Transcriptional regulation*

Peter G. Okkema§, Laboratory of Molecular Biology, University of Illinois at Chicago, Chicago, IL 60607, USA

Michael Krause§, Laboratory of Molecular Biology/NIDDK, Bethesda, MD 20892-0510, USA

Table of Contents

1. Introduction .......................................................................................................................... 2
2. Tools to study transcriptional regulation ............................................................................ 2
3. Locating cis-acting regulatory elements ........................................................................... 4
4. Simple promoters .............................................................................................................. 4
5. Complex promoters ........................................................................................................... 5
   5.1. myo-2: activation of a terminal differentiation gene by the combined activities of organ- and cell type-specific regulatory elements .............................................................. 5
   5.2. hlh-1: activation of gene expression by lineage-preference regulatory elements. .......... 6
   5.3. lin-26: activation of gene expression by tissue-specific regulatory elements. ............. 7
6. Trans-acting factors .......................................................................................................... 8
7. Spatial specificity ............................................................................................................... 29
   7.1. Organ specificity: control of pharyngeal gene expression by a master regulator .......... 29
   7.2. Tissue specificity: regulation of gut gene expression by a cascade of redundant GATA factors. ................................................................. 30
   7.3. Cell type specificity: regulation of A1Y neuronal expression by a single core motif .......... 31
8. Future .................................................................................................................................. 31
9. Acknowledgements ......................................................................................................... 31
10. References ......................................................................................................................... 31

Abstract

The regulation of transcription in C. elegans shares many similarities to transcription in other organisms. The details of how specific transcription factors bind to target promoters and act as either activators or repressors are still being examined in many cases, but an increasing number of factors and their binding sites are being characterized. This chapter reviews the general concepts that have emerged with regards to promoter function in C. elegans. Included are the methods that have been successfully employed as well as limitations encountered to date. Specific cis-acting promoter elements from myo-2, hlh-1 and lin-26

---

*Edited by Thomas Blumenthal. Last revised April 12, 2005. Published December 23, 2005. This chapter should be cited as: Okkema, P. G. and Krause, M. Transcriptional regulation (December 23, 2005), WormBook, ed. The C. elegans Research Community, WormBook, doi/10.1895/wormbook.1.45.1, http://www.wormbook.org.

Copyright: © 2005 Peter G. Okkema and Michael Krause, This is an open-access article distributed under the terms of the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

§To whom correspondence should be addressed. E-mail: okkema@uic.edu or mwkrause@helix.nih.gov
are discussed as examples of complex promoters regulated by multiple sequence elements. In addition, examples of organ-, tissue-, and cell type-specific mechanisms for generating spatial specificity in gene expression are discussed.

1. Introduction

Regulation of Polymerase II (Pol II) transcription in C. elegans can be described as typical for eukaryotes. Pol II appears to act in concert with TATA Binding Protein (TBP) and TBP-Associated Factors (TAFs) at the promoter of protein coding genes (Dantonel et al., 2000; Kaltenbach et al., 2000; Lichtsteiner and Tjian, 1993; Walker et al., 2004). Active Pol II is phosphorylated on the C-terminal domain (CTD) at serine 2 and 5 like other eukaryotes (Seydoux and Dunn, 1997; Wallenfang and Seydoux, 2002; Zhang et al., 2003). The functions of these proteins at the core of transcription are beginning to be defined and are reviewed in Transcription mechanisms. However, there are many things about transcription in C. elegans that we do not yet know for sure. For example, putative TATA and CAT boxes upstream of the coding region are often described, but this is largely done subjectively without any firm experimental evidence for the function of these elements. We also have not fully explored the role of histone modifications and chromatin organization in somatic cell transcription. Progress on these fronts has primarily been made in the areas of dosage compensation (see X-chromosome dosage compensation) and germline chromatin organization (see Germline chromatin). For these cases, the evolutionary conservation suggests that somatic cell transcription will similarly be influenced by typical eucaryotic mechanisms of chromatin organization. In many ways then, our understanding of transcription in C. elegans is still in its infancy, reflecting the fact that C. elegans, as a model biological system, is still a growing field that has primarily been exploited for its genetics.

The purpose of this chapter is to provide an overview of transcriptional regulation in C. elegans. It is geared towards an audience that is naive in the ways of C. elegans gene regulation, but, it also includes information that should be helpful even to the seasoned veteran. The tools for studying transcription in C. elegans will be described in an effort to illustrate successful approaches and highlight techniques that, while useful in other systems, are challenging in the nematode. A review of the general trends in regulatory elements is followed by specific examples of spatial and temporal regulatory strategies. The hope is that this information will serve as both a useful review and an entry point into literature appropriate for specific applications.

2. Tools to study transcriptional regulation

Reporter genes are the most commonly used method to study transcriptional regulation in C. elegans. It is straightforward to generate transgenic lines (see Transformation and microinjection), and, as C. elegans is transparent throughout its life, it is easy to visualize reporter gene expression in all cells. Early studies of gene expression relied on lacZ reporter genes and were aided by the development of a set of vectors by the Fire lab (Fire et al., 1990). These lacZ reporters were very useful for determining cis-acting transcriptional control elements (Fire and Waterston, 1989; MacMorris et al., 1994; Okkema et al., 1993) and lacZ continues to be a robust marker that can be pushed to detect even low levels of expression (Wilkinson and Greenwald, 1995).

More recently, Green Fluorescent Protein (GFP), or one of its variants, serves as the common reporter. A GFP coding cassette can be inserted in different locations within a large genomic clone (tens of kilobases) to generate transcriptional and/or translational fusions. These constructs provide the greatest chance of capturing required cis-acting regulatory elements. However, it is more common to make the assumption that genomic sequences 5′ to the coding region represent the core promoter. This region, usually a few kilobases in length, can be PCR amplified and easily cloned into the reporter gene backbone of choice. Alternatively, the Promoterome Project can serve as a source for many promoter regions and is useful if the reporter gene cloning strategy is Gateway (Invitrogen) compatible (Dupuy et al., 2004). These constructs are commercially available through Open Biosytems.

There are several considerations to take into account when making reporter genes. One is the distinction between transcriptional and translational reporters; often one would like to have both. For transcriptional reporters, expression can be engineered to highlight the cytoplasm, nucleus or other cellular compartments in the expressing cell. Nuclear localized reporters are useful for embryonic cell identification whereas cytoplasmic reporters are often more useful for larval cells, particularly in neurons where they highlight the axonal and dendritic tracks. The Chalfie lab is developing a two-part fluorescent expression system that has the potential to simplify cell type identification (Zhang et al., 2003). Translational reporter genes can provide information on the subcellular localization of the endogenous gene product. In this case, it is advisable to use the fusion protein to rescue a mutant phenotype, thus demonstrating that some or all aspects of the expression pattern and subcellular localization are biologically relevant.
Once a pattern of expression is determined, promoter analyses can be used to home in on important regulatory elements. Sequential deletions of putative promoter regions linked to a GFP reporter gene are easily made by traditional cloning or Splicing by Overlap Extension (SOEing) Polymerase Chain Reaction (PCR) amplification (Horton et al., 1990). The latter technique allows high throughput and convenience as PCR reactions can be injected directly into animals without purification or cloning (Hobert, 2002). As the control elements become localized to small genomic regions (several hundred base pairs or less), they can be placed upstream of “basal” promoters to assay for enhancer activity. These approaches are common in the C. elegans literature and can be very successful in defining important cis-acting regulatory sequences.

The use of reporter genes has several important caveats. First and foremost is that they are artificial and can easily misrepresent the pattern of gene expression. Important positive- and negative-acting control elements can be excluded by assuming promoter location leading to mosaicism, loss of expression, or ectopic expression (Krause et al., 1994). For example, reporter genes that lack sufficient control elements for fidelity are often expressed in the anterior and posterior intestinal cells or a small set of head neurons. Moreover, small changes in promoter regions can dramatically alter expression patterns as illustrated by the studies of the ges-1 gene (Egan et al., 1995). It is critical to confirm reporter gene expression patterns with an independent technique such as in situ hybridization, antibody staining, or mutant phenotype.

A second concern of reporter genes is the nature of any additional modules in the construct or in a co-injected marker that may have an effect on expression. For example, many of the standard constructs available from the Fire Lab have a 3′ untranslated region (UTR) derived from the unc-54 gene encoding muscle myosin heavy chain. These sequences may not be neutral when combined with promoters from other cell types. There are also reports of dramatic effects of co-transformation markers on expression levels suggesting that you should not rely on a single co-transformation marker when exploring the expression of a novel gene (Fukushige and Siddiqui, 1995). Similarly, “basal” promoters (e.g., pes-10) may be biased in working with certain types of genomic elements. This may cause them to fail to respond in certain cell types resulting in false information about a particular enhancer element (Natarajan et al., 2004). As long as you keep these limitations in mind, reporter genes can be very helpful in characterizing transcriptional control elements for the gene of interest.

Genome-wide approaches provide a more global assessment of transcriptional regulation and have begun to become more common in C. elegans. The availability of both spotted cDNA (see Kim Lab) and oligonucleotide microarrays (e.g., Affymetrix) for C. elegans has given birth to a large amount of gene expression data in response to tissue type (e.g., germline-enriched; Reinke et al., 2004), growth conditions (e.g., dauer; Liu et al., 2004), or mutant background (e.g., DAF-16; McElwee et al., 2003). Much of this data is available on web sites (e.g., http://genome-www5.stanford.edu/cgi-bin/login.pl) or is linked in Wormbase to individual genes. A second global approach for gene expression profiling is Serial Analysis of Gene Expression (SAGE) that has recently been combined with tissue isolation or cell type sorting (McKay et al., 2003; http://elegans.bcgsc.ca/home/sage.html). These approaches give an overview of expression. For specific genes, this data should be validated by an independent method, such as reverse transcriptase (RT)-PCR or reporter genes.

Bioinformatics provides another way to study gene regulation, either alone or in combination with other methods. Currently, the genome sequence of two Caenorhabditis species are finished (elegans and briggae), one is in draft (remanei), and two are planned (japonica and CB5161). Interspecific comparisons of non-coding regions provides a powerful tool in identifying important cis-acting regulatory elements controlling gene expression, as functional elements will remain constrained through evolution. Comparisons between C. elegans and C. briggae revealed important cis-acting sequences controlling the vitellogenin genes and helped to identify GATA-type transcription factors as likely regulators (MacMorris et al., 1994; Spieth et al., 1991; Winter et al., 1996; Zucker-Aprison and Blumenthal, 1989). Such comparison continue to provide valuable information about cis-acting sequences within gene promoter regions with many examples in the literature (Culetto et al., 1999; Kirouac and Sternberg, 2003; Marshall and McGhee, 2001; Natarajan et al., 2004; Teng et al., 2004). The power of these comparisons is increased as the number of species is increased and will thus become more informative as sequences of additional species are finished. Recently developed programs, such as FamilyJewels, provide methods for sophisticated multiple alignments (Brown et al., 2002). This approach will become widely exploited in coming years to pinpoint regulatory promoter elements.

Bioinformatic analysis of known transcription factor binding sites upstream of coding regions has also been successful. Given a known binding consensus site of sufficient length, the CisOrtho program can be used to ferret out a list of potential genes sharing expression patterns (Bigelow et al., 2004). Bioinformatic comparisons of
promoters from genes with the same or overlapping expression patterns can also be informative to home in on potential regulatory elements (for example, see Chang et al., 2004; Guhathakurta et al., 2004). A nice combination of bioinformatics and in vitro studies used the DNA binding properties of DAF-12, a regulator of dauer development and lifespan, to define potential binding sites and gene targets (Shostak et al., 2004). Regardless of the method used, candidate elements and gene targets should be validated experimentally by an independent means.

There are also several techniques for studying gene expression that, while commonplace in other organisms, are not routinely used in the worm. For example, one would ideally isolate pure populations of cells and tease apart transcriptional regulation at a biochemical level. At only 1mm in length as an adult, C. elegans makes tissue dissections tedious or impossible for generating enough homogeneous tissue for biochemical analysis. The recent development of cell culture techniques (see Methods in Cell biology), coupled with cell sorting, may make biochemical approaches more feasible in the future. However, the technique is still challenging enough that most researchers have opted for other methods to study transcription.

In situ hybridization is another common technique for cataloging transcriptional profiles in many organisms but it is less often used in C. elegans studies. The impermeable egg shell of C. elegans embryos and the cuticle of larvae and adults often lead to background hybridization or partially permeabilized animals, making it difficult to get in situ hybridization signals that are reproducible or trustworthy. Despite these difficulties, a genome-scale effort to catalog gene transcription profiles using in situ hybridization by the Kohara group is now underway. Their protocols and data are useful and can be accessed at http://nematode.lab.nig.ac.jp/db2/index.php.

3. Locating cis-acting regulatory elements

The majority of protein coding genes in C. elegans are within gene-dense regions of the genome. Consequently, cis-acting regulatory regions are usually close to the coding region. The minimal promoter region required for proper expression of most Pol II transcripts lies within a couple of kilobases upstream of the start codon. There are notable exceptions to this compact view of cis-acting sequences. For example, egl-l expression is controlled, in part, by an element located greater than 2 kb downstream of the coding region and beyond an unrelated, intervening gene (Thellmann et al., 2003). For lin-39, proper reporter gene expression required inclusion of ~30 kb of genomic DNA that extended upstream and downstream of the protein coding region (Wigmaister and Eisenmann, personal communication). Clearly C. elegans genes can have complex and distant control regions. However, a rule-of-thumb of 2 kb upstream of the ATG works well as a starting point in the search for cis-acting control elements.

It is important to remember that the minimal promoter region is not synonymous with the natural promoter. The natural promoter may span a much larger region due to redundancy in the function of regulatory elements that ensure proper and robust regulation of the endogenous gene. One common site of additional control elements is within introns. Most C. elegans introns are small (e.g., <100 bp; see Alternative splicing in C. elegans) and are thus unlikely to contain elements controlling expression. However, introns larger than several hundred base pairs do often have such elements (e.g., Nam et al., 2002; Okkema et al., 1993). Therefore, intron size can provide a clue in searching for transcriptional control sequences. Large introns, particularly at the beginning of a coding region, may also provide a clue to promoter organization and the presence of multiple transcriptional initiation sites. For example, nhr-23 has a 1.8 kb intron at the start of the gene that is included in one transcript and absent in a second (Kostrouchova et al., 1998). In cases such as this, the presence of a trans-spliced leader (see Trans-splicing and operons) on two or more different transcripts from a single gene can be an indicator of multiple messages, possibly encoding different protein isoforms.

