were noted. With this information, a target score was generated for each gene in two ways and the maximum was kept. The first relies on the gene satisfying both criteria in a pair of adjacent time points. Genes with

a *P*-value below 10^{-2} for two adjacent time points were given a score of 1, those below 10^{-3} were given a score of 2, those below 10^{-4} were given a score of 3, those below 10^{-5} were given a score of 4, and those below 10^{-6} were given a score of 5. The adjacency requirement excludes late genes that show differences between genotypes in only the last time point, and so the second target score requires the gene to satisfy both criteria in only a single time point. Genes with a *P*-value below 10^{-4} were given a score of one, those below 10^{-5} were given a score of two, and those below 10^{-6} were given a score of three. A nominal *P*-value of 10^{-4} (or 10^{-2} twice) with ~10,000 genes considered at ten time points with two models (pair of adjacent time points and single time point) should result in about 20 false positives; however, the actual number of false positives should be less given the additional requirement that the genotypes differ in specific ways.

To assign each gene a final target score the max of the ANOVA-based models is added to the score from cluster analysis (1 or 0) and correlation to known targets (1 or 0) producing a maximum score of 7. The target score therefore relies heavily on the ANOVA analysis, with a score of 7 indicating that the gene is in a potential target cluster, correlates well with either of the known targets, and has a *P*-value below 10^{-6} in a pair of adjacent time points. By contrast, genes with a score of 1 are either in a potential target cluster or well correlated with a known target or have a minimally sufficient *P*-value.

Reporter analysis

Reporter constructs were made by PCR (Hobert, 2002). 5′ genomic sequence, either 5 kb or up to the next gene, was used as promoter. YFP was from pPD132.112, which includes *C. elegans* introns, a nuclear localization signal and the *unc-54* 3′UTR; for additional information see http://www.ciwemb.edu/pages/firelab.html (Fire et al., 1990). Either 4 kb from pRF4 containing a dominant *rol-6* (Mello et al., 1991) or a 2.2 kb sequence *unc-119* rescue sequence from pDPMM051 (Maduro and Pilgrim, 1995) was included in the final PCR construct as co-transformation marker. The *vab-7* reporter (HC14) was made by ligating a 5 kb *Age*I/*Pst*I GFP fragment from

pPD104.53 (Fire et al., 1990) to a 14 kb *Age*I/*Pst*I digestion pJA15 (Ahringer, 1996) to produce pHC16. JK3363 (Mathies et al., 2003) was used rather than our *hnd-1* reporter, as our reporter was too dim to score. Either N2 or CB4845 [*unc-119*(e2498)] was injected with PCR product diluted 10-fold in water as described (Mello et al., 1991); *rol-6* plasmid (pRF4) was co-injected at 50 ng/μl where *rol-6* was used as co-transformation marker, including HC14. Stable extrachromosomal lines were used for initial reporter analysis, and select reporters were chosen for chromosomal integration by gamma irradiation as described (Inoue et al., 2002). Table S1 in the supplementary material provides strain names and oligonucleotide sequences used.
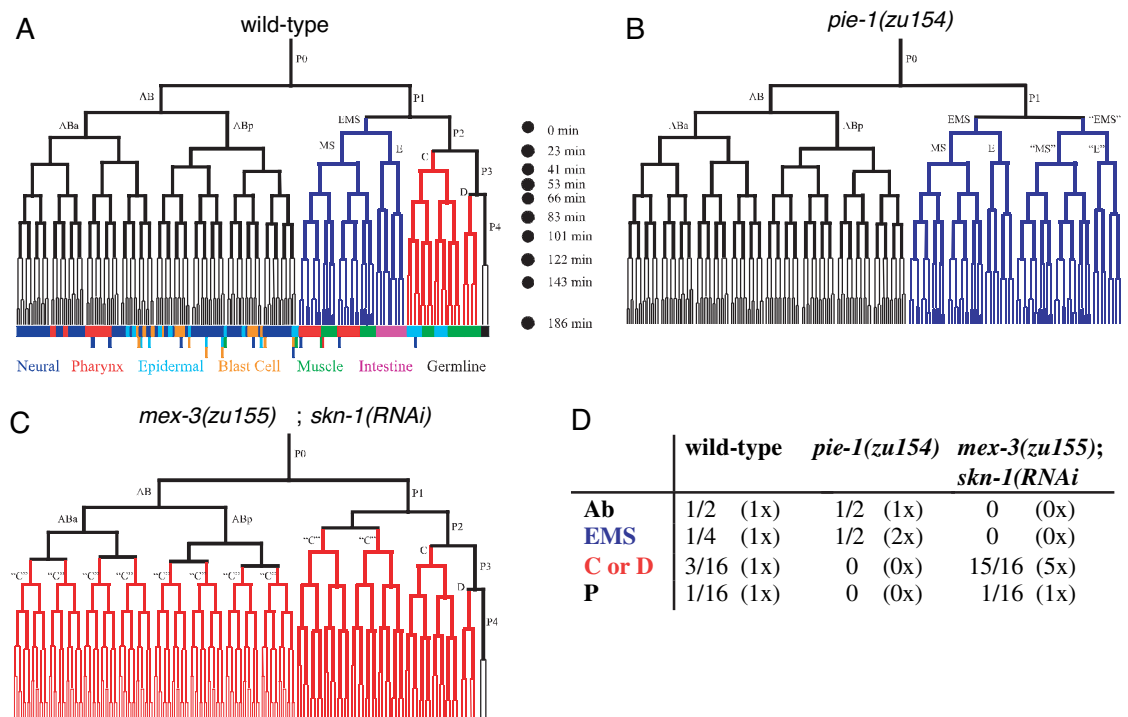
RNAi

Hairpin RNAi feeding vectors were made as described (Winston et al., 2002) for *pal-1*, *mex-3* and *skn-1*, and transformed into *E. coli* HT115; for protocol see http://mcb.harvard.edu/hunter/Protocols/ protocols.htm. JJ532 was grown on both OP50 and HT115 expressing double-stranded *pal-1* RNA, but no differences were detected by microarray (data not shown) and the data were merged. JJ518 was only grown on HT115 expressing double-stranded *skn-1* RNA. For the reporter assay, in addition to being grown on RNAi food, worms were soaked in double-stranded RNA as described (Maeda et al., 2001) (Table 1).

3D imaging

Four-cell embryos were collected from cut mothers by mouth pipette and mounted on 2.0% agarose pads. An Olympus IX70 microscope equipped with Nomarski and fluorescence optics and DeltaVision Spectris software was used to image embryos 210 minutes after the four-cell stage (22°C). Generally, about 40 optical sections along the *z*-axis (ranging from 0.7 to 1.2 μm) were gathered at exposure times ranging from 0.2 to 1.3 seconds. Sections were rendered into three-dimensional images with softWorx Volume Viewer, and projections were generated from maximum intensity voxels every 12° around the

Fig. 1. Use of homeotic mutants to deconvolve lineage-specific expression from transcript profiles. Wild-type and mutant embryos were staged in small cohorts by morphology at the four-cell stage and collected in replicate at the ten time points indicated by black dots in A. The complete lineage through the 190-cell stage is depicted for wild type in A, and hypothetical lineages are depicted for *pie-1(zu154)* in B and *mex-3(zu155)*; *skn-1(RNAi)* in C. Lineages specified by SKN-1 and PAL-1 are blue and red, respectively, and the pattern of cell fates produced is indicated below the wild-type lineage in A. (D) Approximate fraction of the embryo with different lineage identities for each genotype, with expected changes in transcript abundance for lineage-specific genes in parentheses.

	wild-type		*pie-1(zu154)*		*mex-3(zu155)*; *skn-1(RNAi*	
Ab	1/2	(1x)	1/2	(1x)	0	(0x)
EMS	1/4	(1x)	1/2	(2x)	0	(0x)
C or D	3/16	(1x)	0	(0x)	15/16	(5x)
P	1/16	(1x)	0	(0x)	1/16	(1x)

anteroposterior axis of the embryo for a total of thirty projections. Dorsal and lateral projections are presented in Fig. 8, and all 30 projections were assembled as a .mov file in Graphic Converter to be viewed as a movie in Quicktime (see Movies 1-13 in supplementary material).

