Chapter 1

Transcriptional *cis*-Regulation

Introduction

The process of differential gene expression, or the selective activation of different subsets of genes, leads to unique populations of cells that are terminally differentiated. Selective activation is carefully regulated and, ultimately, controls all functions of cells, tissues and organs. Central to the process of differential gene expression and cell fate specification are the *cis*-regulatory elements of genes that are responsible for determining the temporal and spatial domains of gene expression. These *cis*-regulatory elements are part of the larger transcriptional machinery that controls the production of gene products that establish and maintain unique cell populations**.**

Caenorhabditis elegans is a free-living, soil-dwelling nematode. All 959 somatic cells of its transparent, 1mm-long body are visible with a microscope. It has a rapid life cycle (14-hour embryogenesis and 36-hour postembryonic development through four larval stages, L1-L4, to the adult) (reviewed in Riddle *et al.*, 1997). The development and function of this organism is encoded by an estimated 19,476 genes (www.wormbase.org; release WS84). Within this genome are the genes that encode the developmental program of the vulva. The vulva of *C. elegans* provides an excellent system to study the mechanisms by which *cis*-regulatory controls are utilized in establishing differential gene expression and terminal differentiation**.**

cis-acting regulatory elements of transcription in eukaryotes

The typical eukaryotic gene consists of up to four distinct *cis*-regulatory transcriptional control elements: the promoter itself, the upstream promoter elements (UPEs), elements

adjacent to the promoter that are interspersed with the UPEs, and distinct enhancer elements (reviewed in Latchman, 1998).

Upstream elements contain two types of sequences. The first type are those sequences, which are found in many genes that exhibit distinct patterns of regulation, and are likely to be involved in the basic process of transcription. These are referred to as the basal transcription machinery. The second type of sequences are those that are only in genes transcribed in a particular tissue, or in response to a specific signal. This type of transcription is referred to as regulated transcription (reviewed in Latchman, 1998).

Several sequences characterize the typical eukaryotic basal transcription machinery. The first is the TATA box element. This TATA sequence is found 25-30 bp upstream of the transcriptional start site in most genes, although it is sometimes absent, as in many housekeeping genes. The region delimited by the TATA box and the sites of transcriptional initiation (the cap site) has been defined as the gene promoter (reviewed in Latchman, 1998). The promoter probably binds several proteins essential for transcription, as well as RNA polymerase II, the enzyme that is responsible for the transcription of the genes (reviewed in Sentenac, 1985). Genes may also contain UPEs, such as the CCAAT and Sp1 boxes, which, if found, are typically upstream of the TATA box (reviewed in McKnight and Tjian, 1986). In every instance that they have been found, they are essential for the transcription of the genes (reviewed in Latchman, 1998).

The binding of particular proteins to specific upstream sequences in order to confer on a gene the ability to respond to particular stimuli is known as regulated transcription. To prove that an element found in one group of common genes is important for that group's transcriptional activity, the sequence must confer the same response or

expression to an unrelated gene. A classic example of regulated transcription was characterized in the *hsp70* gene. In this case, the heat-shock element, when transferred to an unrelated gene, the non-heat-shock inducible thymidine kinase gene, conferred on it the ability to respond to a heat-shock stimulus (Pelham, 1982). Such DNA sequence elements in the promoters of tissue-specific genes play a critical role in producing their tissue-specific pattern of expression.

These tissue-specific elements are not confined to the promoters of genes; they may be found at great distances from the transcriptional start sites (Grosschedl and Birnstiel, 1980). Even at great distances and, in any orientation with respect to the transcriptional start site, these elements may affect the level of gene expression whether located upstream, downstream, or within the coding region. Although they lack promoter activity by themselves, these sequences act by increasing or decreasing the activity of a promoter, and hence are referred to as enhancers (reviewed in Muller *et al.*, 1988). Enhancers may increase the activity of a promoter in all cell types, or they may activate a particular promoter only in a select cell type (reviewed in Latchman, 1998). Enhancers usually contain multiple binding sites for transcription factors that cooperatively act to alter gene transcription (reviewed in Carey, 1998). These combinations of binding sites may be found in similarly regulated enhancers and promoters (co-regulation), and may also be present in multiple copies (e.g. Sen and Baltimore, 1986).