4. Simple promoters

A simple promoter is defined here as one in which the cis-acting control elements necessary for proper expression are confined to a small region (a few hundreds of bp) of the genome. Housekeeping genes expressed in all tissues might be good candidates for regulation by simple promoters. Unfortunately, few housekeeping genes in C. elegans have been characterized. Among the best characterized simple promoters are those of the hsp-16 family of genes. This family consists of pairs of divergently transcribed genes with promoter regions sufficient for heat-regulated expression contained within the short (~350 bp) intragenic regions (Jones et al., 1986; Russnak and Candido, 1985; Stringham et al., 1992). Despite these compact promoters, distinct tissue expression patterns are induced from different hsp-16 promoters (Stringham et al., 1992), suggesting the presence of multiple regulatory sites within these simple promoters. Another excellent example of simple promoters are in the vitellogenin (vit) genes, which exhibit stage-, tissue- and sex-specific expression controlled, in the case of vit-2, by a 247 bp promoter.
vit-2 promoter activity depends on GATA-factor binding sites and a novel VPE2 site (TGTCAT) conserved in vit gene promoters in C. elegans and C. briggsae (Spieth et al., 1985; Zucker-Aprison and Blumenthal, 1989). Certain cell cycle promoters have also been shown to be remarkably simple. Analysis of several genes expressed only in proliferative cells and encoding G1 phase regulators (e.g., cyclin D) revealed that proper regulation minimally required a 67 bp region of the promoter (Brodigan et al., 2003; Park and Krause, 1999). How could genes with such dynamic expression profiles throughout development be regulated in an apparently simple way? The answer is likely that they are end effectors of a cell's decision to divide rather than integrating lineage or temporal information governing proliferation.

5. Complex promoters

The term complex is used here to describe a promoter in which the overall pattern of gene expression is the result of the composite action of several dispersed elements, each influencing or contributing to the overall expression pattern. This piecemeal organization has been described for the promoter region of several genes, including myo-2, hlh-1 and lin-26. These studies reveal examples in which spatial control of transcription is regulated by elements active in groups of cells related by cell-, tissue- and organ-type and by lineage history.

5.1. myo-2: activation of a terminal differentiation gene by the combined activities of organ- and cell type-specific regulatory elements

myo-2 encodes a myosin heavy chain expressed exclusively in the pharyngeal muscles as these cells undergo terminal differentiation (Ardizzi and Epstein, 1987; Miller et al., 1983). Characterization of the myo-2 promoter region in transgenic C. elegans and identification of trans-acting regulators indicates expression is regulated by a combination of organ- and cell type-specific signals targeting distinct regulatory sequences.

High level activity of the myo-2 promoter requires a transcriptional enhancer located approximately 300 bp upstream of the transcriptional start (Okkema et al., 1993). The intact myo-2 enhancer is active exclusively in the pharyngeal muscles, but, surprisingly, its activity depends on distinct cell-type-specific and organ-specific subelements, termed B and C, that can separately activate gene expression either specifically in the pharyngeal muscles, or more globally in all pharyngeal cell types (Okkema and Fire, 1994). In their endogenous context within the myo-2 gene, these subelements synergistically activate pharyngeal muscle gene expression.

Consistent with their distinct activities, the B and C subelements are targeted by transcription factors expressed in different spatial patterns in the pharynx (Figure 1). The cell-type-specific B subelement binds and is activated by the pharyngeal muscle specific NK-2 family homeodomain factor CEH-22 (Okkema and Fire, 1994; Okkema et al., 1997), which is structurally and functionally related to factors controlling cardiac muscle development in other species (Haun et al., 1998). The organ-specific C subelement binds and is activated by the pan-pharyngeal FoxA family transcription factor PHA-4 (Kalb et al., 1998), which is required for formation of pharyngeal muscle and all other pharyngeal cell types during embryonic development (see below).

CEH-22 is not the only factor functioning with PHA-4 to activate myo-2 expression. CEH-22 is expressed in most, but not all, myo-2 expressing pharyngeal muscles (Okkema and Fire, 1994). Likewise a ceh-22 mutant expresses myo-2, although these animals exhibit defects in B subelement activity and pharyngeal muscle development and function (Okkema et al., 1997). Thus, other as yet unidentified factors must contribute to myo-2 expression, and the characterization of these factors will enhance our understanding of pharyngeal muscle development.
Figure 1. CEH-22 and PHA-4 function in combination to activate pharyngeal muscle expression of myo-2. Myo-2 expression is activated by the pharyngeal muscle-specific CEH-22 and the pan-pharyngeal PHA-4, which bind the myo-2 enhancer B and C subelements, respectively. Micrographs indicate transgenic embryos expressing ceh-22::gfp in pharyngeal muscles and pha-4::gfp in all pharyngeal cells (top, delimited by arrowheads), and a transgenic adult expressing myo-2::lacZ in the pharyngeal muscles (bottom). Note pha-4::gfp is also expressed in the gut. GFP and β-galactosidase are targeted to nuclei to facilitate cell identification.

5.2. hlh-1: activation of gene expression by lineage-preference regulatory elements.

Hlh-1 encodes a basic helix-loop-helix transcription factor expressed in all body wall muscle cells and their precursors (Krause et al., 1990). The body wall muscle cells are derived from multiple cell lineages. Of the 81 body wall muscle cells born during embryogenesis, 1 is from the AB lineage, 28 are from the MS lineage, 32 are from the C lineage and 20 are from the D lineage (Sulston et al., 1983). An additional 14 body wall muscle cells (and other cell types) are born postembryonically from the M mesoblast (Sulston and Horvitz, 1977).

Dissection of the hlh-1 promoter shows that gene expression can be properly regulated by multiple elements spanning ~3 kb upstream of the ATG (Figure 2; Krause et al., 1994). A core element required for all expression resides just upstream of the ATG. In addition, there are several individual elements that drive expression preferentially in one or more lineages. However, no single element is specific for expression in just one lineage. In addition, the expression during embryogenesis is controlled by a different region than that controlling postembryonic expression. The overall pattern of hlh-1 expression is thus a composite of the action of several lineage-preference elements with overlapping domains of action, working in concert with an essential core element. Superimposed on this spatial pattern of regulation are distinct temporal control elements regulating timing of expression during development. As yet, no trans-acting factors have been identified that bind to the defined cis-acting elements, illustrating the difficulty in using promoter analysis alone to identify trans-acting factors.
Figure 2. Regulation of \textit{hlh-1} expression by lineage-preference elements. A) A schematic of body wall muscle nuclei is super-imposed on an image of a comma stage embryo. Each of four different lineages of origin is color-coded as shown in (B) (adapted from (Salston et al., 1983). C) The promoter and partial coding region (exons 1 and 2) of \textit{hlh-1} are shown (adapted from (Krause et al., 1990). All expression is dependent on a “core” element (star) located upstream of the ATG of exon 1. Below the gene structure diagram are color-coded elements that can direct lineage-preference expression of transgenes during embryogenesis; color coding as in (A) and (B). Mature body wall muscle is dependent on distinct temporal elements (purple boxes) that do not have lineage preferences.

5.3. \textit{lin-26}: activation of gene expression by tissue-specific regulatory elements.

\textit{lin-26} encodes a predicted zinc-finger transcription factor expressed in a broad range of ectodermally derived epithelial tissues, the somatic gonad and uterus (Labouesse et al., 1996; Labouesse et al., 1994). Within these ectodermally-derived epithelial tissues are the major hypodermis surrounding the body of the animal, specialized hypodermal cells located at the anterior and posterior ends of the body, and interfacial cells such as rectal cells connecting the external epithelium to the endoderm. A recent characterization of the \textit{lin-26} promoter region revealed this gene is regulated by a core element required for all expression working in concert with tissue-specific elements, rather than lineage-preference elements as discussed above for \textit{hlh-1} (Landmann et al., 2004).

\textit{lin-26} is the downstream gene in an alternatively spliced operon including \textit{lir-1} (Dufourcq et al., 1999), and proper expression of \textit{lin-26} requires an 11 kb upstream region including most of the \textit{lir-1} gene itself (den Boer et al., 1998). Within this region are tissue specific regulatory modules that activate gene expression in subsets of \textit{lin-26} expressing tissues (Figure 3; Landmann et al., 2004). For example, separable modules control expression in the major hypodermal cells, in the minor hypodermal cells and sheath and socket support cells, in rectal cells, or in the somatic gonad. In some cases, redundant elements contribute to expression in particular tissues (e.g., major hypodermal cells), and, in the case of the minor hypodermis and support cells located at the worms anterior and posterior ends, separable elements active either in anterior or posterior ends were identified. Thus, the \textit{lin-26} promoter region contains cis-regulatory elements active in cells that belong to the same organ, are functionally related, or have similar positions along the body (Landmann et al., 2004), and these elements together produce the full \textit{lin-26} expression pattern in a piecemeal fashion.
Regulation of *lin-26* by tissue-specific elements. A) A diagram of some of the cell types expressing *lin-26* (adapted from (Landmann et al., 2004)) is shown super-imposed on an image of a comma stage embryo. B) A partial embryonic lineage showing the origin of *lin-26* expressing cells with color coding matching the cell types shown in (A) (adapted from (Landmann et al., 2004)); major hypodermal cells include hyp 7 (green), seam cells (orange), and P cells (purple); support cells (red); somatic gonad precursors Z1 and Z4 (yellow). C) The promoter elements for *lin-26*. All expression is dependent on a "core" element (star) located in the intergenic region between *lin-26* and its upstream neighbor *lir-1*. Tissue-specific control elements, located within a *lir-1* intron, are shown below the gene structure diagram with color-coding as in (A) and (B). Most control elements function in cells related by tissue-type but not by lineage. Note also that temporal control is achieved by sequentially acting elements that are progressively further upstream from the ATG of *lin-26*.

One common theme to emerge from these three examples is redundancy of regulatory elements. In most cases, even when sub-elements are identified with specific tissue, lineage or organ activity, their loss does not prevent all expression in that region. Clearly endogenous gene regulation has evolved to include multiple and overlapping regulatory regions to ensure proper expression during development. The deconstruction of a promoter is most useful in showing a minimal set of cis-acting control elements. As studies employ more sophisticated techniques and assays, we may learn how extensive this redundancy is.

6. Trans-acting factors

The completion of the *C. elegans* genome makes it possible, in theory, to define all transcription factors in the worm. In practice, this effort is more difficult because of several uncertainties when surveying the properties of a given gene. For example, zinc finger motifs can bind DNA but also can serve other functions including RNA binding and protein-protein interactions. It is therefore difficult to conclude that a given gene product is indeed a transcription factor based solely on the presence of signature motifs. For factors that modify chromatin or participate in a transcription complex, the definition of a transcription factor often lies in the eyes of the investigator. A first-pass attempt at defining a list of *C. elegans* transcription factors is presented in Table 1. Originally compiled in the Sternberg Lab (courtesy of T. Ririe and J. Fernandes), we present a modified version of their list with the understanding that it will necessarily need refinement over time to correct inaccuracies and omissions. The current list includes 664 genes representing only about 3.5% of the predicted genes in *C. elegans*. This number is surprisingly low and about one half the number of transcription factors estimated previously (McGhee and Krause, 1997).
| Gene      | Affy probe set          | Description                                                                 |
|-----------|-------------------------|-----------------------------------------------------------------------------|
| gfl-1     | 190531_at               | AF-9-like                                                                   |
| taf-11.3  | 188157_at               | an ortholog of human TATA-binding protein associated factor TAF11         |
| R07C12.4  | 185116_s_at             | AP-1-like                                                                   |
| F28C6.1   | 191919_at               | AP-2-like                                                                   |
| F28C6.2   | 191940_at               | AP-2-like                                                                   |
| K06A1.1   | 191145_at               | AP-2-like                                                                   |
| Y62E10A.17| 186925_at               | AP-2-like                                                                   |
| Y73E7A.2  | 176887_at               | Apoptosis antagonizing transcription factor                                 |
| aha-1     | 172967_x_at, 172057_x_at| bHLH                                                                        |
| ahr-1     | 193149_at               | bHLH                                                                        |
| C15C8.2   | 192573_at               | bHLH                                                                        |
| cnd-1     | 187594_at               | bHLH                                                                        |
| F38C2.8   | 172921_x_at             | bHLH                                                                        |
| hif-1     | 183824_s_at             | bHLH                                                                        |
| hlh-1     | 193759_at               | bHLH                                                                        |
| hlh-10    | 189550_at               | bHLH                                                                        |
| hlh-11    | 171841_x_at             | bHLH                                                                        |
| hlh-12    | 186469_at               | bHLH                                                                        |
| hlh-13    | 182960_at               | bHLH                                                                        |
| hlh-14    | 192984_at               | bHLH                                                                        |
| hlh-15    | 182141_at               | bHLH                                                                        |
| hlh-16    | 192331_at               | bHLH                                                                        |
| hlh-17    | 172921_x_at             | bHLH                                                                        |
| hlh-19    | 193041_at               | bHLH                                                                        |
| hlh-2     | 193176_s_at             | bHLH                                                                        |
| hlh-21    | 183551_at               | bHLH                                                                        |
| hlh-25    | 184961_s_at             | bHLH                                                                        |
| hlh-26    | 183674_at               | bHLH                                                                        |
| hlh-27    | 184961_s_at             | bHLH                                                                        |
| hlh-28    | 179991_s_at             | bHLH                                                                        |
| hlh-29    | 179991_s_at             | bHLH                                                                        |
| hlh-3     | 192523_at               | bHLH                                                                        |
| hlh-4     | 190064_at               | bHLH                                                                        |
| hlh-6     | 193106_at               | bHLH                                                                        |
| hlh-8     | 193985_at               | bHLH                                                                        |
| hnd-1     | 192707_at               | bHLH                                                                        |
| lin-32    | 188671_at               | bHLH                                                                        |
| mdl-1     | 193723_s_at             | bHLH                                                                        |
### Gene Affy probe set Description