Results

### Identification of lineage-specific targets of PAL-1

We have previously analyzed transcript abundance in precisely staged embryos, roughly two time points per cell cycle from the one-cell stage to mid-gastrulation (190-cell stage), using custom synthesized high-density oligonucleotide microarrays (Baugh et al., 2003). Because these wild-type data represent the baseline for all subsequent comparisons, we re-hybridized these published samples to commercially available microarrays used for our current analysis and also by other members of the *C. elegans* community. The data are marginally improved by hybridization to the newer microarrays (lower coefficient of variation), and in general they are in very good agreement with their published counterpart. The overall correlation coefficient comparing the previously published data with that resulting from re-hybridization to the commercial microarray is 0.90. The differences observed tend to be small in magnitude and distributed over many genes (data not shown), though there are exceptions (including *end-3*). Because the precision and density of sampled time points in the wild-type data greatly increased sensitivity (Baugh et al., 2003), we collected multiple replicates at similar time points for the mutant embryos beginning at the four-cell stage (Fig. 1).

Embryos from homozygous *pie-1(zu154)* mothers lack *pal-1*-dependent C and D blastomeres because the identity of the P2 blastomere is transformed to that of its somatic sister, EMS (Fig. 1) (Mello et al., 1992). In the absence of the C and D lineages, PAL-1 targets, whether direct or indirect, should have reduced expression in *pie-1(zu154)* relative to wild type, and genes that are exclusively expressed in the C or D lineage should not be detected. A *pal-1* mutant could not be used because maternal activity specifies the C and D fates, and *pal-1(null)* mutants are zygotic lethal. We chose to use *pie-1(zu154)* rather than *pal-1(RNAi)* embryos as the mutation

results in higher penetrance than RNAi of *pal-1* (data not shown). *pie-1(zu154)* embryos contain wild-type levels of PAL-1, but PAL-1 is unable to specify C-lineage identity because, in the absence of PIE-1 function, SKN-1 dominantly controls the development of the P2 lineage (Hunter and Kenyon, 1996). To be certain that the PAL-1 protein present in *pie-1* mutant embryos does not affect transcript abundance, we measured transcript abundance in parallel in *pie-1(zu154); pal-1(RNAi)* embryos. However, the results from the two genotypes were indistinguishable, indicating that PAL-1 is impotent in the *pie-1(zu154)* embryos (data not shown). The aggregate *pie-1(zu154)* and *pie-1(zu154); pal-1(RNAi)* replicates are combined in the *pie-1(zu154)* data we report.

Embryos from homozygous *mex-3(zu155)* mothers translate *pal-1* mRNA throughout the embryo, transforming the eight great-granddaughters of the AB blastomere into C-like blastomeres (Fig. 1) (Draper et al., 1996). These eight anterior blastomeres, born at approximately the same time as the C blastomere, produce eight serially homologous lineages giving rise to the muscle and epidermal cell types characteristic of the C lineage. This transformation requires *pal-1* function (Hunter and Kenyon, 1996), thus expression of PAL-1 targets should be greater in *mex-3(zu155)* than in wild type. To sensitize our ability to detect PAL-1 targets further, *skn-1* RNAi was also used in the *mex-3(zu155)* background, because, in the absence of *skn-1* function, PAL-1 is active in the EMS lineage transforming its daughters into C-like blastomeres (Bowerman et al., 1992; Hunter and Kenyon, 1996).

To test the feasibility of using mutants to deconvolve lineage-specific patterns of gene expression and to develop computational rules to enrich for lineage-specific transcripts, we analyzed a known, *skn-1*-dependent lineage-specific transcriptional cascade. In contrast to PAL-1 targets (genes regulated directly or indirectly by PAL-1), SKN-1 targets should be more abundant in *pie-1(zu154)* and less abundant in *mex-3(zu155); skn-1(RNAi)* (Fig. 1). SKN-1 directly activates its earliest zygotic targets, the redundant GATA transcription factors *med-1* and *med-2*, in the EMS blastomere, initiating a transcriptional cascade resulting in expression of new genes after each cell division (Maduro, 2001; Maduro and Rothman, 2002). *med-1* and *med-2* activate the expression of two more
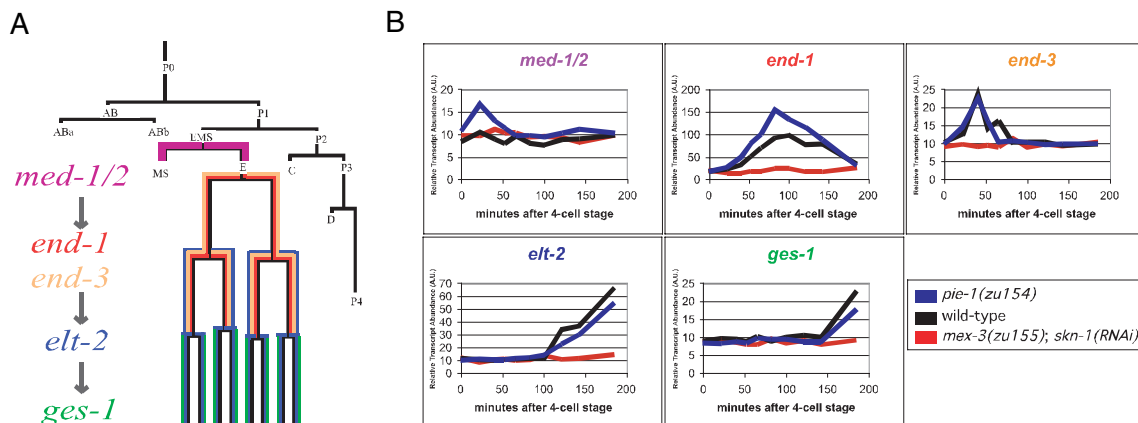
Fig. 2. A *skn-1*-dependent transcriptional cascade. (A) Published expression patterns and regulatory relationships for four transcription factors and a differentiation gene involved in endoderm development. (B) Temporal expression patterns in wild type, *pie-1(zu154)* and *mex-3(zu155); skn-1(RNAi)* measured by microarray. *med-1* and *med-2* cannot be distinguished on the microarray.

A

B

High
Low
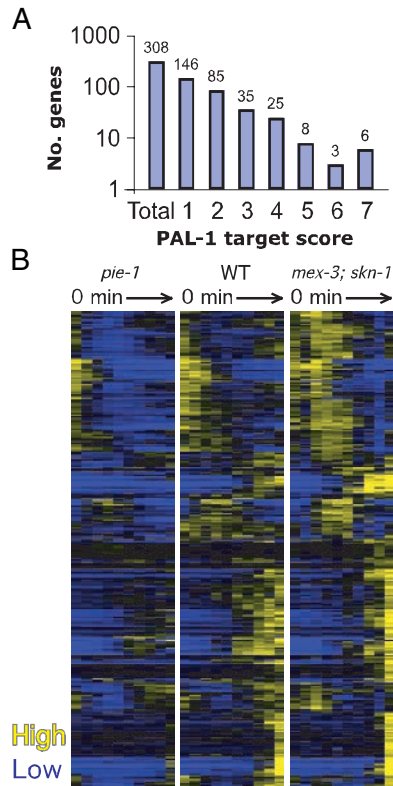
Fig. 3. Identification and rank of candidate PAL-1 target genes. (A) A log-scale histogram of the number of genes versus cumulative scores (1-7) from three distinct tests designed to capture a large fraction of candidate PAL-1 target genes (see text, and Materials and methods for details). (B) Hierarchical clustergram for 308 candidate genes; yellow indicates high and blue low relative expression. The three time courses were concatenated end to end for hierarchical clustering but are separated here for clarity.