The balance between positive- and negative-acting transcription factors that bind to these regulatory regions determines the rate of the gene's transcription. One piece of the puzzle that effects this balance is the access of a transcription factor to its appropriate binding site. This is turn is affected by the manner in which that site is packaged in the

chromatin. A nucleosome, the fundamental unit of chromatin, consists of eight histone molecules around which the DNA wraps. Genes that are about to be transcribed undergo a reorganization of the chromatin (reviewed in Felsenfeld, 1996; Latchman, 1998). While the regulation of chromatin structure is necessary for proper gene expression, it is not sufficient. Distinct multiprotein complexes are needed to alter chromatin structure, to bind to promoters and enhancers, and to communicate between the activators and repressors (reviewed in Narlikar *et al.*, 2002). There are two classes of complexes that regulate the accessibility of the DNA to these various factors. The first class is ATP-dependent complexes that can move the nucleosome positions to expose or hide specific DNA sequences. The second class is those complexes that covalently modify the nucleosomes by adding or removing chemical moieties: acetylation, phosphorylation, and methylation of histone N-termini (reviewed in Narlikar *et al.*, 2002). One of the most studied chromatin-remodeling complexes that utilizes ATP hydrolysis is the SWI/SNF complex in yeast (reviewed in Pazin and Kadonaga, 1997; Tsukiyama and Wu, 1997). The most studied modification of the histone tail involves its acetylation, which in vitro has been shown to enhance accessibility of the DNA to restriction enzymes and transcription factors. There are several hypotheses as to why acetylation may have this effect. The first is that the lowered positive charge on the acetylated N-termini may cause a decrease in the stability of interaction with the DNA (Sewack *et al.*, 2001). The second is that the histone acetylation may decrease the compaction of the nucleosomes by interrupting the internucleosomal interactions made via the histone tails (Tse *et al.*, 1998). Finally, a third hypothesis is that these tail modifications might interact and physically recruit additional transcription factors (Strahl and Allis, 2000). Evidence

indicates that some transcription factors may bind directly to both ATP-dependent chromatin remodeling and histone acetyltransferase complexes, to "target" these activities to specific locations (reviewed in Narlikar *et al.*, 2002).

Gene transcription is initiated through the recruitment of RNA polymerase II (Pol II) to the promoters of target genes, the modification of nucleosomes, and the remodeling of chromatin. This occurs in conjunction with the assembly of multiple components of the basal transcription machinery, including the general transcription factors (GTFs) TFIIA, TFIIB, TFIID, TFIIE, TFIIF and TFIIH, and the transcriptional mediator complex (reviewed in Rachez and Freedman, 2001).

Transcriptional regulation in *C. elegans*

When *C. elegans* transcription is compared to other eukaryotic organisms, there are two major differences; the ability to *trans*-splice and the arrangement of some genes into operons (Krause and Hirsh, 1987; Zorio *et al.*, 1994). Many of the basics of the transcriptional machinery, like RNA polymerase II and the TATA-binding protein function, appear to be well conserved between *C. elegans* and other species (Bird and Riddle, 1989; Roberts *et al.*, 1987, 1989; Sanford *et al.*, 1983, 1985; Sanicola *et al.*, 1990; Dantonel, *et al.*, 2000; Vanfleteren and Van, 1983; Vanfleteren *et al.*, 1989). While the details of chromatin structure re-organization are not known, proteins like *dpy-27* belong to a family of chromosome-condensation proteins (Chuang *et al.*, 1994), and studies on dosage compensation have provided a link between chromatin structure and transcriptional activity (Meyer, 2000). Additionally, the complexes involved in nucleosome remodeling appear to have been conserved in *C. elegans*. For example, the

nucleosome remodeling and histone deacetylase (NURD) complex antagonizes vulval development (Solari and Ahringer, 2000), which is induced by the Ras signal transduction pathway (see discussion below). Inhibition of Ras signaling occurs in part through the action of the synthetic multivulval (synMuv) genes, which comprise two functionally redundant pathways (synMuvA and synMuvB) (Ferguson and Horvitz, 1989). The synMuvA and synMuvB pathways function redundantly to recruit or activate a core NURD complex, which has been hypothesized to repress vulval developmental target genes by local histone deacetylation (Solari and Ahringer, 2000).