| Gene   | Affy probe set | Description |
|--------|----------------|-------------|
| ngn-1  | 176272_at      | bHLH        |
| T01D3.2 | 190290_at      | bHLH        |
| W02C12.3 | 190299_at    | bHLH        |
| Y105C5.29 | 172921_x_at | bHLH        |
| Y39A3.6  | 186966_at      | bHLH        |
| mxl-1   | 193662_at      | bHLH/ZIP    |
| mxl-2   | 184737_s_at    | bHLH/ZIP    |
| mxl-3   | 192645_at      | bHLH/ZIP    |
| bra-2   | 172599_x_at    | BMP receptor associated protein family |
| taf-3   | 177494_at      | bromodomain |
| atf-2   | 189943_at      | bZIP        |
| atf-5   | 192027_s_at    | bZIP        |
| atf-6   | 188462_at      | bZIP        |
| atf-7   | 173892_s_at, 172407_x_at | bZIP |
| C27D6.4 | 188956_at, 187831_at, 176715_at, 193833_s_at | bZIP |
| C34D1.5 | 193844_at      | bZIP        |
| C48E7.11 |             | bZIP        |
| ces-2   | 193437_s_at    | bZIP        |
| crh-1   | 189423_s_at    | bZIP        |
| F17A9.3 | 191134_at      | bZIP        |
| F23F12.9 | 187591_at    | bZIP        |
| F23F12.9 | 171733_x_at   | bZIP        |
| F29G9.4 | 191088_at      | bZIP        |
| F57B10.1 | 185841_s_at  | bZIP        |
| K02F3.4 | 181230_at      | bZIP        |
| mgl-2   | 173929_s_at, 193356_s_at | bZIP |
| pha-1   | 188075_at      | bZIP        |
| R07H5.10 | 181517_at     | bZIP        |
| skn-1   | 188421_at      | bZIP        |
| srx-41  | 183336_at      | bZIP        |
| T24H10.7 | 180818_at, 188946_at | bZIP |
| T27F2.4 | 179204_s_at    | bZIP        |
| W02H5.7 | 182521_at      | bZIP        |
| W07G1.3 | 179163_at, 179658_at | bZIP |
| W08E12.1 | 184700_at    | bZIP        |
| xbp-1   | 190863_at      | bZIP        |
| Y75B8A.29 | 185348_at    | bZIP        |
| ZC376.7 | 189598_s_at    | bZIP        |
| ZC8.4a.1 | 191675_s_at  | bZIP        |
| zip-1   | 173457_s_at    | bZIP        |
| Gene     | Affy probe set | Description                                                      |
|----------|----------------|------------------------------------------------------------------|
| T05C1.4  | 179917_at, 182254_s_at | calmodulin-binding transcription activator (CAMTA)                |
| cbp-1    | 173017_at, 191123_s_at | CBP/p300 homolog                                                  |
| lpd-2    | 182912_at        | CCAAT-binding                                                     |
| T08D10.1 | 189859_s_at      | CCAAT-binding                                                     |
| F23F1.1  | 174270_s_at      | CCAAT-binding, subunit C (HAP5)                                   |
| Y51H1A.5 | 182089_at        | CCAAT-binding                                                     |
| F22F1.3  | 194040_at        | KIX domain, coactivator CBP                                       |
| lag-1    | 175617_at, 192000_s_at | CSL                                                               |
| cdk-8    | 190709_at        | cyclin C interactor                                               |
| pqn-45   | 177704_at        | DEC-1-like                                                       |
| pqn-47   | 182066_s_at      | DEC-1-like                                                       |
| mab-23   | 192765_at        | DM DNA-binding                                                    |
| mab-3    | 192765_at        | DM DNA-binding                                                    |
| C27C12.6 | 173904_at        | DM DNA-binding                                                    |
| Y53F4B.3 | 181571_s_at      | DNA Polymerase epsilon subunit C                                  |
| F10C1.5  | 192453_at        | Doublesex                                                        |
| dpl-1    | 191593_s_at      | E2F/DP1                                                          |
| F49E12.6 | 175552_at, 189678_at | E2F/DP1                                                          |
| elf-1(mex-2)| 186476_s_at       | E2F/DP1                                                          |
| elf-2    | 186271_at        | E2F/DP1                                                          |
| C24A1.2  | 174325_at        | ETS domain                                                       |
| C33A11.4 | 188664_at        | ETS domain                                                       |
| C42D8.4  | 189894_at        | ETS domain                                                       |
| C50A2.4  | 183564_at        | ETS domain                                                       |
| C52B9.2  | 180922_at        | ETS domain                                                       |
| F19F10.1 | 190163_at        | ETS domain                                                       |
| F19F10.5 | 190230_at        | ETS domain                                                       |
| F22A3.1  | 192249_at        | ETS domain                                                       |
| lin-1    | 173446_s_at, 175607_s_at | ETS domain                                                   |
| T08H4.3  | 188791_s_at      | ETS domain                                                       |
| Y73F8A.14|                | ETS-related                                                      |
| peb-1    | 188015_at        | Zn-finger, FLYWCH                                                |
| C34B4.2  | 192149_at        | Forkhead                                                         |
| daf-16   | 188176_s_at, 181992_s_at | Forkhead                                                   |
| pes-1    | 188273_at        | Forkhead                                                         |
| pha-4    | 193962_at        | Forkhead                                                         |
| T27A8.2  | 189248_at        | Forkhead                                                         |
| unc-130  | 190731_s_at      | Forkhead                                                         |
| fkh-10   | 190179_at        | Forkhead                                                         |
| fkh-2    | 193957_s_at, 174011_at | Forkhead                                                   |
## Transcriptional regulation