GATA factors, *end-1* and *end-3*, specifically in the E lineage, which activate another GATA factor, *elt-2*, that activates the expression of the gut esterase gene *ges-1* (Fig. 2). *med-1* and *med-2* are expressed at too low abundance to detect quantitatively in wild type (Baugh et al., 2003), but they show an expected increase in *pie-1(zu154)* at 23 minutes (eight-cell stage). *end-1* transcripts are about 1.5-fold more abundant in *pie-1(zu154)* and are not detected following *skn-1* RNAi. Although *end-3*, *elt-2* and *ges-1* are not detected at increased abundance in *pie-1(zu154)*, all three are not detected following *skn-1* RNAi. These observations indicate that *skn-1(RNAi)* appears to phenocopy a null mutant with respect to target gene expression. Furthermore, the times of activation and maximum expression of the *skn-1* target genes are equivalent in the mutants and wild type, suggesting that the transformed lineage is developing at the same molecular rate as its native

sister. However, the fact that only *end-1* shows elevated expression in *pie-1(zu154)* indicates that few target genes will behave as expected and that we must use flexible rules to identify candidate PAL-1 target genes.

To identify candidate PAL-1 target genes, we filtered the data by cluster analysis, a model-directed approach based on analysis of variance (ANOVA) and expression correlation to known targets. In order to maximize identification of true PAL-1 targets, we combined all three approaches, complementing the biases of each. In order to accommodate the resulting high false-positive rate, we assigned a cumulative numerical score from zero to seven for all genes (see Materials and methods) so that strong candidates can be discerned from weak ones. Genes scoring one or higher are considered candidate PAL-1 targets, of which 308 were identified. A score of one indicates that the gene appears to be a target by only one of the three analytical approaches, while a score of seven indicates that a gene appears to be a target by all three approaches. The distribution of scores for candidate target genes is highly skewed, with nearly half of them (146) scoring one and only 17 scoring five or greater (Fig. 3). Hierarchical clustering of expression patterns for the 308 candidate targets shows that expression is generally lower in *pie-1(zu154)* and greater in *mex-3(zu155)*; *skn-1(RNAi)*, and it reveals distinct patterns across time and genotype (Fig. 3), consistent with PAL-1 targets being expressed in diverse spatial patterns and encoding a variety of developmental functions.

Validation of PAL-1 targets

We expect PAL-1 targets to be expressed in complex expression patterns, some of which may correlate with the C-lineage and or muscle or epidermal cell fates. The summary of expectations given in Fig. 1D is for truly lineage-specific genes. Because of secondary and confounded effects of the homeotic transformations used to identify candidate targets genes, true targets may not have received high scores and many false targets likely received low scores. To validate select candidates, we used 'reporter analysis' to determine approximately where each candidate target is expressed and whether that expression is appropriately responsive to a lack [*pal-1(RNAi)*] or excess [*mex-3(RNAi)*] of *pal-1* activity. For
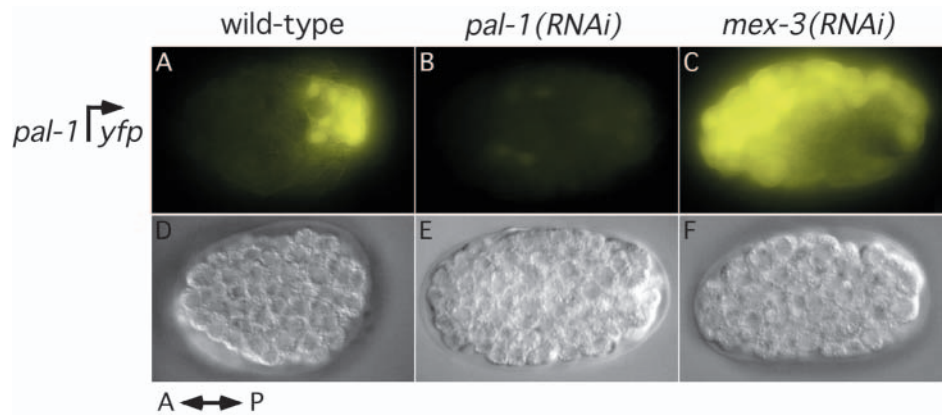
Fig. 4. Target validation. Embryonic expression of *pal-1*, assayed by YFP reporter, is dependent on maternal PAL-1 activity. (A-C) Expression of a zygotic *pal-1* reporter is shown in a mid-gastrula embryo with corresponding Nomarski images for wild-type (A,D), *pal-1(RNAi)* (B,E) and *mex-3(RNAi)* (C,F).

Table 1. Results of target validation

ORF	Gene	Target score	Promoter size (kb)	Wild-type expression pattern	*pal-1* (RNAi)	*mex-3* (RNAi)	Target
C28C12.7	*spp-10*	7	1.7	C lineage	NE	EE	Yes
F01F1.6		7	0.3	No expression	ND	ND	No
R02D3.1		7	0.8	C lineage	NE	EE	Yes
K10B4.6	*cwn-1*	6	1.9	Posterior C lineage	NE	EE	Yes
B0304.1	*hlh-1*	6	3.0	Muscle	NPE	EE	Yes
C09D4.2		5	0.8	C lineage	NE	EE	Yes
C46H11.2		5	2.1	C epidermis	NE	EE	Yes
R07C3.11		5	1.1	Epidermis	LE	EE	Yes
T22B7.3		5	1.8	C epidermis then more epidermis	NE	EE	Yes
ZK1307.1		5	1.1	C epidermis	NE	EE	Yes
C55C2.1		4	6.1	C and D muscle (left side)	NE	EE	Yes
F20D1.4		4	0.4	No expression	NE	NE	No
F45E4.2	*plp-1*	4	1.1	No expression	NE	NE	No
F23H12.4	*sqt-3*	4	2.0	C epidermis then more epidermis	EE	EE	Maybe
ZK829.4		4	0.6	No expression	NE	NE	No
T07C4.6	*tbx-9*	3	3.8	C and AB epidermis, MS	NPE	EE	Yes
C38D4.6	*pal-1*	3	1.4	C and D lineage	NE	EE	Yes
C54F6.8		3	5.0	Epidermis	LE	EE	Yes
D1081.2	*unc-120*	3	5.0	Muscle	NPE	EE	Yes
F35D2.3		3	5.0	Many cells	ND	ND	No
C32F10.5	*hmg-3*	3	5.0	No expression	NE	NE	No
C44C10.8	*hnd-1*	3	1.5	Muscle*	NPE	EE	Yes
Y75B8A.2	*nob-1*	3	4.9	Posterior epidermis and posterior E	LE	EE	Yes
C32E8.8	*ptr-2*	3	3.0	Too dim to score	EE	EE	No
T07C4.2	*tbx-8*	2	3.8	C and AB epidermis, MS	NPE	EE	Yes
ZK270.1	*ptr-23*	2	5.0	C epidermis then all epidermis	EE	ND	No
F35G12.6	*mab-21*	2	4.7	C and AB epidermis	NPE	EE	Yes
W09C2.1	*elt-1*	1	NA	Epidermis†	Unk	Unk	Probably
M142.4	*vab-7*	1	13.8	Posterior C lineage	NE	EE	Yes
T27B1.2		1	5.0	Epidermis	ND	EE	Maybe
F57A10.5	*nhr-60*	1	2.0	Epidermis	ND	EE	Maybe
F11C1.6	*nhr-25*	1	5.0	Epidermis‡	NPE	ND	Yes
F18A1.2	*lin-26*	1	0.8	C epidermis then all epidermis	ND	ND	No
F45E4.9	*hmg-5*	1	0.3	Late epidermis	ND	ND	No
K02B9.4	*elt-3*	1	4.9	Non-seam epidermis	LE	EE	Yes
C13G5.1	*ceh-16*	1	4.1	No expression	ND	ND	No
C30G4.3	*gcy-11*	1	3.2	No expression	NE	NE	No
C25G6.5		1	4.4	Too dim to score	EE	ND	No
W03D2.5	*wrt-5*	1	0.8	No expression	ND	ND	No
Y71F9AL.17		1	0.4	Late epidermis	ND	ND	No

Thirty-nine genes were selected from the list of 308 candidate targets for validation by the assay presented in Fig. 4. Genes were selected either because they have high candidate target scores, they are predicted to be involved in transcription or signaling, or they have known developmental phenotypes.

NE, no expression; NPE, no posterior expression; LE, less expression; EE, ectopic expression; ND, no difference from wild type; Unk, not determined.

Expression in the C and D lineages was not distinguished. Additional, relatively rare expression may not be reported. Results are for early embryo (~4 hours) only (i.e. where 'No expression' is reported).

*Data for *hnd-1* expression were obtained with JK3363 (Mathies et al., 2003).