The gene-specific function of the Mediator as an integrator of transcriptional regulatory signals between multiple inputs and the RNA Polymerase is conserved, and is essential for *C. elegans* development. RNA interference assays have shown that the CeMed6, CeMed7, and CeMed10/CeNut2 gene products form two mediator complexes, and both interact with Pol II via its largest subunit. These components are required *in vivo* for the transcriptional activation of several genes, including *ceh-13* and *nhr-2*, during specific stages of development in the worm, but are not required for the expression of two ubiquitously expressed genes, *rps-5* and *sur-5* (Kwon and Lee, 2001).

In addition, SOP-1/TRAP230 may be a Mediator target of pathways regulating transcriptional response to the Wnt pathway. Widely expressed *sop-1* appears to block action of the Wnt signal transduction pathway, suggesting that its effect must be relieved wherever the Wnt pathway acts (Zhang and Emmons, 2000).

Conservation of *trans*-acting transcriptional regulators in *C. elegans*

Most of the traditional transcription factor families have been identified and characterized in *C. elegans*. In the homeodomain superfamily, members of the HOX, POU, LIM, Paired, and NK subclasses have all been identified (Burglin *et al.*, 1991; Chisholm and Horvitz, 1995; Finney *et al.*, 1988; Herr *et al.*, 1988; Hobert *et al.*, 1998; Hunter and Kenyon, 1995; Okkema and Fire, 1994; Wang *et al.*, 1993; Way and Chalfie, 1988). The zinc finger family (including GATA family members), the helix-loop-helix family, the hormone receptor family, the forkhead family, the bzip family, the ETS family, and a variety of other families of transcription factors are all represented in *C. elegans* (Beitel *et al.*, 1995; Bowerman *et al.*, 1992; Kostrouch *et al.*, 1995; Krause *et al.*, 1990; Labouesse *et al.*, 1994; Miller *et al.*, 1993; Spieth *et al.*, 1991b; reviewed in McGhee and Krause, 1997).

Vulva cell specification and intracellular signaling pathways

Like more complicated organisms, *C. elegans* has a vulva connecting its uterus to the outside world to allow egg laying, and copulation with males. The development of this organ provides an excellent opportunity to study how cell-fate specification is controlled during development.

It is clear that pattern formation of the vulva involves the initiation, integration, and termination of many signals that work in concert to produce a final invariant lineage. In the *C. elegans* vulval ectoderm, at least three known intercellular signaling pathways, the inductive (EGF), lateral (NOTCH), and the WNT pathways, induce six multipotential Vulval Precursor Cells (VPCs) to generate an invariant spatial pattern of cell fates. These signaling pathways stimulate both the division of the VPC cells and the emergence of a

precise pattern (reviewed in Greenwald, 1997; Sternberg and Han, 1998). The VPCs are of three types: 1° and 2° VPCs, which can be distinguished by their division pattern and differential expression of marker genes, and 3° VPCs, which generate non-vulval epidermis (Burdine *et al.*, 1997; Greenwald, 1997; Kimble *et al.*, 1979; Sternberg and Horvitz, 1986; Sulston and Horvitz, 1977). The morphogenetic interactions of the 1° and 2° VPCs lead to the development of seven toroidal cells that connect the endothelium of the uterus to the external epithelium. These seven toroidal cells are the terminally differentiated VPCs: vulF, E, D, C, B2, B1, and A (Figure 1; Sharma-Kishore *et al.*, 1999).

The formation of competent multipotential cells is the first step in vulva formation. The twelve P cells that are present at hatching divide once; the anterior cells become neuroblasts and the posterior cells other than P3-P8.p fuse with the hypodermal syncytium in the L1 stage (Horvitz and Sternberg, 1991). Members of the homeotic gene family, the HOM-C gene cluster, are thought to play a critical role in establishing VPC competency (Clandinin *et al.*, 1997). In loss of function *lin-39* mutants, a Hom-C gene, P3-P8.p cells fuse with the hypodermal syncytium (Maloof and Kenyon, 1998). Since P3-P8.p cells have the ability to assume any of the vulval fates in response to an inductive signal LIN-3, all six cells must be competent to assume these cellular fates, and are considered developmentally equivalent (Katz *et al.*, 1995; Sternberg and Horvitz, 1986; Sulston and White, 1980). Therefore there is no strong intrinsic difference that pre-ordains the cells to a particular fate, and it does not appear as if cell fate specification in the vulva is dependent on some initial bias in competency. If it is not some initial bias

built into the cell that specifies the terminal cell fate, then there must be a mechanism that distinguishes the P3-P8.p cells such that an invariant lineage of cell fates is established.