| Gene     | Affy probe set                  | Description               |
|----------|---------------------------------|---------------------------|
| fkh-3    | 190724_s_at                     | Forkhead                  |
| fkh-4    | 172877_x_at                     | Forkhead                  |
| fkh-5    | 187411_at                       | Forkhead                  |
| fkh-6    | 191946_at                       | Forkhead                  |
| fkh-7    | 187837_at                       | Forkhead                  |
| fkh-8    | 181677_at                       | Forkhead                  |
| fkh-9    | 188781_at                       | Forkhead                  |
| let-381  | 175474_at                       | Forkhead                  |
| lin-31   | 188600_at                       | Forkhead                  |
| C18G1.2  | 191147_at, 183542_at, 174326_s_at| GATA                      |
| egl-18   | 190079_at                       | GATA                      |
| elt-1    | 192655_s_at                     | GATA                      |
| elt-2    | 193259_at                       | GATA                      |
| elt-3    | 193640_s_at                     | GATA                      |
| elt-4    |                                  | GATA                      |
| elt-6    | 185110_at                       | GATA                      |
| elt-7    |                                  | GATA                      |
| end-1    | 193618_at                       | GATA                      |
| end-3    | 193616_at                       | GATA                      |
| med-1    | 188376_at                       | GATA                      |
| med-2    |                                  | GATA                      |
| F55A3.3  | 171809_s_at, 190137_s_at         | global transcriptional regulator |
| F31F7.3  | 182888_at                       | Golden2-like              |
| unc-37   | 173188_s_at, 193614_s_at         | Groucho                   |
| egr-1    | 188149_s_at, 188166_s_at         | HDAC, GATA-like)          |
| K08B5.2  | 182979_at                       | heat shock                 |
| dro-1    | 192105_at                       | histone-like               |
| W10D9.4  | 180226_s_at                     | histone-like               |
| lin-22   | 175682_at                       | HLH                       |
| ref-1    | 193898_at                       | HLH                       |
| shp-1    | 192735_s_at                     | HLH                       |
| hmg-1.1  | 175799_at                       | HMG box                   |
| hmg-1.2  | 175806_at                       | HMG box                   |
| hmg-11   | 193301_at                       | HMG box                   |
| hmg-12   | 188169_at                       | HMG box                   |
| hmg-3    | 193484_s_at                     | HMG box                   |
| hmg-4    | 188651_s_at                     | HMG box                   |
| hmg-5    | 187989_at                       | HMG box                   |
| pop-1    | 188002_at                       | HMG box                   |
| gei-3    | 178044_at, 192981_at             | HMG-box                   |
| C02F12.5 | 192268_at                       | homeobox domain            |
| Gene     | Affy probe set          | Description  |
|----------|-------------------------|--------------|
| C07E3.6  | 189406_at, 178029_at    | homeobox domain |
| C09G12.1 | 189632_at               | homeobox domain |
| C12D12.4 | 184383_at               | homeobox domain |
| C12D12.5 | 184573_at               | homeobox domain |
| C17H12.9 | 189669_at               | homeobox domain |
| C18B12.3 | 189424_at               | homeobox domain |
| C36F7.1  | 189594_at               | homeobox domain |
| C49C3.5  | 189172_s_at             | homeobox domain |
| ceh-1    | 176628_at               | homeobox domain |
| ceh-10   | 176624_at               | homeobox domain |
| ceh-12   | 192019_at               | homeobox domain |
| ceh-13   | 176615_at               | homeobox domain |
| ceh-14   | 192631_s_at             | homeobox domain |
| ceh-16   | 188939_at               | homeobox domain |
| ceh-17   | 191721_at               | homeobox domain |
| ceh-19   | 192315_at               | homeobox domain |
| ceh-2    | 191750_at               | homeobox domain |
| ceh-22   | 175701_at, 193693_s_at   | homeobox domain |
| ceh-23   | 189145_at               | homeobox domain |
| ceh-24   | 193511_at               | homeobox domain |
| ceh-26   | 176620_at               | homeobox domain |
| ceh-27   | 192317_at               | homeobox domain |
| ceh-28   | 193479_at               | homeobox domain |
| ceh-30   | 189850_at               | homeobox domain |
| ceh-31   | 189951_at               | homeobox domain |
| ceh-32   | 193576_at               | homeobox domain |
| ceh-33   | 188645_at               | homeobox domain |
| ceh-34   | 192166_at               | homeobox domain |
| ceh-35   | 192928_at               | homeobox domain |
| ceh-36   | 173826_s_at             | homeobox domain |
| ceh-37   | 192014_s_at             | homeobox domain |
| ceh-40   | 192022_at               | homeobox domain |
| ceh-41   | 193249_at               | homeobox domain |
| ceh-43   | 193833_at               | homeobox domain |
| ceh-5    | 191117_at               | homeobox domain |
| ceh-6    | 194028_s_at             | homeobox domain |
| ceh-7    | 191844_at               | homeobox domain |
| ceh-8    | 193515_at               | homeobox domain |
| ceh-9    |                        | homeobox domain |
| egl-5    | 175712_s_at             | homeobox domain |
| F17A9.6  | 189546_s_at             | homeobox domain |
| Gene          | Affy probe set | Description     |
|--------------|----------------|-----------------|
| F22A3.5      | 190371_at      | homeobox domain |
| F42G2.6      | 181119_s_at    | homeobox domain |
| F45C12.15    | 186689_s_at    | homeobox domain |
| K02F3.8      | 181123_at      | homeobox domain |
| lin-39       | 189637_at      | homeobox domain |
| M6.3         | 189007_at      | homeobox domain |
| mab-18       | 194012_s_at    | homeobox domain |
| mab-5        | 173387_at, 176579_s_at | homeobox domain |
| mls-2        | 173754_at, 189498_s_at, 182023_at | homeobox domain |
| nob-1        | 191468_s_at    | homeobox domain |
| pal-1        | 193341_at, 193342_s_at | homeobox domain |
| pax-1        | 192418_at      | homeobox domain |
| pax-3        | 193866_at      | homeobox domain |
| php-3        | 191509_at      | homeobox domain |
| R04A9.5      | 193949_s_at    | homeobox domain |
| R06F6.6      | 189665_at      | homeobox domain |
| T13C5.4      | 189464_at      | homeobox domain |
| tab-1        | 189979_at      | homeobox domain |
| ttx-1        | 186833_s_at    | homeobox domain |
| unc-39       | 192928_at      | homeobox domain |
| unc-4        | 193581_at      | homeobox domain |
| unc-42       | 191585_at      | homeobox domain |
| vab-7        | 193846_at      | homeobox domain |
| Y38E10A.6    | 171992_x_at, 186209_s_at | homeobox domain |
| Y80D3A.3     | 182349_s_at    | homeobox domain |
| zag-1        | 193363_at      | homeobox domain |
| ceh-20       | 177324_s_at    | homeobox domain, cofactor |
| ceh-21       | 190942_at      | homeobox domain, CUT |
| ceh-38       | 189668_at      | homeobox domain, CUT |
| ceh-39       | 174323_at      | homeobox domain, CUT |
| ceh-44       | 176483_at, 176499_at | homeobox domain, CUT |
| ZK1193.5     | 185625_at      | homeobox domain, CUT |
| egl-13       | 188167_s_at, 193037_s_at | homeobox domain, HMG |
| sox-2        | 191681_at      | homeobox domain, HMG |
| C25G4.4      | 188729_at      | homeobox domain, HMG, SAND domain |
| eat-1        | 174122_s_at, 188775_s_at | homeobox domain, LIM |
| vab-15       | 192609_at      | homeobox domain, MSH |
| K03A11.5     | 180454_s_at    | homeobox domain, Nkx2 |
| cog-1        | 189136_at      | homeobox domain, Nkx6.1 |
| R08B4.2      | 192223_at      | homeobox domain, Paired domain |
| unc-30       | 174442_s_at, 193622_at | homeobox domain, Paired domain |
| Gene          | Affy probe set                      | Description                          |
|--------------|-------------------------------------|--------------------------------------|
| Y53C12C.1    | 192900_at                           | homeobox domain, Paired domain       |
| egl-38       | 193140_at                           | homeobox domain, Paired domain       |
| vab-3        | 194012_s_at                         | homeobox domain, Paired domain       |
| ceh-18       | 187890_at                           | homeobox domain, POU                 |
| unc-86       | 187875_s_at                         | homeobox domain, POU                 |
| unc-62       | 192045_s_at                         | homeobox domain, TALE                |
| B0496.7      | 181555_at                           | LIM domain                           |
| B0496.8      | 189866_s_at                         | LIM domain                           |
| C26C6.6      | 188741_at                           | LIM domain                           |
| C28H8.6      | 189207_at                           | LIM domain                           |
| C34B2.4      | 181869_at                           | LIM domain                           |
| F20D12.5     | 184796_s_at                         | LIM domain                           |
| F25H5.1      | 180879_at, 193411_s_at, 193324_at, 181484_s_at, 181483_at | LIM domain |
| F28F5.3      | 181521_at                           | LIM domain                           |
| F33D11.1     | 188952_at                           | LIM domain                           |
| F42H10.4     | 176229_at                           | LIM domain                           |
| lim-4        | 191102_at                           | LIM domain                           |
| lim-6        | 187894_at                           | LIM domain                           |
| lim-7        | 191686_at                           | LIM domain                           |
| lin-11       | 194088_at                           | LIM domain                           |
| ltd-1        | 188644_at                           | LIM domain                           |
| mec-3        | 194097_s_at                         | LIM domain                           |
| ttx-3        | 172067_x_at, 194096_s_at             | LIM domain                           |
| unc-115      | 190640_s_at                         | LIM domain                           |
| unc-95       | 176968_s_at                         | LIM domain                           |
| unc-97       | 193526_s_at                         | LIM domain                           |
| Y1A5A.1      | 193268_at                           | LIM domain                           |
| Y57G11A.1    | 192213_s_at                         | LIM domain                           |
| Y57G11A.3    | 192211_at                           | LIM domain                           |
| Y65B4A.7     | 176112_at, 176123_at                | LIM domain                           |
| ZK381.5      | 180400_at, 181271_s_at              | LIM domain                           |
| ZK622.4      | 187981_at                           | LIM domain                           |
| zyx-1        | 192073_at                           | LIM domain                           |
| pin-2        | 189051_at                           | LIM domain (Focal adhesion protein PINCH-1) |
| B0379.4a     | 193257_at, 178388_at                | LIM domain, (basal component)        |
| grh-1        | 186593_at                           | LSF/GRH-like                         |
| unc-120      | 188706_at                           | MADS box                             |
| mef-2        | 190718_s_at                         | MADS-box                             |
| dpy-22       | 187986_s_at, 173764_at              | mediator                             |
| sur-2        | 172005_x_at, 190922_s_at            | mediator                             |
| Gene       | Affy probe set                                      | Description                                                                                           |
|------------|-----------------------------------------------------|-------------------------------------------------------------------------------------------------------|
| Y62F5A.1   | 172520_x_at, 185066_s_at                            | mediator                                                                                             |
| F58H1.2    | 179510_at                                           | Metencephalon-mesencephalon-olfactorystranscription factor 1                                           |
| C50F4.12   | 183467_at                                           | Mitochondrial transcription termination factor, mTERF                                                 |
| egl-27     | 188531_s_at, 188530_at, 187748_s_at, 175679_at     | Myb-like DNA binding, GATA Zn finger                                                                |
| K11H12.8   | 184428_at, 184429_s_at                              | Myb-related                                                                                          |
| F40F9.7    | 190650_at                                           | NC2-like transcriptional repressor                                                                    |
| R11H6.5    | 190505_at                                           | NFAT-like                                                                                             |
| C13C4.1    | 194101_at                                           | NHR                                                                                                   |
| C14C6.4    | 190050_at                                           | NHR                                                                                                   |
| C17A2.1    | 190255_s_at                                         | NHR                                                                                                   |
| C25E10.1   | 192196_at                                           | NHR                                                                                                   |
| C26B2.4    | 190868_at                                           | NHR                                                                                                   |
| C33G8.10   | 190253_at                                           | NHR                                                                                                   |
| C33G8.12   | 191187_at                                           | NHR                                                                                                   |
| C33G8.7    | 190077_at                                           | NHR                                                                                                   |
| C33G8.8    | 190084_at                                           | NHR                                                                                                   |
| C41G6.5    | 192406_at                                           | NHR                                                                                                   |
| C49D10.9   | 190255_s_at                                         | NHR                                                                                                   |
| C50B6.8    | 194133_at                                           | NHR                                                                                                   |
| C54E10.5   | 192526_at                                           | NHR                                                                                                   |
| dpr-1      | 192538_at                                           | NHR                                                                                                   |
| F16B4.1    | 190004_at                                           | NHR                                                                                                   |
| F16B4.11   | 190260_at                                           | NHR                                                                                                   |
| F16H9.2    | 194105_at                                           | NHR                                                                                                   |
| F38H12.3   | 191042_at                                           | NHR                                                                                                   |
| F41D3.3    | 192456_at                                           | NHR                                                                                                   |
| F44A2.4    | 190088_at                                           | NHR                                                                                                   |
| F44E7.8    | 189031_at                                           | NHR                                                                                                   |
| F47C10.1   | 186434_at                                           | NHR                                                                                                   |
| F47C10.3   | 191031_at                                           | NHR                                                                                                   |
| F47C10.4   | 185749_at                                           | NHR                                                                                                   |
| F47C10.7   | 185192_at                                           | NHR                                                                                                   |
| F47C10.8   | 191126_at                                           | NHR                                                                                                   |
| F59E11.10  | 190380_at                                           | NHR                                                                                                   |
| F59E11.11  | 1872090_at                                          | NHR                                                                                                   |
| fax-1      | 175757_s_at                                         | NHR                                                                                                   |
| M02H5.5    | 176016_at                                           | NHR                                                                                                   |
| nhr-1      | 192168_at                                           | NHR                                                                                                   |
| nhr-10     | 191884_s_at                                         | NHR                                                                                                   |
| Gene  | Affy probe set                                                                 | Description |
|-------|-------------------------------------------------------------------------------|-------------|
| nhr-100 | 194112_at                                                                    | NHR         |
| nhr-101 | 172572_x_at                                                                 | NHR         |
| nhr-102 | 173975_s_at                                                                  | NHR         |
| nhr-103 | 190850_at                                                                    | NHR         |
| nhr-104 | 190875_at                                                                    | NHR         |
| nhr-105 | 189902_at                                                                    | NHR         |
| nhr-106 | 189964_at                                                                    | NHR         |
| nhr-107 | 191205_s_at                                                                  | NHR         |
| nhr-108 | 192601_s_at                                                                  | NHR         |
| nhr-109 | 175463_at, 182895_at                                                          | NHR         |
| nhr-11  | 174023_at, 192634_at                                                          | NHR         |
| nhr-110 | 188585_at, 188586_s_at                                                         | NHR         |
| nhr-111 | 192431_at                                                                    | NHR         |
| nhr-112 | 173159_s_at                                                                  | NHR         |
| nhr-113 | 192511_s_at                                                                  | NHR         |
| nhr-114 | 177144_at                                                                    | NHR         |
| nhr-115 | 189980_at                                                                    | NHR         |
| nhr-116 | 190812_at                                                                    | NHR         |
| nhr-117 | 190287_at                                                                    | NHR         |
| nhr-118 | 189855_at                                                                    | NHR         |
| nhr-119 | 187893_s_at, 190018_at, 176194_at                                              | NHR         |
| nhr-12  | 193687_at                                                                    | NHR         |
| nhr-120 | 180497_at                                                                    | NHR         |
| nhr-121 | 181125_at                                                                    | NHR         |
| nhr-122 | 176072_at                                                                    | NHR         |
| nhr-123 | 175973_at                                                                    | NHR         |
| nhr-124 | 190854_at                                                                    | NHR         |
| nhr-125 | 182971_at                                                                    | NHR         |
| nhr-126 | 188808_at                                                                    | NHR         |
| nhr-127 | 192488_at                                                                    | NHR         |
| nhr-128 | 189920_at                                                                    | NHR         |
| nhr-129 | 194133_at                                                                    | NHR         |
| nhr-13  | 176625_at                                                                    | NHR         |
| nhr-130 | 190001_at                                                                    | NHR         |
| nhr-131 | 189948_at                                                                    | NHR         |
| nhr-132 | 190237_at                                                                    | NHR         |
| nhr-133 | 189999_at                                                                    | NHR         |
| nhr-134 | 190842_at                                                                    | NHR         |
| nhr-135 | 177337_at                                                                    | NHR         |
| nhr-136 | 194102_at                                                                    | NHR         |
| nhr-137 | 185968_at                                                                    | NHR         |
| Gene  | Affy probe set       | Description |
|-------|----------------------|-------------|
| nhr-138 | --                   | NHR         |
| nhr-14  | 193175_at            | NHR         |
| nhr-15  | 192935_at            | NHR         |
| nhr-16  | 190172_at            | NHR         |
| nhr-17  | 192635_at            | NHR         |
| nhr-18  | 192777_at            | NHR         |
| nhr-19  | 193637_at            | NHR         |
| nhr-2   | 193015_s_at          | NHR         |
| nhr-20  | 193638_s_at          | NHR         |
| nhr-21  | 194090_s_at          | NHR         |
| nhr-22  | 193245_s_at          | NHR         |
| nhr-23  | 194130_s_at          | NHR         |
| nhr-25  | 193001_s_at          | NHR         |
| nhr-28  | 173241_at            | NHR         |
| nhr-3   | 193679_s_at          | NHR         |
| nhr-31  | 192541_s_at          | NHR         |
| nhr-32  | 192460_at            | NHR         |
| nhr-34  | 174018_at, 193049_at  | NHR         |
| nhr-35  | 190373_at            | NHR         |
| nhr-38  | 193582_at            | NHR         |
| nhr-4   | 194123_s_at          | NHR         |
| nhr-40  | 175720_at, 194035_s_at| NHR         |
| nhr-41  | 176582_at, 186589_at  | NHR         |
| nhr-42  | 192603_at            | NHR         |
| nhr-43  | 193483_at            | NHR         |
| nhr-44  | 192173_s_at          | NHR         |
| nhr-45  | 193486_at            | NHR         |
| nhr-46  | 192835_at            | NHR         |
| nhr-47  | 192271_s_at          | NHR         |
| nhr-48  | 193671_at            | NHR         |
| nhr-49  | 194120_at            | NHR         |
| nhr-5   | 188223_at            | NHR         |
| nhr-50  | 172035_x_at          | NHR         |
| nhr-51  | 193556_at            | NHR         |
| nhr-52  | 193670_at            | NHR         |
| nhr-53  | 193701_s_at          | NHR         |
| nhr-54  | 193559_s_at          | NHR         |
| nhr-55  | 192136_at            | NHR         |
| nhr-56  | 192231_at            | NHR         |
| nhr-57  | 192189_at            | NHR         |
| nhr-58  | 192282_at            | NHR         |
| Gene   | Affy probe set                      | Description |
|--------|-------------------------------------|-------------|
| nhr-59 | 172034_x_at, 172722_x_at            | NHR         |
| nhr-6  | 194121_s_at                         | NHR         |
| nhr-60 | 194128_s_at                         | NHR         |
| nhr-61 | 193603_s_at                         | NHR         |
| nhr-62 | 192819_at                           | NHR         |
| nhr-63 | 193566_at                           | NHR         |
| nhr-64 | 193768_at                           | NHR         |
| nhr-65 | 173003_s_at, 193567_at              | NHR         |
| nhr-66 | 192260_at                           | NHR         |
| nhr-67 | 192717_at                           | NHR         |
| nhr-68 | 192782_at                           | NHR         |
| nhr-69 | 194132_at                           | NHR         |
| nhr-7  | 192920_s_at                         | NHR         |
| nhr-70 | 193658_at                           | NHR         |
| nhr-71 | 193513_s_at                         | NHR         |
| nhr-72 | 192209_at                           | NHR         |
| nhr-73 | 192748_s_at                         | NHR         |
| nhr-74 | 192640_at, 192641_s_at              | NHR         |
| nhr-75 | 192291_at                           | NHR         |
| nhr-76 | 182867_at, 192270_at                | NHR         |
| nhr-77 | 193530_at                           | NHR         |
| nhr-78 | 188348_at                           | NHR         |
| nhr-79 | 193540_s_at                         | NHR         |
| nhr-8  | 193749_at                           | NHR         |
| nhr-80 | 190236_s_at, 191574_at              | NHR         |
| nhr-81 | 193507_at                           | NHR         |
| nhr-82 | 193508_at                           | NHR         |
| nhr-83 | 192858_at                           | NHR         |
| nhr-84 | 193542_at                           | NHR         |
| nhr-85 | 193753_at                           | NHR         |
| nhr-86 | 176830_s_at                         | NHR         |
| nhr-87 | 176042_s_at                         | NHR         |
| nhr-88 | 189858_at, 190898_at                | NHR         |
| nhr-89 | 193577_at                           | NHR         |
| nhr-9  | 188454_at                           | NHR         |
| nhr-90 | 189961_at                           | NHR         |
| nhr-91 | 193433_at                           | NHR         |
| nhr-92 | 174019_s_at, 176050_at              | NHR         |
| nhr-94 | 184113_at                           | NHR         |
| nhr-95 | 183502_at                           | NHR         |
| nhr-96 | 187786_at                           | NHR         |
| Gene         | Affy probe set | Description                   |
|--------------|----------------|-------------------------------|
| nhr-97       | 194122_at      | NHR                           |
| nhr-98       | 175979_at      | NHR                           |
| nhr-99       | 176003_at      | NHR                           |
| R11G11.12    | 190367_at      | NHR                           |
| R13D11.8     | 185889_at      | NHR                           |
| sex-1        | 172940_x_at, 188033_s_at, 171736_x_at | NHR |
| T01G6.5      | 189965_at      | NHR                           |
| T01G6.6      | 191004_at      | NHR                           |
| T03E6.3      | 189871_at      | NHR                           |
| T09D3.4      | 190944_at      | NHR                           |
| Y116A8C.18   | 186548_at      | NHR                           |
| Y17D7A.1     | 192543_at      | NHR                           |
| Y17D7B.1     | 192565_s_at    | NHR                           |
| Y22F5A.1     | 192564_at      | NHR                           |
| Y41D4B.21    | 176108_at      | NHR                           |
| Y54F10AM.1   | 176503_at      | NHR                           |
| Y80D3A.4     | 182290_at      | NHR                           |
| ZK455.6      | 192466_at      | NHR                           |
| ZK488.4      | 189947_at      | NHR                           |
| ZK697.2      | 190022_at      | NHR                           |
| lin-14       | 193913_s_at    | novel nuclear protein         |
| unc-3        | 193538_s_at    | HLH                           |
| F21A10.2     | 180693_at      | p53-like                      |
| Y51H4A.19    | 184538_s_at    | p53-like                      |
| pax-2        | 192273_at      | homeobox domain, Paired domain|
| C28H8.9      | 187205_at      | PHD-finger                    |
| C36C5.13     | 185704_at      | PHD-finger                    |
| C44B9.4      | 193372_at      | PHD-finger                    |
| F17A2.3      | 186658_at      | PHD-finger                    |
| lin-49       | 175774_at, 192161_s_at | PHD-finger                  |
| lin-59       | 188202_s_at    | PHD-finger                    |
| phf-5        |                | PHD-finger                    |
| T06A10.4     | 174921_s_at, 186168_at | PHD-finger                  |
| T23B12.1     | 185082_at      | PHD-finger                    |
| Y51H1A.4     | 188811_s_at    | PHD-finger                    |
| Y51H4A.12    | 184300_s_at    | PHD-finger                    |
| Y53G8AR.2    | 174165_at, 187052_at | PHD-finger                  |
| ZC132.2      | 179825_at      | PHD-finger                    |
| K04C1.2      | 182794_s_at    | Polycomb-group                |
| mes-2        | 190261_s_at    | Polycomb-group                |
| mes-6        | 187971_at      | Polycomb-group                |
## Transcriptional regulation