†We were unable to establish *elt-1::yfp* reporter lines (consistent with B. Page's rescue experiments) and therefore rely on the published expression pattern (Page et al., 1997).

‡*nhr-25* is reported to be expressed in the posterior epidermis but also the endoderm (Asahina et al., 2000); however, we do not see endodermal expression with either the published strain or ours.

example, *pal-1* received a moderate target score of three, and maternal PAL-1 has been suspected to activate the transcription of zygotic *pal-1* (Hunter and Kenyon, 1996). We find that a *pal-1* promoter::YFP fusion is expressed exclusively in the C and D lineage of early wild-type embryos, but is not detected in *pal-1(RNAi)* embryos, and is expressed ectopically in the anterior of *mex-3(RNAi)* embryos (Fig. 4). Because this reporter contains no *pal-1*-coding sequence, RNAi of *pal-1* does not directly effect reporter expression, but only the endogenous gene. This result demonstrates that maternal *pal-1* activates zygotic *pal-1* expression and suggests that zygotic *pal-1* auto-regulates its expression.

We analyzed 39 candidate target genes by reporter analysis in wild type and following *pal-1* and *mex-3* RNAi (Table 1).

Among the 308 unique candidate target genes, all predicted transcription factors or signaling molecules were selected for the reporter assay, as well as a random eight out of 17 high-scoring genes (scored as five or greater). Expression was detected for 31 (79%) of the reporters and 21 (68%) of those were confirmed as targets, with those scoring higher in the cumulative index validating more frequently (Table 2). Temporal expression profiles for each of the validated targets can be found in Fig. S2 in the supplementary material. Failure to be validated does not necessarily mean that the gene is not a target as there are multiple reasons why the reporter assay may result in false negatives. Nevertheless, the validation efficiency suggests that roughly 130 of the 308 candidates would be validated by this reporter assay.

Table 2. Efficiency of target validation

Target score	Number of genes	Number tested by reporter	Expression detected	Validated as target	Validation efficiency
5, 6 or 7	17	10	9	9	90%
2, 3 or 4	145	17	13	9	53%
1	146	12	9	3	25%
All	308	39	31	21	54%

Only genes with a 'yes' in Table 1 are included in the percentage.

Lineage-based regulation of tissue identity genes

With the exception of intestine and germline, cell fates in the *C. elegans* embryo are polyclonal (Sulston et al., 1983). Nevertheless, precursor cells for a given tissue or organ are typically born near each other such that a fate map can be drawn for the mid-gastrula embryo. It has been suggested that this regional positioning of tissue and organ precursors enables their coherent development as tissue and organ primordia (Labouesse and Mango, 1999). However, it is not known what mechanisms direct the expression of tissue and organ identity genes in the appropriate cells. It is possible that the position of a cell in the embryo could influence the decision to express a given tissue identity gene (Schnabel, 1996). Alternatively, different cell-autonomous mechanisms could direct the expression of the tissue identity gene in different lineages.

The muscle-specific genes *hnd-1*, *hlh-1* and *unc-120* are expressed in muscle precursors in the MS, C and D lineages (Krause et al., 1990; Mathies et al., 2003) (Fig. 5, Fig. 8). Although the expressing cells are derived from three lineages, they are born in the ventral-posterior region of the embryo and within two cell divisions form four longitudinal stripes pre-ordaining the four quadrants of body wall muscle present at hatching (Fig. 5, Fig. 8). To determine if the decision to express these tissue identity genes is controlled in a cell-autonomous fashion, we examined reporter gene expression more carefully following RNAi of *pal-1*. Consistent with a lineage-based mechanism, *pal-1* function is required for expression of *hnd-1*, *hlh-1* and *unc-120* reporter genes in the C and D lineages but not MS (Fig. 5), where we presume an unknown lineage-specific factor is required for their expression.

Regulation of a posterior HOX gene in lineage-specific fashion

pal-1, also known as *nob-2*, was isolated from the same forward genetic screen as the Abd-b/Hox9-13 ortholog *nob-1* (Van Auken et al., 2000). The Nob (no back end) phenotype suggests that the two genes may function in a common pathway

to pattern posterior development. Consistent with this hypothesis, a transcriptional reporter for *nob-1* is expressed in the extreme posterior of the embryo in cells derived from the AB, E and C lineages (see Figs 7 and 9). This expression pattern suggests regional, as opposed to lineage-based, regulation of *nob-1* because regional regulation need only require a single mechanism, while lineage-based regulation would require *nob-1* expression to be activated independently in the most posterior descendants of each lineage. We find that expression of a *nob-1* reporter within the C lineage requires *pal-1* function and that *mex-3(RNAi)* results in ectopic anterior expression of *nob-1*::YFP in a relatively small number of cells consistent with *nob-1* being expressed in one-eighth of the C lineage (Fig. 6). These results indicate that *nob-1* expression in the C lineage requires lineage-specific *pal-1*-dependent information, although we cannot rule out the possibility that *nob-1* expression requires *pal-1*-dependent signaling activity. However, given that in *pal-1(RNAi)* embryos *nob-1*::YFP is expressed in the AB and E lineages and that in *mex-3(RNAi)* embryos *nob-1*::YFP is expressed in sparse pairs of anterior cells, it is unlikely that *nob-1* expression is activated by a *pal-1*-dependent signaling activity.

PAL-1 targets affect specification, differentiation and morphogenesis of C-lineage cells

Candidate PAL-1 targets selected for validation by reporter gene analysis were biased toward potential developmental regulators (Table 3). Twelve transcription factors representing many families, including homeodomain, zinc-finger, GATA, MADS domain, bHLH and T-box are included, consistent with PAL-1 targets controlling diverse aspects of development. That so many previously characterized transcription factors are among the targets of PAL-1 suggests that *pal-1* controls a transcriptional regulatory network.

To begin to dissect the function and regulatory relationships among the validated PAL-1 targets, we measured embryonic lethality following RNAi by soaking. Given the pleiotropy and

Fig. 5. Reporters for muscle-specific transcription factors are regulated in a lineage-specific fashion. Expression of *hnd-1*::GFP::*lacZ* (A,D), *hlh-1*::YFP (B,E) and *unc-120*::YFP (C,F) in wild-type (A-C) and *pal-1*(RNAi) (D-F) embryos. *pal-1* RNAi eliminates expression in the C and D lineages (posterior). *hnd-1*::*gfp*::*lacZ* embryos were imaged 210 minutes after the four-cell stage and *hlh-1*::*yfp* and *unc-120*::*yfp* at 260 minutes (22°C). Because of the larger size of the reporter protein the signal for *hnd-1*::*gfp*::*lacZ* is better localized to the nucleus than in the other two strains.
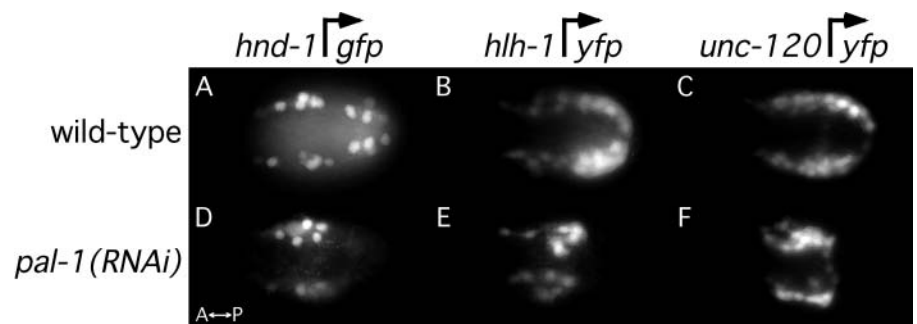
Fig. 6. Expression of the posterior HOX gene *nob-1* is regulated by cell lineage. (A) Wild-type expression of a *nob-1* YFP reporter in a ~200-cell embryo (210 minutes after the four-cell stage at 22°C). Arrows indicate the two posterior-most C-lineage cells. Expression of the *nob-1* reporter in similarly staged *pal-1(RNAi)* (B) and *mex-3(RNAi)* (C) embryos shows loss of and ectopic expression, respectively.
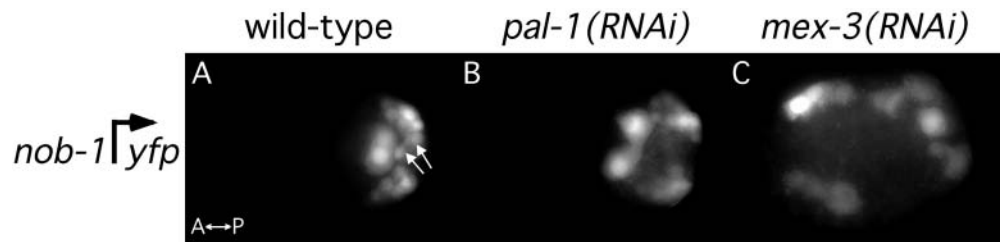
Table 3. Phenotypic analysis of validated target genes