We know that three signaling pathways, EGF, Notch and Wnt, play a critical role in specifying the cell fate of the Pn.p cells. In a canonical RAS signaling pathway, a growth factor stimulates a receptor tyrosine kinase (RTK) to activate Ras GTPase and the downstream kinases Raf, MEK, and MAP kinase/ERK, ultimately regulating the activities of transcription factors in the nucleus (reviewed in Sternberg and Alberola-Ila, 1998). In *C. elegans*, the receptor-tyrosine kinase LET-23 is stimulated by the growth factor ligand LIN-3 (Aroian *et al.*, 1990; Ferguson and Horvitz, 1985; Ferguson *et al.*, 1987; Hill and Sternberg, 1992; Horvitz and Sulston, 1980). The anchor cell (AC) serves as the source of the inductive signal, LIN-3 (Hill and Sternberg, 1992; Katz *et al.*, 1995; Kimble, 1981). Following stimulation of the RTK, LET-60 RAS activates the downstream kinases LIN-45 (RAF), MEK-2 (MAP kinase kinase) and MPK-1/SUR-1 (MAP kinase) (Church *et al.*, 1995; Han *et al.*, 1993; Kornfeld *et al.*, 1995; Lackner *et al.*, 1994; Wu and Han, 1994; Wu *et al.*, 1995), which ultimately alter the activities of transcription factors like LIN-1 (ETS), LIN-31 (a winged-helix transcription factor), and LIN-25 (a novel protein) (Beitel *et al.*, 1995; Miller *et al.*, 1993; Tan *et al.*, 1998; Tuck and Greenwald, 1995). There are many downstream positive regulators of *let-60 ras* signaling, including *ptp-2* (a SH2-containing protein tyrosine phosphatase), *ksr-1* (a novel protein kinase), *sur-6* (a subunit of the protein phosphatase 2A PPP2A-B), and *sur-8*/*soc-2* (a novel protein containing a leucine-rich repeat) (Gutch *et al.*, 1998; Kornfeld *et al.*, 1995; Sieburth *et al.*, 1998, 1999; Sundaram and Han, 1995). There are also several downstream negative regulators of EGF pathway, including the synthetic multivulva

genes (synMuv genes), *unc-101*, *sli-1*, *gap-1*, *ark-1* and *sur-5* (Beitel *et al.*, 1990; Clark *et al.*, 1994; Ferguson and Horvitz, 1985, 1989; Gu *et al.*, 1998; Hajnal *et al.*, 1997; Horvitz and Sulston, 1980; Hsieh *et al.*, 1999; Huang *et al.*, 1994; Jongeward *et al.*, 1995; Lee *et al.*, 1994; Lu and Horvitz, 1998; Solari and Ahringer, 2000; Thomas and Horvitz, 1999; Yoon *et al.*, 1995).

In the canonical model for Notch signaling, a number of proteolytic cleavages within NOTCH release the Notch intracellular domain (NICD) from the plasma membrane following ligand binding. This regulated intramembrane proteolysis allows NOTCH to function as a receptor in ligand binding, and also as a signal transducer, since the NICD translocates to the nucleus to directly interact with the DNA binding factor CSL (CBF-1, Suppressor of Hairless, LAG-1, also known as RBP-J) to regulate Notch target genes. In the absence of NICD, CSL acts as a transcriptional repressor (reviewed in Baron *et al.*, 2002). The existence of a lateral (NOTCH) signaling pathway in *C. elegans* vulva development between the VPCs was suggested of multivulva animals, in which all the VPCs adopt vulval fates independent of the inductive pathway (Sternberg, 1988). LIN-12/NOTCH appears to perform two functions during vulval induction that are separated by the phase of the VPC cell cycle (Ambros, 1999). Before completion of the S phase, LIN-12 is thought to inhibit the specification of the 1° fate and maintain the VPCs in an uncommitted state. After completion of the S phase, LIN-12 promotes the specification of the 2° fate. A*notch*-like mediated *lin-12* signal induces secondary fate (vulA, B1, B2, C, and D), and prevents any two adjacent VPCs from becoming primary (vulE and F; Sternberg, 1988; Sternberg and Horvitz, 1989). It was recently discovered that the MAP kinase phosphatase LIP-1 appears to mediate this lateral inhibition of the