| Gene   | Affy probe set                     | Description                                      |
|--------|------------------------------------|--------------------------------------------------|
| *sop-2*| 185770_at, 186277_at               | Polycomb-group                                   |
| *mix-1*| 173070_s_at, 193889_at             | possible T.F.                                    |
| T12A7.6| 175158_s_at                       | possible T.F.                                    |
| *arx-6*| 187615_at                         | possible T.F., ARp2/3 complex component          |
| *cdk-9*| 193890_s_at                       | P-TEF-b component                                |
| *lin-35*| 188392_s_at                     | RB-like                                          |
| *daf-19*| 190486_s_at                     | RFX                                             |
| **lin-41**| 187703_s_at                  | RING finger                                      |
| C36E8.1| 186636_s_at                       | RNA polymerase I transcription factor           |
| C15H11.8| 192503_at                      | RNA polymerase I transcription factor TFIIS     |
| *icd-1*| 174688_at, 192927_s_at            | RNA polymerase II BTF3 (basal component)        |
| Y73B3A.8| 175882_at                     | RNA polymerase II subunit 9                     |
| R03D7.4| 186827_s_at                       | RNA polymerase II transcription elongation factor|
| *spt-5*| 193747_s_at                       | RNA polymerase II transcription elongation factor DSIF/SUPT5/SPT5 |
| *pqn-51*| 185179_s_at                     | RNA polymerase II transcription initiation factor TFIIB, large chain|
| T16H12.4| 174400_s_at                     | RNA polymerase II transcription initiation TFIIH |
| ZK1128.4| 180510_s_at                     | RNA polymerase II transcription initiation TFIIH |
| ZK856.13| 189860_s_at                     | RNA polymerase III transcription factor TFIIIC   |
| C01B12.2| 186210_at                      | SAND domain                                      |
| C44F1.2| 188653_at                        | SAND domain                                      |
| T21B10.5| 192141_at                       | SET domain                                       |
| *dac-1*| 189843_at                        | SKI/SNO domain                                   |
| *elc-1*| 192714_at                        | SKP1 component                                   |
| *daf-14*| 175659_at, 190867_s_at           | SMAD                                            |
| *daf-3*| 188906_s_at                      | SMAD                                            |
| *daf-8*|                                  | SMAD                                            |
| R05D11.1| 188788_at                      | SMAD                                            |
| Y113G7B.14| 187096_at                  | SNF2-related                                     |
| R07E5.3| 175088_at, 188857_at             | SWI/SNF                                          |
| *taf-11.2*| 192267_at                   | TAFII28-like protein                             |
| Y37E11B.2| 186683_at                     | TATA binding factor                              |
| *mab-9*| 192729_at                        | T-box                                           |
| mls-1   | 186222_at                        | T-box                                           |
| *tbx-11*| 187486_at                        | T-box                                           |
| *tbx-18*| 185577_s_at                      | T-box                                           |
| *tbx-2* | 188633_at                        | T-box                                           |
| *tbx-30*| 186686_s_at                      | T-box                                           |
| Gene     | Affy probe set                                      | Description          |
|----------|----------------------------------------------------|----------------------|
| tbx-31   | 182311_at                                         | T-box                |
| tbx-32   | 188373_at, 188374_s_at                            | T-box                |
| tbx-33   | 176669_at                                         | T-box                |
| tbx-34   | 186275_s_at                                       | T-box                |
| tbx-35   | 187661_at                                         | T-box                |
| tbx-36   | 178364_at                                         | T-box                |
| tbx-37   | 186121_at                                         | T-box                |
| tbx-38   | 188023_at                                         | T-box                |
| tbx-39   | 184577_at                                         | T-box                |
| tbx-40   | 184249_at, 184250_s_at                            | T-box                |
| tbx-41   | 189631_at                                         | T-box                |
| tbx-7    | 188450_at                                         | T-box                |
| tbx-8    | 190477_at                                         | T-box                |
| tbx-9    | 190539_s_at                                       | T-box                |
| Y59E9AR.5|                                                   | T-box                |
| taf-5    | 193468_at                                         | TBP-associated T.F.  |
| taf-6.1  | 188098_s_at                                       | TBP-associated T.F.  |
| taf-7.1  | 188233_at                                         | TBP-associated T.F.  |
| taf-8    |                                                   | TBP-associated T.F.  |
| egl-44   | 193230_at                                         | TEA/ATTS domain      |
| Y73F8A.24| 184394_s_at                                       | Tfb2 T.F.            |
| B0336.13 | 176733_at                                         | TFIIA, (basal component) |
| taf-10   | 188274_at                                         | TFIID                |
| taf-13   | 190935_at                                         | TFIID                |
| tbp-1    | 188606_at                                         | TFIID                |
| tlf-1    | 187785_s_at, 193370_s_at, 193369_at                | TFIID                |
| taf-4    | 174570_s_at, 187952_at                             | TFIID component      |
| taf-9    | 178055_at                                         | TFIID; TAFII-31      |
| cdk-7    | 172068_x_at, 176484_s_at                           | TFIIH complex (basal component) |
| brf-1    | 193633_s_at                                       | TFIIB (basal component) |
| Y77F11A.6| 186842_at                                         | TFIIS                |
| Y97E10AR.5| 176591_s_at                                       | TFIIS                |
| F10D7.4  | 180109_at                                         | Transcription elongation factor (basal component) |
| F59E12.9 | 185204_s_at                                       | Transcription elongation factor S-II, central region |
| sqt-4    | 188021_at                                         | Transcription elongation factor SPT4 |
| T24H10.1 | 191738_at                                         | Transcription elongation factor TFIIIS |
| Y51B9A.5 | 182001_at                                         | Transcription factor 21-related motif |
| Y38E10A.1| 185362_at                                         | Transcription factor Sp3  |
| C25H3.6  | 175492_at                                         | Transcription factor TFIIIS elongin A-like |
| Y38F2AR.13| 176361_at                                         | Transcription factorsIIIC-alpha subunit |
## Transcriptional regulation

| Gene          | Affy probe set                          | Description                                |
|---------------|-----------------------------------------|--------------------------------------------|
| Y38F2AR.5     | 176179_at, 176323_at                    | Transcription factorsIIIC-alpha subunit    |
| Y111B2A.13    | 193715_at                               | Transcription initiation factor IIA, gamma subunit |
| F54D5.11      | 193310_s_at                             | Transcription initiation factor IIE, beta subunit |
| C01F1.1       | 190074_s_at, 173313_s_at                | Transcription initiation factor IIF, alpha subunit |
| Y39B6A.36     | 183088_at                               | Transcription initiation factor IIF, small subunit (RAP30) |
| tnb-1         | 172103_x_at, 192210_at                  | Transcription initiation factor TFIIB       |
| Y56A3A.4      | 183864_s_at                             | transcription initiation factor TFIIID 20/15 kDa subunits |
| taf-1         | 193610_s_at                             | Transcription initiation factor TFIIID, subunit TAF1 |
| taf-12        | 183864_s_at                             | Transcription initiation factor TFIIID, subunit TAF12 |
| F01G4.1       | 174598_s_at                             | trithorax family                           |
| sdc-2         | 173447_s_at                             | X chromosome transcription repressor       |
| eor-1         | 189910_s_at                             | Zn finger                                  |
| eor-2         | 177725_at                               | Zn finger                                  |
| F22D6.2       | 179366_at                               | Zn finger                                  |
| arc-1         | 194111_at                               | Zn-finger                                  |
| asc-1         | 184727_s_at                             | Zn-finger                                  |
| C03G6.12      | 191749_at                               | Zn-finger                                  |
| C06E1.8       | 176000_at                               | Zn-finger                                  |
| C09F5.3       | 187378_at                               | Zn-finger                                  |
| C28G1.6       | 192344_at                               | Zn-finger                                  |
| C38D4.7       | 178513_at                               | Zn-finger                                  |
| C55C2.1       | 192307_at                               | Zn-finger                                  |
| ces-1         | 192612_at                               | Zn-finger                                  |
| che-1         | 188612_at                               | Zn-finger                                  |
| D1046.2       | 189843_at                               | Zn-finger                                  |
| daf-12        | 187714_at, 192797_s_at                  | Zn-finger                                  |
| dpy-20        | 191799_at                               | Zn-finger                                  |
| egl-43        | 193452_at                               | Zn-finger                                  |
| egl-46        | 192272_at                               | Zn-finger                                  |
| egl-9         | 187973_s_at                             | Zn-finger                                  |
| F33E11.2      | 174322_s_at, 184231_at                  | Zn-finger                                  |
| F36F12.8      | 190224_at                               | Zn-finger                                  |
| F45H11.1      | 190574_at                               | Zn-finger                                  |
| F47H4.1       | 173993_s_at                             | Zn-finger                                  |
| F53B3.1       | 182205_s_at                             | Zn-finger                                  |
| F53B3.1       | 182204_at                               | Zn-finger                                  |
| Gene     | Affy probe set | Description |
|----------|----------------|-------------|
| F53F8.1  | 192448_at      | Zn-finger   |
| F56F11.3 | 190273_at      | Zn-finger   |
| ham-2    | 190176_at      | Zn-finger   |
| hbl-1    | 192791_s_at    | Zn-finger   |
| K01H12.1 | 180815_at      | Zn-finger   |
| K11D2.4  | 192478_at      | Zn-finger   |
| lin-13   | 189214_s_at    | Zn-finger   |
| lin-26   | 190309_at      | Zn-finger   |
| lin-28   | 172855_x_at, 191973_s_at | Zn-finger |
### Transcriptional regulation

The goal in studying transcription is to make the link between transcription factors and their target genes. For a small number of genes in *C. elegans*, this connection has been made and a chart showing some of these is presented in Table 2. Notice that most transcription factors have been defined as either activators or repressors. However, for some, both modes of action have been described highlighting the importance of co-factors and promoter context within chromatin in determining the transcriptional outcome of DNA binding by these proteins. The list of potential target genes for several transcription factors will explode over the coming years with the application of microarray methods. However, most of these will not be specifically tested to determine if the regulation is direct and which cis-acting elements mediate the effect.

#### Table 2. Transcription factor target genes

| Class          | Number of genes | Factor type | Factor | Partner | Activator | Repressor | Putative target genes | DNA binding sequence | References            |
|----------------|-----------------|-------------|--------|---------|-----------|-----------|------------------------|----------------------|-----------------------|
| Homeo-domain   | 83              | HOX         | LIN-39 | CEH-20  | YES       | Possible  | *hlh-8, egl-17, egl-18, elt-5,* | TGATTAAT (G/T) (G/A) | Ruvkun and Hobert, 1998 |
|                |                 |             |        |         |           |           |                        |                      | Cui and Han, 2003; Koh et al., |
| Class         | Number of genes | Factor type | Factor | Partner | Activator | Repressor | Putative target genes | DNA binding sequence | References                           |
|--------------|-----------------|-------------|--------|---------|-----------|-----------|-----------------------|----------------------|--------------------------------------|
| Paired-like  |                 | CEH-10      | TTX-3  | YES     | N.D.      |           | elt-6, possible direct repressor of eff-1 | elt-6, possible direct repressor of eff-1 | 2002; Liu and Fire, 2000; Shemer and Podbilewicz, 2002 |
|              |                 | UNC-4       | UNC-37 | YES     |           |           | VB motor neuron genes | TAATY-NR-ATTA         | Winnier et al., 1999                  |
|              |                 | UNC-30      | N.D.   | YES     | N.D.      | unc-25, unc-47 | unc-25, unc-47 | unc-25, unc-47 | Eastman et al., 1999; Jin et al., 1994 |
|              |                 | EGL-38      | N.D.   | YES     | N.D.      | lin-48    | lin-48    | lin-48    | Johnson et al., 2001                  |
| Even-skipped | VAB-7           | N.D.        | N.D.   | YES     | unc-4 (may be indirect) | N.D. | unc-4 (may be indirect) | unc-4 (may be indirect) | Esmaeili et al., 2002                  |
| LIM          | TTX-3           | CEH-10      | YES    | N.D.    | ceh-23 and 38 other AIY terminal genes | AATTGG (C/T) TT (A/C) (G/A) TTA (G/A) | AATTGG (C/T) TT (A/C) (G/A) TTA (G/A) | Wenick and Hobert, 2004                 |
|              | UNC-86          | YES         | N.D.   | mec-4, mec-7; see also UNC-86 | CATNNNN-AATGCAT | CATNNNN-AATGCAT | CATNNNN-AATGCAT | Duggan et al., 1998; Way and Chalfie, 1988 |
| POU          | UNC-86          | MEC-3       | YES    | N.D.    | mec-3, 50 plus candidate genes from screen; snap-25 | TNNAGTG | TNNAGTG | TNNAGTG | Zhang et al., 2002; Hwang and Lee, 2003 |
| Zn-Finger    | CHE-1           | N.D.        | YES    | N.D.    | cog-1, ceh-36, gcy-5 other undefined ASE genes | N.D. | N.D. | N.D. | Chang et al., 2003                  |
| NK Class     | CEH-22          | N.D.        | YES    | N.D.    | myo-1, myo-2 (B element) | TNNAGTG | TNNAGTG | TNNAGTG | Okkema and Fire, 1994; Okkema et al., 1997 |
| EXD          | CEH-20          | LIN-39, UNC-62 (genetic evidence only) | YES | N.D. | hhl-8 | TGATTAAT | TGATTAAT | TGATTAAT | Liu and Fire, 2000; Van Auken et al., 2002 |
| Zn-Finger    | GATA            | ELT-1       | N.D.   | YES     | N.D.      | msp-113   | msp-113   | msp-113   | Shim, 1999; Shim et al., 1995           |
|              |                 | ELT-2       | N.D.   | YES     | N.D.      | ges-1, pho-1, mtl-1, mtl-2 | ges-1, pho-1, mtl-1, mtl-2 | ges-1, pho-1, mtl-1, mtl-2 | Egan et al., 1995; Fukushige et al., 1995 |
| Class          | Number of genes | Factor type | Factor | Partner | Activator | Repressor | Putative target genes | DNA binding sequence | References                                                                 |
|---------------|-----------------|-------------|--------|---------|-----------|-----------|-----------------------|---------------------|---------------------------------------------------------------------------|
|               |                  |             |        |         |           |           |                       |                     |                                                                            |
| END-1         | N.D.             | YES         | N.D.   | elt-2   | GATA      |           |                       |                     | 1998; Moilanen et al., 1999                                              |
| MED-1, MED-2  | N.D.             | YES         | N.D.   | end-1, end-2 | GATA      |           |                       |                     | Fukushige et al., 1998; Shoichet et al., 2000                              |
| SAL           | SEM-4            | N.D.        | N.D.   | YES     | elg-5, mec-3 | ACACAA    |                       |                     | Maduro et al., 2001; Maduro and Rothman, 2002; JHR unpublished            |
|               | CES-1            | N.D.        | N.D.   | YES     | egl-1     | CACCTG    |                       |                     | Toker et al., 2003                                                       |
| Gli           | TRA-1A           | N.D.        | N.D.   | YES     | egl-1, mab-3 | TGTGAGG-TC |                       |                     | Zarkower and Hodgkin, 1993; Conradt and Horvitz, 1999; Yi et al., 2000 |
| DM domain     | MAB-3            | N.D.        | N.D.   | YES     | vit-2 (probably all vits) | AATGTTG-CGA (T/A) NT |                     |                     | Yi and Zarkower, 1999                                                    |
| O/E           | UNC-3            | PAG-3?      | YES    | Possible acr-2 (indirect activator?) | N.D. |                     |                     | Prasad et al., 1998                                                     |
| GFI           | PAG-3            | UNC-3?      | N.D.   | YES     | pag-3     | TAAATCAC (A/T) GCA |                       |                     | McDermott and Aamodt, unpublished; Zweidler-Mckay et al., 1996          |
|               |                  |             |        |         |           |           |                       |                     |                                                                            |
| LIN-29        | N.D.             | YES         | N.D.   | YES     | col-19    | N.D.      |                       |                     | Rougvie and Ambros, 1995                                                 |
| NHR           | 284              |             |        |         |           |           |                       |                     | Maglich et al., 2001                                                     |
| FTZ-F1        | NHR-25           | N.D.        | YES    | N.D.    | lin-3     | TCAGGTC-CA |                       |                     | Hwang and Sternberg, 2004                                                |
| ROR/ RZR      | NHR-23           | N.D.        | YES    | N.D.    | dpy-7?    | AGGTCCAN-NNNNAG-GTCA |                       |                     | Kostrouchova et al., 1998; Kostrouchova et al., 2001                     |
| Vit D         | DAF-12           | N.D.        | YES    | YES     | ceh-22, myo-2, many others | CA(C/G)AC (A/G); AGT-GCANNNNN-NAGTGA |                     |                     | Ao et al., 2004; Shostak et al., 2004                                   |
| bHLH          | 39               |             |        |         |           |           |                       |                     | Ledent et al., 2002                                                      |
| MyoD          | HLH-1            | HLH-1       | YES    | N.D.    | N.D.      | CAGCTG    |                       |                     | Krause et al., 1992; Blackwell et al., 1994                               |
| E/Daughterless| HLH-2            | HLH-2; HLH-3; | YES    | N.D.    | lin-3, lag-2 | CACCTG    |                       |                     | Hwang and Sternberg, 2004;                                                |
| Class | Number of genes | Factor type | Factor | Partner | Activator | Repressor | Putative target genes | DNA binding sequence | References |
|-------|-----------------|-------------|--------|---------|-----------|-----------|-----------------------|----------------------|------------|
|       | LIN-32          | Achaete-scuter | HLH-3  | HLH-2   | YES       | N.D.      | egl-1                 | CACCTG              | Karp and Greenwald, 2003 |
|       |                 | Twist       | HLH-8  | HLH-8, HLH-2 | YES | N.D. | ceh-24, egl-15, mls-1 | CATATG; CAGGTG | Krause et al., 1997; Thellmann et al., 2003 |
| Twist |                 | Atonal        | LIN-32 | HLH-2   | YES       | N.D.      | N.D.; several genes identified that respond to overexpression | CACGTG | Portman and Emmons, 2004 |
|       |                 | bZIP         | 19     |         |           |           |                      | WWTRTC-AT          | Ruvkun and Hobert, 1998 |
|       |                 | Cap’n’ Collar | SKN-1  | N.D. | YES       | N.D. | med-1, med-2, end-1?, gcs-1 |                     | An and Blackwell, 2003; Blackwell et al., 1994; Carroll et al., 1997; Maduro et al., 2001; Walker et al., 2000 |
|       | Forkhead        | Forkhead    | 15     |         |           |           |                      |                     | Hope et al., 2003 |
|       | FOXO            | FOXO        | DAF-16 | N.D. | YES       | YES       | microarray            | TTGTTTAC           | Furuyama et al., 2000; Lee et al., 2003; McElwee et al., 2003 |
|       | PHA-4           | PHA-4       | PEB-1  | N.D. | YES       | N.D. | myo-2 (C element) | TGTGTGC           | Gaudet and Mango, 2002; Kalb et al., 1998; Okkema and Fire, 1994; Villimas et al., 2004 |
|       | UNC-1-30        | UNC-1-30    | N.D.  | N.D. | YES       |          | unc-129               | WTRTTNN-NNY        | Nash et al., 2000 |
|       | ETS             | ETS         | 10     |         |           |           |                      |                     | Hart et al., 2000 |
|       | T-box           | T-box       | 20     |         |           |           |                      |                     | Pocock et al., 2004 |
7. Spatial specificity