ORF	Gene	Identification	Published RNAi*	Embryonic lethality (n)	Embryonic function	Larval phenotype
C28C12.7	*spp-10*	Predicted prosaposin	Wild type	1% (787)		
R02D3.1		Dehydrogenase	Wild type	0% (731)		
C09D4.2		Uncharacterized	Wild type	2% (657)		
C38D4.6	*pal-1*	Homeobox transcription factor (cad subfamily)	Emb	100% (1038)	Specification and morphogenesis of C lineage[†,‡]	Nob, Vab[‡]
M142.4	*vab-7*	Homeobox TF (even-skipped subfamily)	Bmd	2% (512)	Morphogenesis of C lineage[§]	Nob, Vab[§]
C55C2.1		Zinc-finger transcription factor	Wild type	2% (1032)		
B0304.1	*hlh-1*	bHLH transcription factor	Wild type	9% (716)	Differentiation of muscle[¶]	Unc[¶]
D1081.2	*unc-120*	MADS domain transcription factor	Unc	3% (680)	Differentiation of muscle**	Unc**
C44C10.8	*hnd-1*	Hand bHLH transcription factor	Bmd	1% (516)		Posterior bulges***
K10B4.6	*cwn-1*	Putative wingless ligand	NA	2% (507)		Rare tail defects[†††]
R07C3.11		Uncharacterized	Wild type	1% (699)		
C54F6.8		Nuclear hormone receptor	Wild type	2% (875)		
F11C1.6	*nhr-25*	Nuclear hormone receptor	Wild type	1% (718)		Rare tail defects[‡‡]
W09C2.1	*elt-1*	GATA transcription factor	Emb, Unc	100% (487)	Specification of epidermis[††]	
K02B9.4	*elt-3*	GATA transcription factor	Wild type	1% (833)		
F35G12.6	*mab-21*	Highly conserved novel protein	Wild type	2% (711)		Sqt, male tale defects[§§]
T22B7.3		Uncharacterized	Wild type	1% (657)		
C46H11.2		Uncharacterized	Wild type	1% (791)		
ZK1307.1		Uncharacterized	Wild type	1% (698)		
T07C4.2	*tbx-8*	T-box transcription factor (Brachyury)	Wild type	3% (502)		Vab, posterior bulges, tail defects[†††]
T07C4.6	*tbx-9*	T-box transcription factor (Brachyury)	Wild type	1% (520)		Vab, posterior bulges, tail defects[†††]
Y75B8A.2	*nob-1*	Homeodomain transcription factor (posterior Hox paralog)	Emb, Bmd	36% (761)	Morphogenesis of posterior epidermis[¶¶]	Nob, Vab[¶¶]

Protein identification, published phenotypes from a genome-wide RNAi screen, embryonic lethality following RNAi, and known embryonic functions and larval phenotypes are summarized.

Emb, embryonic lethal; Bmd, body morphology defective; Unc, uncoordinated; Nob, no back end; Vab, variably abnormal morphogenesis; Sqt, squat.

*Kamath and Ahringer, 2003; [†]Hunter and Kenyon, 1996; [‡]Edgar et al., 2001; [§]Ahringer, 1996; [¶]Chen et al., 1994; **Williams and Waterston, 1994; [††]Page et al., 1997; [‡‡]Asahina et al., 2000; [§§]Chow et al., 1995; [¶¶]Van Auken et al., 2000; ***Mathies et al., 2003; [†††]This work (data not shown).

Published RNAi results are only shown for Emb, Bmd and Unc only (wild type only relates to those three phenotypes).

severity of *pal-1* loss-of-function phenotypes, we were surprised that RNAi of only two of the 22 genes resulted in 100% embryonic lethality (*pal-1* and *elt-1*), while another two resulted in significant lethality (*hlh-1* and *nob-1*) (Table 3). The embryonic function of these four genes has been described in the literature, and they affect specification, differentiation or morphogenesis of C-lineage cells. Zygotic *pal-1* function is required for proper morphogenesis of the C lineage (Edgar et al., 2001), *elt-1* is required for specification of the epidermis (Page et al., 1997), *hlh-1* is required for differentiation of the muscle (Chen et al., 1994) and *nob-1* is required for morphogenesis of the posterior epidermis (Van Auken et al., 2000). RNAi of several of the other genes produced post-embryonic phenotypes indicative of a function in posterior morphogenesis. For example, tail defects in larvae lacking *vab-7*, *pal-1*, *nob-1*, *tbx-8* and *tbx-9* function have been shown to

result from improper morphogenesis of C-lineage cells (Ahringer, 1996; Edgar et al., 2001; Pocock et al., 2004; Van Auken et al., 2000), and the rare larval defects seen following loss of function of *hnd-1*, *nhr-25* and *cwn-1* are also likely to result in part from improper morphogenesis of C-lineage cells (Table 3). Although additional subtle phenotypes may be detected in more sensitive phenotypic assays, the limited number of penetrant phenotypes following RNAi may be attributable to ineffective depletion by RNAi, functional redundancy of individual genes or network level compensatory mechanisms.

Predicting regulatory network structure from expression

To model the regulatory network specified by PAL-1, we focused on the expression patterns of validated targets most
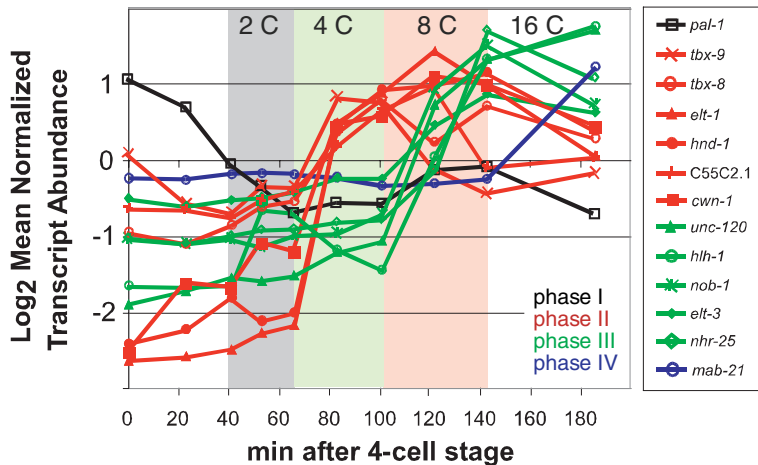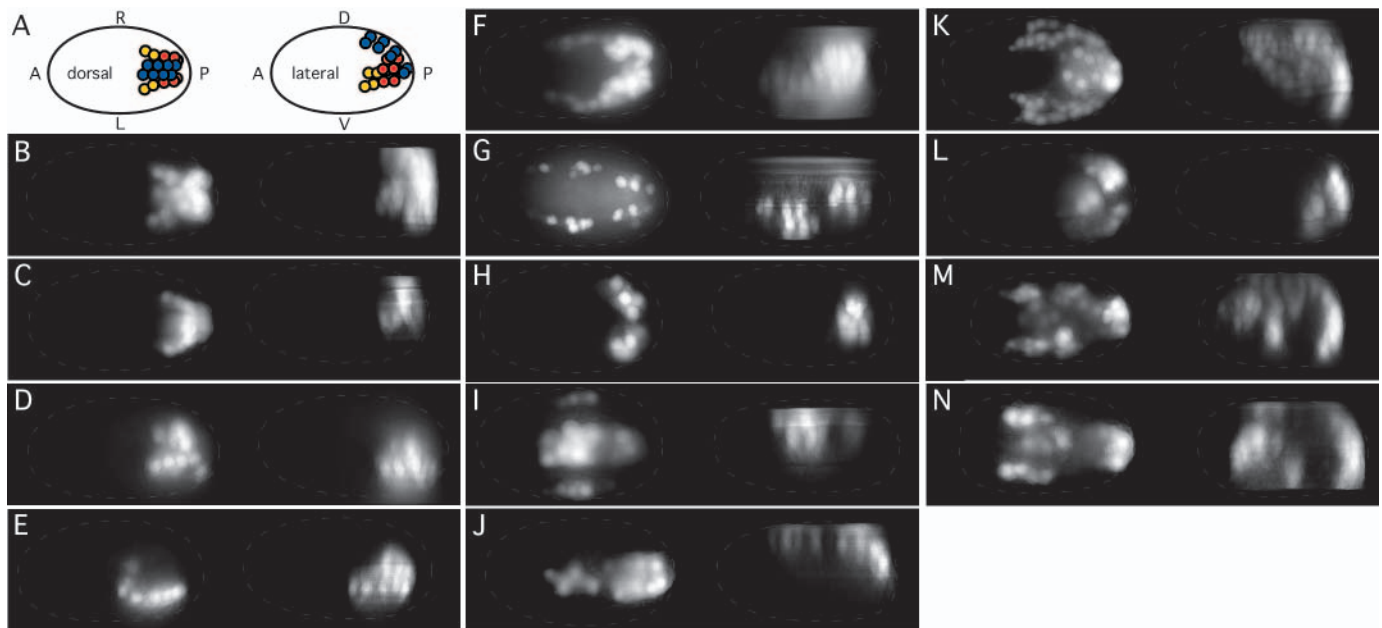
Fig. 7. PAL-1 target genes likely involved in patterning are expressed in four temporal phases. Wild-type temporal expression patterns are plotted for 13 validated target genes that encode 11 transcription factors, a wingless ligand (*cwn-1*) and a novel protein (*mab-21*) known to be involved in a cell fate switch in the male tail. Developmental stages are indicated as the number of C-lineage cells across the top of the graph, and the phase each gene belongs to is indicated by color (based on when zygotic expression is first detected). *vab-7* is omitted from the graph because it is not detected in wild type by microarray, but we know it is a phase II gene based on *lacZ* reporter and in situ hybridization (Ahringer, 1996), as well as its detection in the *mex-3(zu155)*; *skn-1(RNAi)* timecourse data.

decisions in the male tail (*mab-21*). Global analysis of transcript abundance in wild-type embryos has shown that most temporally modulated embryonic genes are expressed transiently, with abundance increasing for one cell cycle and decreasing for one after that (Baugh et al., 2003), consistent with the timing of gene expression during specification of endodermal cell fate (Maduro and Rothman, 2002). The first three divisions of the C blastomere are asymmetric in either cell fate or gene expression (Ahringer, 1996; Sulston et al., 1983), suggesting that the PAL-1 network may also operate on a one cell cycle time scale. Consistent with this expectation, transcription of these target genes is activated in one of four temporal phases, each a cell cycle apart (Fig. 7). Twelve out of the 14 selected targets are expressed in two waves corresponding to the 4C-cell stage (phase II) and the 8C-cell stage (phase III). *mab-21* is the only phase IV gene among the 14, and *pal-1* is considered phase I because its zygotic transcripts are first detected by in situ hybridization at the 2C-cell stage in Ca and Cp (Hunter and Kenyon, 1996).

To generate a working model of the regulatory network, we began with the simple notion that targets from one phase regulate targets of the next, with transcription factors regulating targets expressed in the same cells and signaling molecules regulating targets in adjacent cells from where each is expressed. As YFP expression perdures, we compared spatial expression patterns of each target at the end of phase IV, capturing the spatial expression patterns initiated earlier. Fig. 8 shows dorsal and lateral views of volume rendered images

likely to directly effect gene expression. This includes 12 transcription factors, an uncharacterized wingless ligand (*cwn-1*) and a conserved novel protein known to affect cell fate

Fig. 8. Projection of three-dimensional images of reporter expression patterns for PAL-1 target genes likely to be involved in patterning. Dorsal and lateral views of chromosomally integrated transcriptional reporters 210 minutes after the four-cell stage (22°C; ~200 cells, 16 C-lineage cells) are shown for 13 validated targets. Images are orthogonal perspectives following three-dimensional volume rendering of many optical sections. A broken gray line marks the approximate outline of each egg case. (A) Ordination: 8 C ectodermal nuclei in blue, 8 C muscle nuclei in red and 4 D muscle nuclei in yellow; (B) *pal-1*; (C) *vab-7*; (D) *cwn-1*; (E) C55C2.1; (F) *unc-120*; (G) *hnd-1*; (H) *hlh-1*; (I) *elt-3*; (J) *mab-21*; (K) *nhr-25*; (L) *nob-1*; (M) *tbx-8*; (N) *tbx-9*. Movies for each reporter are available in the supplementary material.

Table 4. A simple model for predicting regulatory relationships from temporal and spatial expression patterns

Gene	Temporal phase	Spatial expression pattern	Supporting data	Refs	Best candidate upstream activator
pal-1	I	16 C, 4 D	Reporter, antibody, in situ hybridization	Edgar et al., 2001; Hunter and Kenyon, 1996	*pal-1*
cwn-1	II	C and D muscle, 2 posterior C ectoderm			*pal-1*
C55C2.1	II	Left posterior muscle (C and D), unidentified ventral cells			*pal-1*
hnd-1	II	All muscle (including 8 C muscle, 4 D muscle)	Reporter	Mathies et al., 2003	*pal-1*
tbx-8	II	Dorsal epidermis, unidentified lateral cells (including 8 C ectoderm)	Reporter	Pocock et al., 2004; Andachi, 2004	*pal-1*
tbx-9	II	Dorsal epidermis, unidentified lateral cells (including 8 C ectoderm)	Reporter, in situ hybridization	Pocock et al., 2004; Andachi, 2004	*pal-1*
elt-1	II	All major epidermis	Antibody	Page et al., 1997	*pal-1*
vab-7	III	8 C: four posterior muscle, four posterior ectoderm	Reporter, in situ hybridization	Ahringer 1996	*cwn-1*
unc-120	III	All muscle (including 8 C muscle, 4 D muscle)	Reporter	Waterston (unpublished)	*hnd-1*
hlh-1	III	All muscle (including 8 C muscle, 4 D muscle)*	Reporter, antibody, in situ hybridization	Krause et al., 1990; Seydoux and Fire, 1994	*hnd-1*
elt-3	III	Non-seam major epidermis (including 8 C ectoderm)	Reporter	Gilleard et al., 1999	*elt-1*
nhr-25	III	Dorsal epidermis (including 8 C ectoderm)	Reporter	Asahina et al., 2000	*elt-1*
nob-1	III	Posterior endoderm (E) and ectoderm – mostly non-C	In situ hybridization	Kohara (unpublished)	*cwn-1*
mab-21	IV	Dorsal-most epidermis (including 8 C ectoderm)	Reporter	Ho et al., 2001	*elt-3*

Temporal phase and spatial expression pattern are summarized for 14 target genes likely to be involved in patterning. The best candidate upstream activator for each gene is the gene from the previous temporal phase with the most similar spatial expression pattern. References with supporting data measured by a variety of techniques are included where available. All reporter expression patterns are consistent with published data. *Expression is seen only in posterior muscle (C and D), where it is initiated (see Fig. 7).

of transcriptional reporters (see also Movies 1-13 in the supplementary material). The expression patterns are summarized in Table 4, and are consistent with published descriptions. *hnd-1*, *hlh-1* and *unc-120* are expressed in all of the embryonic myoblasts, but *hnd-1* is phase II and *hlh-1* and *unc-120* are phase III, suggesting that *hnd-1* is a positive regulator of *hlh-1* and *unc-120*. We have extended this logic, combining spatial expression with quantitative temporal data to identify the best candidate upstream activator for each of the 14 target genes (Table 4). This set of predictions provides a first draft model of the transcriptional network specified by PAL-1 in the C lineage (Fig. 9). Expression of all fourteen targets is by definition dependent on *pal-1* function, and we

presume that both maternal and zygotic PAL-1 act directly on phase II targets, but we do not know if PAL-1 acts directly on phase III or IV targets. The model predicts which phase II targets activate which phase III targets, as well as the activation of the phase IV target *mab-21* by the phase III target *elt-3*, and it can easily be extended for other validated targets.