Spatial specificity refers to a pattern of gene expression that is limited to one or a few organs, tissues, or cell types. Examples of control mechanisms governing these types of spatial restriction are presented to show the logic underlying these patterns. Our current understanding shows that spatial specificity can be achieved by multiple mechanisms, ranging from the combinatorial action of overlapping transcription factors to transcriptional cascades.

7.1. Organ specificity: control of pharyngeal gene expression by a master regulator

The *C. elegans* pharynx is a complex organ consisting of five very different cell types, including muscles, neurons, epithelia, glands and marginal cells (Albertson and Thomson, 1976). The pharynx initially forms as a primordium of undifferentiated cells around mid-embryogenesis, and these cells subsequently differentiate and express cell type-specific genes (Sulston et al., 1983).

Formation of the pharynx and differentiation of all pharyngeal cell types depends on a single FoxA family transcription factor PHA-4 (Horner et al., 1998; Kalb et al., 1998; Mango et al., 1994). PHA-4 is expressed in all pharyngeal cells beginning at the time these cells become committed to a pharyngeal cell fate, as well as in the hindgut and intestine (Horner et al., 1998; Kalb et al., 1998). PHA-4 is believed to directly regulate most or all genes specifically expressed in the pharynx, including both early genes specifying fate of different pharyngeal cell types and late genes expressed during terminal differentiation (Gaudet and Mango, 2002). A major question in understanding pharyngeal development is how does PHA-4 regulate genes expressed in different pharyngeal cell types and at different times in pharyngeal development.
The function of PHA-4 in cell-type specific differentiation is best understood in the pharyngeal muscles. As discussed above, PHA-4 functions with the pharyngeal muscle-specific homeodomain factor CEH-22 to activate myo-2 expression during muscle cell differentiation (Kalb et al., 1998; Okkema and Fire, 1994). PHA-4 and CEH-22 similarly target a late functioning auto-regulatory enhancer from ceh-22 itself (Gaudet and Mango, 2002; Kuchenthal et al., 2001), suggesting these factors function together to control many regulatory sequences that function during terminal differentiation of the pharyngeal muscles. Earlier in pharyngeal muscle development, PHA-4 is also required for the initiation of ceh-22 expression (Mango et al., 1994; Vilimas et al., 2004), but the mechanism by which PHA-4 initially activates ceh-22 in the pharyngeal muscles remains unknown.

Less is known of how PHA-4 regulates specific gene expression in other pharyngeal cell types, largely because other pharyngeal specific promoters have not been extensively characterized. This situation, however, appears soon to be changed based on pioneering microarray studies that have identified >300 genes preferentially expressed in the pharynx (Ao et al., 2004; Gaudet and Mango, 2002). Analyses of these genes’ known expression patterns, in situ hybridization patterns, and placement on the Gene Expression Topo Map have identified clusters of genes expressed preferentially in subsets of pharyngeal cells, and comparisons of the promoters of genes within these clusters have identified conserved regulatory elements that likely impart positional or cell type specificity to PHA-4 target genes (Ao et al., 2004).

PHA-4 also regulates genes in temporally distinct patterns in the pharynx. This temporal regulation may involve both the affinity of PHA-4 for its binding sites in various gene promoters (Gaudet and Mango, 2002), and the presence of binding sites for additional factors functioning with PHA-4 (Gaudet et al., 2004). These mechanisms are likely not mutually exclusive and may be interdependent, as additional factors could affect PHA-4 binding affinity by cooperative binding.

7.2. Tissue specificity: regulation of gut gene expression by a cascade of redundant GATA factors.

The E blastomere is the clonal precursor of the gut, and the maternal factors specifying E blastomere identity are well understood. One effect of these maternal factors is to initiate zygotic gene expression, including expression of a series of sequentially functioning GATA family transcription factors expressed exclusively in the gut lineage (reviewed in Maduro and Rothman, 2002; Figure 4). These GATA factors bind WGATAR motifs required in many gut specific genes and directly activate gut gene expression (e.g., Britton et al., 1998; Egan et al., 1995; MacMorris et al., 1992; Nam et al., 2002).

The first of these gut-specific GATA factors is END-1, which is expressed transiently in the E lineage, beginning in the E cell itself and continuing until approximately the 8E stage (Figure 4; Zhu et al., 1997). ELT-2 is then expressed on one cell division later, beginning at the 2 E-cell stage (Fukushige et al., 1998). elt-2 expression is activated by END-1 (Zhu et al., 1998), but, unlike end-1, elt-2 remains expressed in the gut throughout the life of the worm through an autoregulatory mechanism (Fukushige et al., 1998; Fukushige et al., 1999). Interestingly, both end-1 and elt-2 appear to be members of redundant gene families. While ectopic expression of either of these genes activates widespread gut differentiation, loss-of-function studies reveal surprisingly mild defects in gut gene expression (Fukushige et al., 1998; Zhu et al., 1998; Zhu et al., 1997). Indeed, end-1 loss-of-function produces no phenotype. In comparison, elt-2 loss produces gut defects and lethality, while the effect on gene expression varies

![Figure 4](https://example.com/figure4.jpg)
from promoter to promoter (Fukushige et al., 2005; Fukushige et al., 1998; Oskouian et al., 2005). In both cases, the likely suspects for redundant genes encode additional GATA factors. end-1 may be redundant with the linked gene end-3, while elt-2 may be partially redundant with elt-7 (Maduro and Rothman, 2002). Thus, end-1 and end-3 are believed to establish endoderm fate in the E lineage, while elt-2 and elt-7 are likely the direct regulators of most genes expressed in the gut (Figure 4).

While most gut-specific promoters contain WGATAR motifs, their accurate regulation depends on more than simply turning on ELT-2. Gut genes are expressed under distinct temporal, sex-specific, and environmental controls, indicating other factors must contribute to gut gene regulation. In the case of the vit-2 gene, which encodes a yolk protein expressed only in the gut of adult hermaphrodites, repression in males requires the MAB-3 DM-domain protein (Yi and Zarkower, 1999). Likewise, ges-1, which encodes a gut-specific esterase, is activated in the gut by ELT-2 while being repressed in other regions of the digestive system by an unknown factor binding near the WGATAR motifs (Fukushige et al., 1996; Marshall and McGhee, 2001). In most cases, the identity of factors functioning with ELT-2 remain unknown, and there remains much to be learned about gut transcription.

7.3. Cell type specificity: regulation of AIY neuronal expression by a single core motif

Cell type specificity of gene expression is best exemplified by studies of neuronal gene expression (for example Chang et al., 2004; Zhang et al., 2004). Hobert and colleagues have studied the mechanism that regulates gene expression in a single pair of bilateral interneurons in the head called AIY left and right (AIYL & AIYR) that function in sensory input processing, learning, and memory (Ishihara et al., 2002; Mori and Ohshima, 1995; Tsalik et al., 2003). Differentiation of these interneurons is dependent on the transcription factors ceh-10 (Paired homeobox) and ttx-3 (LIM homeobox; Altun-Gultekin et al., 2001). Analysis of several promoters of genes expressed in AIY, including ceh-10 and ttx-3, revealed a consensus 16 bp AIY motif responsible for proper expression and comprising the core of an element that functions as an AIY-specific enhancer (Wenick and Hobert, 2004). This enhancer element is active in combination with some non-neuronal cell type promoters but not others demonstrating that promoter context is an important aspect of transcriptional regulation. Both ceh-10 and ttx-3 are part of an autoregulatory loop that activates their own expression, explaining the presence of an AIY motif within each of their promoters.

Control of AIY gene expression by ceh-10 and ttx-3 provides some insight into the logic of cell type-specific gene regulation (Wenick and Hobert, 2004). Cell type specificity is generated by using a combination of transcription factors that are unique to AIY interneurons in concert with a modular AIY response element upstream of target genes. Although some of the identified ceh-10/ttx-3 target genes were AIY specific, others were generally expressed in neurons or non-neuronal tissues. However, in most cases, expression in AIY was dependent on the AIY motif demonstrating that widespread expression may often be the composite action of several cell type-specific cis-acting elements. Finally, cis-acting control of gene targets encoding terminal differentiation products in AIY appeared to lack repressive elements. This suggests that the integration of positive and negative signals influencing cell type specificity is carried out by upstream transcription factors, ceh-10 and ttx-3 in this case. Once these upstream factors are activated, the downstream target gene battery will ensue largely independent of other influences.

8. Future

There is little doubt that the field of transcriptional regulation is on the verge of an information explosion. The combination of genome sequences from multiple Caenorhabditis species, microarray transcriptional profiling, and improved methodology will soon lead to a wealth of information on transcriptional activators and downstream target genes. One challenge will be the experimental verification of the mountains of data that will become available about upstream activators and downstream targets. Can these relationships be confirmed by independent approaches and are the interactions direct or indirect? We are entering an age in which the connections between most trans-acting factors and cis-acting regulatory target elements will be defined. Understanding how these connections regulate development will add an exciting chapter in the study of the worm and for transcriptional regulation in general.

9. Acknowledgements

This research was supported in part by the Intramural Research Program of the NIH, National Institute of Diabetes, Digestive and Kidney Diseases, and by grants from the NIH (GM053996) and the American Heart Association (03505487).
10. References

Albertson, D.G., and Thomson, J.N. (1976). The pharynx of *Caenorhabditis elegans*. Philos. Trans. R Soc. Lond. B Biol. Sci. 275, 299–325. Abstract

Altun-Gultekin, Z., Andachi, Y., Tsalik, E.L., Pilgrim, D., Kohara, Y., and Hobert, O. (2001). A regulatory cascade of three homeobox genes, *ceh-10*, *txr-3* and *ceh-23*, controls cell fate specification of a defined interneuron class in *C. elegans*. Development 128, 1951–1969. Abstract

An, J.H., and Blackwell, T.K. (2003). *SKN-1* links *C. elegans* mesendodermal specification to a conserved oxidative stress response. Genes Dev. 17, 1882–1893. Abstract

Ao, W., Gaudet, J., Kent, W.J., Muttumu, S., and Mango, S.E. (2004). Environmentally induced foregut remodeling by PHA-4/FoxA and DAF-12/NHR. Science 305, 1743–1746. Abstract Article

Ardizzi, J.P., and Epstein, H.F. (1987). Immunohistochemical localization of myosin heavy chain isoforms and paramyosin in developmentally and structurally diverse muscle cell types of the nematode *Caenorhabditis elegans*. J. Cell Biol. 105, 2763–2770. Abstract

Beaster-Jones, L., and Okkema, P.G. (2004). DNA binding and in vivo function of *C. elegans* *PEB-1* require a conserved FLYWCH motif. J. Mol. Biol. 339, 695–706. Abstract Article

Bigelow, H.R., Wenick, A.S., Wong, A., and Hobert, O. (2004). CisOrtho: a program pipeline for genome-wide identification of transcription factor target genes using phylogenetic footprinting. BMC Bioinformatics 5, 27. Abstract Article

Blackwell, T.K., Bowerman, B., Priess, J.R., and Weintraub, H. (1994). Formation of a monomeric DNA binding domain by Skn-1 bZIP and homeodomain elements. Science 266, 621–628. Abstract

Britton, C., McKerrow, J.H., and Johnstone, I.L. (1998). Regulation of the *Caenorhabditis elegans* gut cysteine protease gene *cpr-1*: requirement for GATA motifs. J. Mol. Biol. 283, 15–27. Abstract

Brodigan, T.M., Liu, J., Park, M., Kipreos, E.T., and Krause, M. (2003). Cyclin E expression during development in *Caenorhabditis elegans*. Dev. Biol. 254, 102–115. Abstract

Brown, C.T., Rust, A.G., Clarke, P.J., Pan, Z., Schilstra, M.J., De Buyscher, T., Griffin, G., Wold, B.J., Cameron, R.A., Davidson, E.H., and Bolouri, H. (2002). New computational approaches for analysis of cis-regulatory networks. Dev. Biol. 246, 86–102. Abstract Article

Carroll, A.S., Gilbert, D.E., Liu, X., Cheung, J.W., Michnowicz, J.E., Wagner, G., Ellenberger, T.E., and Blackwell, T.K. (1997). *SKN-1* domain folding and basic region monomer stabilization upon DNA binding. Genes Dev. 11, 2227–2238. Abstract