Discussion

We have shown that lineage-specific patterns of gene expression can be deconvolved from microarray data collected from whole *C. elegans* embryos by analyzing mutants with transformed lineage identity (Figs 2, 3). This approach can be extended with available mutants to identify genes specifically expressed in each of the somatic founder lineages of the embryo. Although reporter genes will eventually be made for the entire genome, scoring all of their embryonic expression patterns will be difficult. By contrast, microarrays can be used to predict the early embryonic expression pattern for each gene in the genome and reporters can be used for validation and refinement.

The identification of PAL-1 targets confirms previous phenotypic analysis (Hunter and Kenyon, 1996) by demonstrating that it controls development of the C lineage, including specification of multiple cell types. PAL-1 activates expression of target genes responsible for various developmental functions such that their disruption phenocopies different aspects of the *pal-1* mutant phenotype (Table 3). Although many
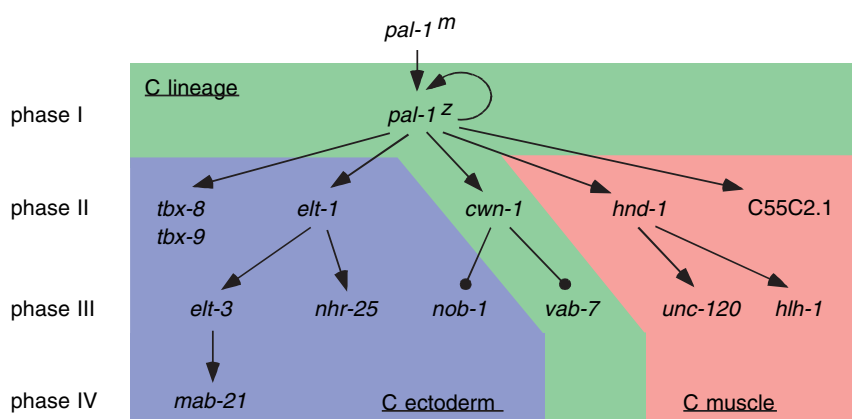
Fig. 9. Proposed structure of the regulatory network specified by *pal-1*. A graph of predicted regulatory relationships based on the best candidate upstream activator for each gene (Table 4) is presented. Temporal phase is indicated on the left. Lines with arrows represent cell-autonomous regulation by transcription factors and lines with dots represent regulation by signaling molecules.

of the target genes function in the descendants of additional founder blastomeres, their expression in the C lineage depends on PAL-1 (Table 1). We specifically show three muscle-specific genes to require *pal-1* function for C-lineage expression (Fig. 5) and find the same to be true for epidermis-specific expression (Table 1). If specification of cell fate were controlled by regionalizing influences as opposed to lineage-based mechanisms, we would not expect expression of fate-specific genes to depend on *pal-1* function (unless *pal-1* were required to respond to regionalizing influences). Furthermore, expression of the posterior HOX gene *nob-1* is complex with respect to lineage but simple with respect to region, yet we show that *nob-1*::YFP expression in the C lineage requires PAL-1.

We propose a model for the structure of the network specified by PAL-1 based only on temporal and spatial expression patterns in wild type (Fig. 9). The rationale behind the model is simple: genes activated in one cell cycle affect the expression of genes expressed in the next cell cycle. This premise is supported by both functional analysis of the endodermal network (Maduro and Rothman, 2002) and global analysis of expression dynamics (Baugh et al., 2003). Furthermore, whereas zygotic *pal-1* transcripts are first detected at the 2C-cell stage in Ca and Cp (phase I), loss of zygotic *pal-1* function results in a detectable mutant phenotype in their daughters at the 4C-cell stage (Edgar et al., 2001; Hunter and Kenyon, 1996). In addition, protein for the phase II gene *elt-1* is first detected at the end of the 4C-cell stage, and transcription of its confirmed phase III target *elt-3* begins in the 8C-cell stage (Fig. 7) (Gilleard and McGhee, 2001; Page et al., 1997).

That ectopic PAL-1 activity in early blastomeres is sufficient to cause complete transformation of one lineage into another indicates that the regulatory network specified by PAL-1 is modular or self-contained (Draper et al., 1996; Hunter and Kenyon, 1996). After maternal PAL-1 specifies the C lineage, embryonically expressed PAL-1 is required for C-lineage development (Edgar et al., 2001). We therefore hypothesize that PAL-1 continuously regulates target genes during patterning of the C lineage, as opposed to simply initiating a transcriptional cascade. Although it is not known how far into development PAL-1 function is required, phenotypically mutant *pal-1* mosaic animals were recovered corresponding to loss of *pal-1* in one Cxx cell at the 4C-cell stage and PAL-1 expression is detectable in the C lineage until the 16C-cell stage (Edgar et al., 2001), leaving open the possibility that PAL-1 directly activates each of the target genes identified here. Combinatorial control of gene expression, where early targets regulate late targets in combination with PAL-1, offers one possible mechanism for the timing of gene expression within this modular network (Mangan and Alon, 2003; Penn et al., 2004).

There must be additional regulation not predicted by our model. Genes that are not PAL-1 targets are likely to participate in transcriptional regulation and patterning of the C lineage. For example, the Homothorax ortholog *unc-62* and the Extradenticle homologs *ceh-20* and *ceh-40* have superficially similar phenotypes to *nob-1* and *pal-1*, suggesting that these co-factor homeodomain proteins interact with and modify the function of PAL-1 and NOB-1 (Van Auken et al., 2002). Likewise, the Tcf/Lef factor *pop-1* is thought to mediate cell-

fate decisions associated with every cell division on the AP axis of the early embryo (Lin et al., 1998), and, as has been shown for development of the E lineage (Calvo et al., 2001; Maduro et al., 2002), we expect POP-1 to contribute to patterning of PAL-1 target expression, in particular where targets are expressed in only the anterior or posterior daughters following a round of C cell divisions (e.g. *hlh-1*, *elt-1* and *vab-7*). Repression is completely ignored in the current model, but is probably crucial for patterning, as indicated by the fact that very few targets are expressed in all PAL-1-expressing cells. So there may also be genes repressed by PAL-1. We have not allowed for genes of the same temporal phase to regulate each other, though it is likely that there is mutual repression between genes specifying muscle and epidermis, leading to insulation of the two states. In addition, genes of the same temporal phase expressed in the same cells may activate the expression of one another, and we imagine multiple auto-regulatory positive feedbacks in addition to the one demonstrated for *pal-1*. It will be interesting to compare the structures of different developmental regulatory networks in an effort to understand better how different topological motifs contribute to the functional properties of the regulatory network (Milo et al., 2002; Shen-Orr et al., 2002) and ultimately how network structure relates to body plan.

Supplementary material

Supplementary material for this article is available at
http://dev.biologists.org/cgi/content/full/132/8/1843/DC1

References

Ahringer, J. (1996). Posterior patterning by the Caenorhabditis elegans even-skipped homolog vab-7. *Genes Dev.* **10**, 1120-1130.

Andachi, Y. (2004). *Caenorhabditis elegans* T-box genes tbx-9 and tbx-8 are required for formation of hypodermis and body-wall muscle in embryogenesis. *Genes Cells* **9**, 331-344.

Asahina, M., Ishihara, T., Jindra, M., Kohara, Y., Katsura, I. and Hirose, S. (2000). The conserved nuclear receptor Ftz-F1 is required for embryogenesis, moulting and reproduction in Caenorhabditis elegans. *Genes Cells* **5**, 711-723.

Baugh, L. R., Hill, A. A., Brown, E. L. and Hunter, C. P. (2001). Quantitative analysis of mRNA amplification by in vitro transcription. *Nucleic Acids Res.* **29**, E29.

Baugh, L. R., Hill, A. A., Slonim, D. K., Brown, E. L. and Hunter, C. P. (2003). Composition and dynamics of the Caenorhabditis elegans early embryonic transcriptome. *Development* **130**, 889-900.