Chang, S., Johnston, R.J., Jr., Frokjaer-Jensen, C., Lockery, S., and Hobert, O. (2004). MicroRNAs act sequentially and asymmetrically to control chemosensory lateralization in the nematode. Nature 430, 785–789. Abstract Article

Chang, S., Johnston, R.J., Jr., and Hobert, O. (2003). A transcriptional regulatory cascade that controls left/right asymmetry in chemosensory neurons of *C. elegans*. Genes Dev. 17, 2123–2137. Abstract Article

Choi, K.Y., Ji, Y.J., Lee, C., Kim do, H., and Ahnn, J. (2002). Characterization of CeHDA-7, a class II histone deacetylase interacting with MEF-2 in *Caenorhabditis elegans*. Biochem. Biophys. Res. Commun. 293, 1295–1300. Abstract Article

Christensen, S., Kodoyianni, V., Bosenberg, M., Friedman, L., and Kimble, J. (1996). *lag-1*, a gene required for *lin-12* and *glp-1* signaling in *Caenorhabditis elegans*, is homologous to human CBF1 and Drosophila Su(H). Development 122, 1373–1383. Abstract

Corsi, A.K., Costas, S.A., Fire, A., and Krause, M. (2000). *Caenorhabditis elegans* twist plays an essential role in non-striated muscle development. Development 127, 2041–2051. Abstract
Cui, M., and Han, M. (2003). Cis regulatory requirements for vulval cell-specific expression of the *Caenorhabditis elegans* fibroblast growth factor gene *egl-17*. Dev. Biol. 257, 104–116. Abstract

Culetto, E., Combes, D., Fedon, Y., Roig, A., Toutant, J.P., and Arpagaus, M. (1999). Structure and promoter activity of the 5' flanking region of *ace-1*, the gene encoding acetylcholinesterase of class A in *Caenorhabditis elegans*. J. Mol. Biol. 290, 951–966. Abstract Article

den Boer, B.G., Sookhareea, S., Dufourcq, P., and Labouesse, M. (1998). A tissue-specific knock-out strategy reveals that *lin-26* is required for the formation of the somatic gonad epithelium in *Caenorhabditis elegans*. Development 125, 3213–3224. Abstract

Dantonel, J.C., Quintin, S., Lakatos, L., Labouesse, M., and Tora, L. (2000). TBP-like factor is required for embryonic RNA polymerase II transcription in *C. elegans*. Mol. Cell 6, 715–722. Abstract

Dichoso, D., Brodigan, T., Chwoe, K.Y., Lee, J.S., Llacer, R., Park, M., Corsi, A.K., Kostas, S.A., Fire, A., Ahnn, J., and Krause, M. (2000). The MADS-Box factor CeMEF2 is not essential for *Caenorhabditis elegans* myogenesis and development. Dev. Biol. 223, 431–440. Abstract Article

Dufourcq, P., Chanal, P., Vicaire, S., Camut, E., Quintin, S., den Boer, B.G., Bosher, J.M., and Labouesse, M. (1999). *lir-2*, *lir-1* and *lin-26* encode a new class of zinc-finger proteins and are organized in two overlapping operons both in *Caenorhabditis elegans* and in *Caenorhabditis briggsae*. Genetics 152, 221–235. Abstract

Duggan, A., Ma, C., and Chalfie, M. (1998). Regulation of touch receptor differentiation by the *Caenorhabditis elegans* *mec-3* and *unc-86* genes. Development 125, 4107–4119. Abstract

Dupuy, D., Li, Q.R., Deplancke, B., Boxem, M., Hao, T., Lamesch, P., Sequerras, R., Bosak, S., Doucette-Stamm, L., Hope, I.A., *et al.* (2004). A first version of the *Caenorhabditis elegans* Promoterome. Genome Res. 14, 2169–2175. Abstract Article

Eastman, C., Horvitz, H.R., and Jin, Y. (1999). Coordinated transcriptional regulation of the *unc-25* glutamic acid decarboxylase and the *unc-47* GABA vesicular transporter by the *Caenorhabditis elegans* UNC-30 homeodomain protein. J. Neurosci. 19, 6225–6234. Abstract

Egan, C.R., Chung, M.A., Allen, F.L., Heschl, M.F., Van Buskirk, C.L., and McGhee, J.D. (1995). A gut-to-pharynx/tail switch in embryonic expression of the *Caenorhabditis elegans* *ges-1* gene centers on two GATA sequences. Dev. Biol. 170, 397–419. Abstract

Esmaeili, B., Ross, J.M., Neades, C., Miller, D.M., 3rd, and Ahringer, J. (2002). The *C. elegans* even-skipped homologue, *vab-7*, specifies DB motoneurone identity and axon trajectory. Development 129, 853–862. Abstract

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. Abstract

Fire, A., and Waterston, R.H. (1989). Proper expression of myosin genes in transgenic nematodes. EMBO J. 8, 3419–3428. Abstract

Fukushige, T., Goszczynski, B., Yan, J., and McGhee, J.D. (2005). Transcriptional control and patterning of the *pho-1* gene, an essential acid phosphatase expressed in the *C. elegans* intestine. Dev. Biol. 279, 446–461. Abstract Article

Fukushige, T., Hawkins, M.G., and McGhee, J.D. (1998). The GATA-factor *elt-2* is essential for formation of the *Caenorhabditis elegans* intestine. Dev. Biol. 198, 286–302. Abstract

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. Abstract

Fukushige, T., Schroeder, D.F., Allen, F.L., Goszczynski, B., and McGhee, J.D. (1996). Modulation of gene expression in the embryonic digestive tract of *C. elegans*. Dev. Biol. 178, 276–288. Abstract
Fukushige, T., and Siddiqui, S.S. (1995). Effect of the dpy-20 and rol-6 cotransformation markers on alpha-tubulin gene expression in C. elegans transformants. Transgenic Res. 4, 332–340. Abstract

Furuyama, T., Nakazawa, T., Nakano, I., and Mori, N. (2000). Identification of the differential distribution patterns of mRNAs and consensus binding sequences for mouse DAF-16 homologues. Biochem. J. 349, 629–634. Abstract

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

Gaudet, J., Muttumu, S., Horner, M., and Mango, S.E. (2004). Whole-genome analysis of temporal gene expression during foregut development. PLoS Biol. 2, e352. Abstract Article

Guhathakurta, D., Schrier, L.A., Waterston, R.H., and Stormo, G.D. (2004). Novel transcription regulatory elements in Caenorhabditis elegans muscle genes. Genome Res. 14, 2457–2468. Abstract Article

Harfe, B.D., and Fire, A. (1998). Muscle and nerve-specific regulation of a novel NK-2 class homeodomain factor in Caenorhabditis elegans. Development 125, 421–429. Abstract

Harfe, B.D., Vaz Gomes, A., Kenyon, C., Liu, J., Krause, M., and Fire, A. (1998). Analysis of a Caenorhabditis elegans Twist homolog identifies conserved and divergent aspects of mesodermal patterning. Genes Dev. 12, 2623–2635. Abstract

Hart, A.H., Reventar, R., and Bernstein, A. (2000). Genetic analysis of ETS genes in C. elegans. Oncogene 19, 6400–6408. Abstract Article

Haun, C., Alexander, J., Stainier, D.Y., and Okkema, P.G. (1998). Rescue of Caenorhabditis elegans pharyngeal development by a vertebrate heart specification gene. Proc. Natl. Acad. Sci. USA 95, 5072–5075. Abstract

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

Hope, I.A., Mounsey, A., Bauer, P., and Aslam, S. (2003). The forkhead gene family of Caenorhabditis elegans. Gene 304, 43–55. Abstract

Horner, M.A., Quintin, S., Domeier, M.E., Kimble, J., Labouesse, M., and Mango, S.E. (1998). pha-4, an HNF-3 homolog, specifies pharyngeal organ identity in Caenorhabditis elegans. Genes Dev. 12, 1947–1952. Abstract

Horton, R.M., Cai, Z.L., Ho, S.N., and Pease, L.R. (1990). Gene splicing by overlap extension: tailor-made genes using the polymerase chain reaction. Biotechniques 8, 528–535. Abstract

Hwang, B.J., and Sternberg, P.W. (2004). A cell-specific enhancer that specifies lin-3 expression in the C. elegans anchor cell for vulval development. Development 131, 143–151. Abstract Article

Hwang, S.B., and Lee, J. (2003). Neuron cell type-specific SNAP-25 expression driven by multiple regulatory elements in the nematode Caenorhabditis elegans. J. Mol. Biol. 333, 237–247. Abstract

Ishihara, T., Iino, Y., Mohri, A., Mori, I., Gengyo-Ando, K., Mitani, S., and Katsura, I. (2002). HEN-1, a secretory protein with an LDL receptor motif, regulates sensory integration and learning in Caenorhabditis elegans. Cell 109, 639–649. Abstract

Jin, Y., Hoskins, R., and Horvitz, H.R. (1994). Control of type-D GABAergic neuron differentiation by C. elegans UNC-30 homeodomain protein. Nature 372, 780–783. Abstract Article

Johnson, A.D., Fitzsimmons, D., Hagman, J., and Chamberlin, H.M. (2001). EGL-38 Pax regulates the ovo-related gene lin-48 during Caenorhabditis elegans organ development. Development 128, 2857–2865. Abstract

Jones, D., Russnak, R.H., Kay, R.J., and Candido, E.P. (1986). Structure, expression, and evolution of a heat shock gene locus in Caenorhabditis elegans that is flanked by repetitive elements. J. Biol. Chem. 261, 12006–12015. Abstract
Kalb, J.M., Lau, K.K., Goszczynski, B., Fukushige, T., Moons, D., Okkema, P.G., and McGhee, J.D. (1998). *pha-4* is Ce-fkh-1, a fork head/HNF-3alpha,beta,gamma homolog that functions in organogenesis of the *C. elegans* pharynx. Development 125, 2171–2180. Abstract

Kaltenbach, L., Horner, M.A., Rothman, J.H., and Mango, S.E. (2000). The TBP-like factor CeTLF is required to activate RNA polymerase II transcription during *C. elegans* embryogenesis. Mol. Cell 6, 705–713. Abstract

Karp, X., and Greenwald, I. (2003). Post-transcriptional regulation of the E/Daughterless ortholog HLH-2, negative feedback, and birth order bias during the AC/VU decision in *C. elegans*. Genes Dev. 17, 3100–3111. Abstract Article

Kirouac, M., and Sternberg, P.W. (2003). cis-Regulatory control of three cell fate-specific genes in vulval organogenesis of *Caenorhabditis elegans* and *C. briggsae*. Dev. Biol. 257, 85–103. Abstract

Koh, K., Peyrot, S.M., Wood, C.G., Wagmaister, J.A., Maduro, M.F., Eisenmann, D.M., and Rothman, J.H. (2002). Cell fates and fusion in the *C. elegans* vulval primordium are regulated by the EGL-18 and ELT-6 GATA factors – apparent direct targets of the LIN-39 Hox protein. Development 129, 5171–5180. Abstract

Kostas, S.A., and Fire, A. (2002). The T-box factor MLS-1 acts as a molecular switch during specification of nonstriated muscle in *C. elegans*. Genes Dev. 16, 257–269. Abstract Article

Kostrouchova, M., Krause, M., Kostrouch, Z., and Rall, J.E. (1998). CHR3: a *Caenorhabditis elegans* orphan nuclear hormone receptor required for proper epidermal development and molting. Development 125, 1617–1626. Abstract

Kostrouchova, M., Krause, M., Kostrouch, Z., and Rall, J.E. (2001). Nuclear hormone receptor CHR3 is a critical regulator of all four larval molts of the nematode *Caenorhabditis elegans*. Proc. Natl. Acad. Sci. USA 98, 7360–7365. Abstract Article

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. Abstract

Krause, M., Fire, A., White-Harrison, S., Weintraub, H., and Tapscott, S. (1992). Functional conservation of nematode and vertebrate myogenic regulatory factors. J. Cell Sci. 16(Suppl.), 111–115. Abstract

Krause, M., Harrison, S.W., Xu, S.Q., Chen, L., and Fire, A. (1994). Elements regulating cell- and stage-specific expression of the *C. elegans* MyoD family homolog hlh-1. Dev. Biol. 166, 133–148. Abstract

Krause, M., Park, M., Zhang, J.M., Yuan, J., Harfe, B., Xu, S.Q., Greenwald, I., Cole, M., Paterson, B., and Fire, A. (1997). A *C. elegans* E/Daughterless bHLH protein marks neuronal but not striated muscle development. Development 124, 2179–2189. Abstract

Kuchenthal, C.A., Chen, W., and Okkema, P.G. (2001). Multiple enhancers contribute to expression of the NK-2 homeobox gene ceh-22 in *C. elegans* pharyngeal muscle. Genesis 31, 156–166. Abstract

Labouesse, M., Hartweg, E., and Horvitz, H.R. (1996). The *Caenorhabditis elegans* LIN-26 protein is required to specify and/or maintain all non-neuronal ectodermal cell fates. Development 122, 2579–2588. Abstract

Labouesse, M., Sookhareea, S., and Horvitz, H.R. (1994). The *Caenorhabditis elegans* gene lin-26 is required to specify the fates of hypodermal cells and encodes a presumptive zinc-finger transcription factor. Development 120, 2359–2368. Abstract

Landmann, F., Quintin, S., and Labouesse, M. (2004). Multiple regulatory elements with spatially and temporally distinct activities control the expression of the epithelial differentiation gene lin-26 in *C. elegans*. Dev. Biol. 265, 478–490. Abstract

Ledent, V., Paquet, O., and Vervoort, M. (2002). Phylogenetic analysis of the human basic helix-loop-helix proteins. Genome Biol. 3, RESEARCH0030. Abstract
Lee, S.S., Kennedy, S., Tolonen, A.C., and Ruvkun, G. (2003). DAF-16 target genes that control C. elegans life-span and metabolism. Science 300, 644–647. Abstract Article

Lichtsteiner, S., and Tjian, R. (1993). Cloning and properties of the Caenorhabditis elegans TATA-box-binding protein. Proc. Natl. Acad. Sci. USA 90, 9673–9677. Abstract

Liu, J., and Fire, A. (2000). Overlapping roles of two Hox genes and the exd ortholog ceh-20 in diversification of the C. elegans postembryonic mesoderm. Development 127, 5179–5190. Abstract

Liu, T., Zimmerman, K.K., and Patterson, G.I. (2004). Regulation of signaling genes by TGFbeta during entry into dauer diapause in C. elegans. BMC Dev. Biol. 4, 11. Abstract Article

MacMorris, M., Broverman, S., Greenspoon, S., Lea, K., Madej, C., Blumenthal, T., and Spieth, J. (1992). Regulation of vitellogenin gene expression in transgenic Caenorhabditis elegans: short sequences required for activation of the vit-2 promoter. Mol. Cell. Biol. 12, 1652–1662. Abstract

MacMorris, M., Spieth, J., Madej, C., Lea, K., and Blumenthal, T. (1994). Analysis of the VPE sequences in the Caenorhabditis elegans vit-2 promoter with extrachromosomal tandem array-containing transgenic strains. Mol. Cell. Biol. 14, 484–491. Abstract

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. Abstract

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. Abstract Article

Maglich, J.M., Sluder, A., Guan, X., Shi, Y., McKee, D.D., Carrick, K., Kamdar, K., Willson, T.M., and Moore, J.T. (2001). Comparison of complete nuclear receptor sets from the human, Caenorhabditis elegans and Drosophila genomes. Genome Biol. 2, RESEARCH0029. Abstract

Mango, S.E., Lambie, E.J., and Kimble, J. (1994). The pha-4 gene is required to generate the pharyngeal primordium of Caenorhabditis elegans. Development 120, 3019–3031. Abstract

Marshall, S.D., and McGhee, J.D. (2001). Coordination of ges-1 expression between the Caenorhabditis pharynx and intestine. Dev. Biol. 239, 350–363. Abstract Article

McElwee, J., Bubb, K., and Thomas, J.H. (2003). Transcriptional outputs of the Caenorhabditis elegans forkhead protein DAF-16. Aging Cell 2, 111–121. Abstract

McGhee, J.D., and Krause, M.W. (1997). Transcription Factors and Transcriptional Regulation. In: C. elegans II, D.L. Riddle, T. Blumenthal, B.J. Meyer, and J.R. Priess, eds., Cold Spring Harbor, NY, CSHL Press, pp. 147–184.

McKay, S.J., Johnsen, R., Khattra, J., Asano, J., Baillie, D.L., Chan, S., Dube, N., Fang, L., Goszczyński, B., Ha, E., et al. (2003). Gene expression profiling of cells, tissues, and developmental stages of the nematode C. elegans. Cold Spring Harb. Symp. Quant. Biol. 68, 159–169. Abstract

Miley, G.R., Fantz, D., Glossip, D., Lu, X., Saito, R.M., Palmer, R.E., Inoue, T., Van Den Heuvel, S., Sternberg, P.W., and Kornfeld, K. (2004). Identification of residues of the Caenorhabditis elegans LIN-1 ETS domain that are necessary for DNA binding and regulation of vulval cell fates. Genetics 167, 1697–1709. Abstract Article

Miller, D.M., 3rd, Ortiz, I., Berliner, G.C., and Epstein, H.F. (1983). Differential localization of two myosins within nematode thick filaments. Cell 34, 477–490. Abstract

Moilanen, L.H., Fukushima, T., and Freedman, J.H. (1999). Regulation of metallothionein gene transcription. Identification of upstream regulatory elements and transcription factors responsible for cell-specific expression of the metallothionein genes from Caenorhabditis elegans. J. Biol. Chem. 274, 29655–29665. Abstract
Mori, I., and Ohshima, Y. (1995). Neural regulation of thermotaxis in *Caenorhabditis elegans*. Nature 376, 344–348. Abstract Article

Nam, S., Jin, Y.H., Li, Q.L., Lee, K.Y., Jeong, G.B., Ito, Y., Lee, J., and Bae, S.C. (2002). Expression pattern, regulation, and biological role of runt domain transcription factor, run, in *Caenorhabditis elegans*. Mol. Cell. Biol. 22, 547–554. Abstract

Nash, B., Colavita, A., Zheng, H., Roy, P.J., and Culotti, J.G. (2000). The forkhead transcription factor UNC-130 is required for the graded spatial expression of the UNC-129 TGF-beta guidance factor in *C. elegans*. Genes Dev. 14, 2486–2500. Abstract

Natarajan, L., Jackson, B.M., Szyleyko, E., and Eisenmann, D.M. (2004). Identification of evolutionarily conserved promoter elements and amino acids required for function of the *C. elegans* beta-catenin homolog BAR-1. Dev. Biol. 272, 536–557. Abstract Article

Okkema, P.G., and Fire, A. (1994). The *Caenorhabditis elegans* NK-2 class homeoprotein CEH-22 is involved in combinatorial activation of gene expression in pharyngeal muscle. Development 120, 2175–2186. Abstract

Okkema, P.G., Ha, E., Haun, C., Chen, W., and Fire, A. (1997). The *Caenorhabditis elegans* NK-2 homeobox gene *ceh-22* activates pharyngeal muscle gene expression in combination with *pha-1* and is required for normal pharyngeal development. Development 124, 3965–3973. Abstract

Okkema, P.G., Harrison, S.W., Plunger, V., Aryana, A., and Fire, A. (1993). Sequence requirements for myosin gene expression and regulation in *Caenorhabditis elegans*. Genetics 135, 385–404. Abstract

Oskouian, B., Mendel, J., Shocron, E., Lee, M.A., Jr., Fyrst, H., and Saba, J.D. (2005). Regulation of sphingosine-1-phosphate lyase gene expression by members of the GATA family of transcription factors. J. Biol. Chem. Abstract Article

Park, M., and Krause, M.W. (1999). Regulation of postembryonic G(1) cell cycle progression in *Caenorhabditis elegans* by a cyclin D/CDK-like complex. Development 126, 4849–4860. Abstract

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. Abstract Article

Portman, D.S., and Emmons, S.W. (2004). Identification of *C. elegans* sensory ray genes using whole-genome expression profiling. Dev. Biol. 270, 499–512. Abstract Article

Powell-Coffman, J.A., Bradfield, C.A., and Wood, W.B. (1998). *Caenorhabditis elegans* orthologs of the aryl hydrocarbon receptor and its heterodimerization partner the aryl hydrocarbon receptor nuclear translocator. Proc. Natl. Acad. Sci. USA 95, 2844–2849. Abstract

Prasad, B.C., Ye, B., Zackhary, R., Schrader, K., Seydoux, G., and Reed, R.R. (1998). *unc-3*, a gene required for axonal guidance in *Caenorhabditis elegans*, encodes a member of the O/E family of transcription factors. Development 125, 1561–1568. Abstract

Reinke, V., Gil, I.S., Ward, S., and Kazmer, K. (2004). Genome-wide germline-enriched and sex-biased expression profiles in *Caenorhabditis elegans*. Development 131, 311–323. Abstract Article

Rougvie, A.E., and Ambros, V. (1995). The heterochronic gene *lin-29* encodes a zinc finger protein that controls a terminal differentiation event in *Caenorhabditis elegans*. Development 121, 2491–2500. Abstract

Russnak, R.H., and Candido, E.P. (1985). Locus encoding a family of small heat shock genes in *Caenorhabditis elegans*: two genes duplicated to form a 3.8-kilobase inverted repeat. Mol. Cell. Biol. 5, 1268–1278. Abstract

Ruvkun, G., and Hobert, O. (1998). The taxonomy of developmental control in *Caenorhabditis elegans*. Science 282, 2033–2041. Abstract
Seydoux, G., and Dunn, M.A. (1997). Transcriptionally repressed germ cells lack a subpopulation of phosphorylated RNA polymerase II in early embryos of *Caenorhabditis elegans* and *Drosophila melanogaster*. *Development* 124, 2191–2201. Abstract

Shaham, S., and Bargmann, C.I. (2002). Control of neuronal subtype identity by the *C. elegans* ARID protein CFI-1. *Genes Dev.* 16, 972–983. Abstract Article

Shemer, G., and Podbilewicz, B. (2002). LIN-39/Hox triggers cell division and represses EFF-1/fusogen-dependent vulval cell fusion. *Genes Dev.* 16, 3136–3141. Abstract Article

Shim, Y.H. (1999). *elt-1*, a gene encoding a *Caenorhabditis elegans* GATA transcription factor, is highly expressed in the germ lines with nsp genes as the potential targets. *Mol. Cells* 9, 535–541. Abstract

Shim, Y.H., Bonner, J.J., and Blumenthal, T. (1995). Activity of a *C. elegans* GATA transcription factor, ELT-1, expressed in yeast. *J. Mol. Biol.* 253, 665–676. Abstract

Shoichet, S.A., Malik, T.H., Rothman, J.H., and Shivdasani, R.A. (2000). Action of the *Caenorhabditis elegans* GATA factor END-1 in Xenopus suggests that similar mechanisms initiate endoderm development in ecdysozoa and vertebrates. *Proc. Natl. Acad. Sci. USA* 97, 4076–4081. Abstract

Shostak, Y., Van Gilst, M.R., Antebi, A., and Yamamoto, K.R. (2004). Identification of *C. elegans* DAF-12-binding sites, response elements, and target genes. *Genes Dev.* 18, 2529–2544. Abstract Article

Spieth, J., Denison, K., Kirtland, S., Cane, J., and Blumenthal, T. (1985). The *C. elegans* vitellogenin genes: short sequence repeats in the promoter regions and homology to the vertebrate genes. *Nucleic Acids Res.* 13, 5283–5295. Abstract

Spieth, J., Shim, Y.H., Lea, K., Conrad, R., and Blumenthal, T. (1991). *elt-1*, an embryonically expressed *Caenorhabditis elegans* gene homologous to the GATA transcription factor family. *Mol. Cell. Biol.* 11, 4651–4659. Abstract

Stringham, E.G., Dixon, D.K., Jones, D., and Candido, E.P. (1992). Temporal and spatial expression patterns of the small heat shock (hsp16) genes in transgenic *Caenorhabditis elegans*. *Mol. Biol. Cell* 3, 221–233. Abstract

Sulston, J.E., and Horvitz, H.R. (1977). Post-embryonic cell lineages of the nematode, *Caenorhabditis elegans*. *Dev. Biol.* 56, 110–156. Abstract

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. Abstract

Teng, Y., Girard, L., Ferreira, H.B., Sternberg, P.W., and Emmons, S.W. (2004). Dissection of cis-regulatory elements in the *C. elegans* Hox gene *egl-5* promoter. *Dev. Biol.* 276, 476–492. Abstract Article

Thatcher, J.D., Fernandez, A.P., Beaster-Jones, L., Haun, C., and Okkema, P.G. (2001). The *Caenorhabditis elegans* peb-1 gene encodes a novel DNA-binding protein involved in morphogenesis of the pharynx, vulva, and hindgut. *Dev. Biol.* 229, 480–493. Abstract

Thatcher, J.D., Haun, C., and Okkema, P.G. (1999). The DAF-3 Smad binds DNA and represses gene expression in the *Caenorhabditis elegans* pharynx. *Development* 126, 97–107. Abstract

Thellmann, M., Hatzold, J., and Conradt, B. (2003). The Snail-like CES-1 protein of *C. elegans* can block the expression of the BH3-only cell-death activator gene *egli-1* by antagonizing the function of bHLH proteins. *Development* 130, 4057–4071. Abstract

Toker, A.S., Teng, Y., Ferreira, H.B., Emmons, S.W., and Chalfie, M. (2003). The *Caenorhabditis elegans* split-like gene *sem-4* restricts touch cell fate by repressing the selector Hox gene *egl-5* and the effector gene *mec-3*. *Development* 130, 3831–3840. Abstract
Tsalik, E.L., Niacaris, T., Wenick, A.S., Pau, K., Avery, L., and Hobert, O. (2003). LIM homeobox gene-dependent expression of biogenic amine receptors in restricted regions of the C. elegans nervous system. Dev. Biol. 263, 81–102. Abstract

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. Abstract

Venkatesan, K., McManus, H.R., Mello, C.C., Smith, T.F., and Hansen, U. (2003). Functional conservation between members of an ancient duplicated transcription factor family, LSF/Grainyhead. Nucleic Acids Res. 31, 4304–4316. Abstract

Vilimas, T., Abraham, A., and Okkema, P.G. (2004). An early pharyngeal muscle enhancer from the Caenorhabditis elegans ceh-22 gene is targeted by the Forkhead factor PHA-4. Dev. Biol. 266, 388–398. Abstract

Walker, A.K., See, R., Batchelder, C., Kophengnavong, T., Gronniger, J.T., Shi, Y., and Blackwell, T.K. (2000). A conserved transcription motif suggesting functional parallels between Caenorhabditis elegans SKN-1 and Cap’n’Collar-related basic leucine zipper proteins. J. Biol. Chem. 275, 22166–22171. Abstract Article

Walker, A.K., Shi, Y., and Blackwell, T.K. (2004). An extensive requirement for transcription factor IID-specific TAF-1 in Caenorhabditis elegans embryonic transcription. J. Biol. Chem. 279, 15339–15347. Abstract Article

Wallenfang, M.R., and Seydoux, G. (2002). cdk-7 Is required for mRNA transcription and cell cycle progression in Caenorhabditis elegans embryos. Proc. Natl. Acad. Sci. USA 99, 5527–5532. Abstract Article

Way, J.C., and Chalfie, M. (1988). mec-3, a homeobox-containing gene that specifies differentiation of the touch receptor neurons in C. elegans. Cell 54, 5–16. Abstract

Wenick, A.S., and Hobert, O. (2004). Genomic cis-regulatory architecture and trans-acting regulators of a single interneuron-specific gene battery in C. elegans. Dev. Cell 6, 757–770. Abstract Article

Wilkinson, H.A., and Greenwald, I. (1995). Spatial and temporal patterns of lin-12 expression during C. elegans hermaphrodite development. Genetics 141, 513–526. Abstract

Winner, A.R., Meir, J.Y., Ross, J.M., Tavermarakis, N., Driscoll, M., Ishihara, T., Katsura, I., and Miller, D.M., 3rd (1999). UNC-4/UNC-37-dependent repression of motor neuron-specific genes controls synaptic choice in Caenorhabditis elegans. Genes Dev. 13, 2774–2786. Abstract

Winter, C.E., Penha, C., and Blumenthal, T. (1996). Comparison of a vitellogenin gene between two distantly related rhabditid nematode species. Mol. Biol. Evol. 13, 674–684. Abstract

Yi, W., and Zarkower, D. (1999). Similarity of DNA binding and transcriptional regulation by Caenorhabditis elegans MAB-3 and Drosophila melanogaster DSX suggests conservation of sex determining mechanisms. Development 126, 873–881. Abstract

Zhang, F., Barboric, M., Blackwell, T.K., and Peterlin, B.M. (2003). A model of repression: CTD analogs and PIE-1 inhibit transcriptional elongation by P-TEFb. Genes Dev. 17, 748–758. Abstract Article

Zhang, S., Ma, C., and Chalfie, M. (2004). Combinatorial marking of cells and organelles with reconstituted fluorescent proteins. Cell 119, 137–144. Abstract Article

Zhu, J., Fukushima, 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.

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. Abstract
Zucker-Aprison, E., and Blumenthal, T. (1989). Potential regulatory elements of nematode vitellogenin genes revealed by interspecies sequence comparison. J. Mol. Evol. 28, 487–496. Abstract

Zweidler-Mckay, P.A., Grimes, H.L., Flubacher, M.M., and Tsichlis, P.N. (1996). Gfi-1 encodes a nuclear zinc finger protein that binds DNA and functions as a transcriptional repressor. Mol. Cell. Biol. 16, 4024–4034. Abstract