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.

Calvo, D., Victor, M., Gay, F., Sui, G., Luke, M. P., Dufourcq, P., Wen, G., Maduro, M., Rothman, J. and Shi, Y. (2001). A POP-1 repressor complex restricts inappropriate cell type-specific gene transcription during Caenorhabditis elegans embryogenesis. *EMBO J.* **20**, 7197-7208.

Chen, L., Krause, M., Sepanski, M. and Fire, A. (1994). The Caenorhabditis elegans MYOD homologue HLH-1 is essential for proper muscle function and complete morphogenesis. *Development* **120**, 1631-1641.

Chow, K. L., Hall, D. H. and Emmons, S. W. (1995). The mab-21 gene of Caenorhabditis elegans encodes a novel protein required for choice of alternate cell fates. *Development* **121**, 3615-3626.

Draper, B. W., Mello, C. C., Bowerman, B., Hardin, J. and Priess, J. R.

(1996). MEX-3 is a KH domain protein that regulates blastomere identity in early C. elegans embryos. *Cell* **87**, 205-216.

Edgar, L. G., Carr, S., Wang, H. and Wood, W. B. (2001). Zygotic expression of the caudal homolog pal-1 is required for posterior patterning in Caenorhabditis elegans embryogenesis. *Dev. Biol.* **229**, 71-88.

Fire, A., Harrison, S. W. and Dixon, D. (1990). A modular set of lacZ fusion vectors for studying gene expression in Caenorhabditis elegans. *Gene* **93**, 189-198.

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

Gilleard, J. S. and McGhee, J. D. (2001). Activation of hypodermal differentiation in the *Caenorhabditis elegans* embryo by GATA transcription factors ELT-1 and ELT-3. *Mol. Cell. Biol.* **21**, 2533-2544.

Gilleard, J. S., Shafi, Y., Barry, J. D. and McGhee, J. D. (1999). ELT-3: a *Caenorhabditis elegans* GATA factor expressed in the embryonic epidermis during morphogenesis. *Dev. Biol.* **208**, 265-280.

Heyer, L. J., Kruglyak, S. and Yooseph, S. (1999). Exploring expression data: identification and analysis of coexpressed genes. *Genome Res* **9**, 1106-1115.

Ho, S. H., So, G. M. and Chow, K. L. (2001). Postembryonic expression of *Caenorhabditis elegans mab-21* and its requirement in sensory ray differentiation. *Dev. Dyn.* **221**, 422-430.

Hobert, O. (2002). PCR fusion-based approach to create reporter gene constructs for expression analysis in transgenic C. elegans. *Biotechniques* **32**, 728-730.

Huang, N. N., Mootz, D. E., Walhout, A. J., Vidal, M. and Hunter, C. P. (2002). MEX-3 interacting proteins link cell polarity to asymmetric gene expression in Caenorhabditis elegans. *Development* **129**, 747-759.

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.

Inoue, T., Sherwood, D. R., Aspock, G., Butler, J. A., Gupta, B. P., Kirouac, M., Wang, M., Lee, P. Y., Kramer, J. M., Hope, I. et al. (2002). Gene expression markers for Caenorhabditis elegans vulval cells. *Mech. Dev.* **119**, S203-S209.

Irizarry, R. A., Hobbs, B., Collin, F., Beazer-Barclay, Y. D., Antonellis, K. J., Scherf, U. and Speed, T. P. (2003). Exploration, normalization, and summaries of high density oligonucleotide array probe level data. *Biostatistics* **4**, 249-264.

Kaletta, T., Schnabel, H. and Schnabel, R. (1997). Binary specification of the embryonic lineage in Caenorhabditis elegans. *Nature* **390**, 294-298.

Kamath, R. S. and Ahringer, J. (2003). Genome-wide RNAi screening in Caenorhabditis elegans. *Methods* **30**, 313-321.

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.

Labouesse, M. and Mango, S. E. (1999). Patterning the C. elegans embryo: moving beyond the cell lineage. *Trends Genet.* **15**, 307-313.

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.

Maduro, M. and Pilgrim, D. (1995). Identification and cloning of unc-119, a gene expressed in the Caenorhabditis elegans nervous system. *Genetics* **141**, 977-988.

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., 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-3β homolog is mediated by MED-1 and -2 in *C. elegans*. *Mol. Cell* **7**, 475-485.

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.

Maeda, I., Kohara, Y., Yamamoto, M. and Sugimoto A. (2001). Large-scale analysis of gene function in *Caenorhabditis elegans* by high-throughput RNAi. *Curr. Biol.* **11**, 171-176.

Mangan, S. and Alon, U. (2003). Structure and function of the feed-forward loop network motif. *Proc. Natl. Acad. Sci. USA* **100**, 11980-11985.

Mathies, L. D., Henderson, S. T. and Kimble, J. (2003). The C. elegans Hand gene controls embryogenesis and early gonadogenesis. *Development* **130**, 2881-2892.

Mello, C. C., Kramer, J. M., Stinchcomb, D. and Ambros, V. (1991). Efficient gene transfer in C.elegans: extrachromosomal maintenance and integration of transforming sequences. *EMBO J.* **10**, 3959-3970.

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.

Mello, C. C., Schubert, C., Draper, B., Zhang, W., Lobel, R. and Priess, J. R. (1996). The PIE-1 protein and germline specification in C. elegans embryos. *Nature* **382**, 710-712.

Milo, R., Shen-Orr, S., Itzkovitz, S., Kashtan, N., Chklovskii, D. and Alon, U. (2002). Network motifs: simple building blocks of complex networks. *Science* **298**, 824-827.

Page, B. D., Zhang, W., Steward, K., Blumenthal, T. and Priess, J. R. (1997). ELT-1, a GATA-like transcription factor, is required for epidermal cell fates in Caenorhabditis elegans embryos. *Genes Dev.* **11**, 1651-1661.

Penn, B. H., Bergstrom, D. A., Dilworth, F. J., Bengal, E. and Tapscott, S. J. (2004). A MyoD-generated feed-forward circuit temporally patterns gene expression during skeletal muscle differentiation. *Genes Dev.* **18**, 2348-2353.

Pocock, R., Ahringer, J., Mitsch, M., Maxwell, S. and Woollard, A. (2004). A regulatory network of T-box genes and the even-skipped homologue vab-7 controls patterning and morphogenesis in C. elegans. *Development* **131**, 2373-2385.

Schnabel, R. (1996). Pattern formation: regional specification in the early C. elegans embryo. *BioEssays* **18**, 591-594.

Seydoux, G. and Fire, A. (1994). Soma-germline asymmetry in the distributions of embryonic RNAs in *Caenorhabditis elegans*. *Development* **120**, 2823-2834.

Seydoux, G., Mello, C. C., Pettitt, J., Wood, W. B., Priess, J. R. and Fire, A. (1996). Repression of gene expression in the embryonic germ lineage of C. elegans. *Nature* **382**, 713-716.

Shen-Orr, S. S., Milo, R., Mangan, S. and Alon, U. (2002). Network motifs in the transcriptional regulation network of Escherichia coli. *Nat. Genet.* **31**, 64-68.

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.

Van Auken, K., Weaver, D. C., Edgar, L. G. and Wood, W. B. (2000). Caenorhabditis elegans embryonic axial patterning requires two recently discovered posterior-group Hox genes. *Proc. Natl. Acad. Sci. USA* **97**, 4499-4503.

Van Auken, K., Weaver, D., Robertson, B., Sundaram, M., Saldi, T., Edgar, L., Elling, U., Lee, M., Boese, Q. and Wood, W. B. (2002). Roles of the Homothorax/Meis/Prep homolog UNC-62 and the Exd/Pbx homologs CEH-20 and CEH-40 in C. elegans embryogenesis. *Development* **129**, 5255-5268.

Williams, B. D. and Waterston, R. H. (1994). Genes critical for muscle development and function in Caenorhabditis elegans identified through lethal mutations. *J. Cell Biol.* **124**, 475-490.

Winston, W. M., Molodowitch, C. and Hunter, C. P. (2002). Systemic RNAi in C. elegans requires the putative transmembrane protein SID-1. *Science* **295**, 2456-2459.

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.