primary fate (Berset *et al.*, 2001). MAP kinase phosphatases inactivate different types of MAP kinases by dephosphorylating the critical phosphotyrosine and phosphothreonine residues of the kinases (Camps *et al.*, 2000). LIP-1 is initially expressed at a low level in all VPCs. The inductive signal is thought to overcome this constitutive inhibition in P6.p to induce the 1° fate, whereas in P5.p and P7.p, LIN-12/NOTCH appears to up-regulate *lip-1* transcription, and this might inactivate MAP kinase and inhibit primary fate specification (Berset *et al.*, 2001). There are both positive regulators (*sup-17*, which encodes a metalloprotease of the ADAM family, and *sel-12*, which encodes presenilin), and negative regulators (*sel-1*, which encodes a novel extracellular protein, and *sel-10*, which encodes an F-box/WD40 repeat-containing protein) of this pathway (Grant and Greenwald, 1996; Hubbard *et al.*, 1997; Levitan and Greenwald, 1995; Sundaram and Greenwald, 1993; Tax *et al.*, 1997; Wen *et al.*, 1997).

The canonical Wnt pathway involves a WNT ligand that stimulates Frizzled (Fz) receptors to antagonize axin and GSK3 and stabilize β-catenin, ultimately regulating the activities of transcription factors of the TCF/LEF family (Cadigan and Nusse, 1997). In *C. elegans*, analysis of the WNT signaling mutants *bar-1* (a β-catenin-related protein) (Eisenmann *et al.*, 1998), *apr-1* (an APC-related protein) (Hoier *et al.*, 2000; Rocheleau *et al.*, 1997), and *mig-1* (which appears to function in many Wnt-mediated processes) (Eisenmann and Kim, 2000; Harris *et al.*, 1996; Thorpe *et al.*, 1997), shows that P4.p–P8.p can fuse instead of adopting the normal 1°, 2°, or 3° fates. Additionally, P5.p–P7.p can adopt the 3° fate instead of the 1° and 2° fates, resulting in too few VPCs adopting induced fates. Maintenance of the Hox gene *lin-39* in VPCs requires *bar-1* and *apr-1*, and cells that lose *lin-39* expression fuse (Eisenmann *et al.*, 1998; Hoier *et al.*,

2000). *lin-39* acts twice in vulval development, first in the L1 stage during generation of the VPCs (Clark *et al.*, 1993; Wang *et al.*, 1993), and later in the L3 stage during adoption of induced cell fates by the VPCs, when LIN-39 protein levels increase in response to activation of the RTK/Ras pathway (Clandinin *et al.*, 1997; Maloof and Kenyon, 1998). These results suggest that a Wnt pathway utilizing MIG-14, BAR-1, and APR-1 is active in the VPCs, and that one target of this pathway is *lin-39*. Hyperactivation of the Wnt pathway via a *pry-1* (axin homolog) (Korswagen *et al.*, 2002) loss-of-function mutation, or expression of an activated BAR-1 protein, leads to a Muv phenotype in which extra VPCs adopt induced cell fates (Gleason *et al.*, 2002). This indicates that *pry-1* may negatively regulate Wnt signaling in the VPCs, and that hyperactivation of the Wnt pathway may cause cells to adopt vulval fates that would not normally do so. However, the hyper-induced phenotype caused by Wnt pathway hyperactivation is not dependent on signaling through the Ras pathway (Gleason *et al.*, 2002).

In the final step of vulval development, the morphogenetic interactions of the primary and secondary VPCs, which migrate relative to their neighbors generate seven rings of toroidal cells (vulF, E, D, C, B2, B1, and A; Figure 1) that join the endothelium of the uterus to the external epithelium. The vulval muscles are attached to these rings, and specific cell attachments are made to lateral epithelial cells. Finally, the vulva partially everts to block the transit of eggs until it is opened by activation of the vulval muscles (Sharma-Kishore *et al.*, 1999). The genetics behind what drives these morphogenetic interactions is not well understood, and is currently being studied.

Historically, the only way to distinguish that a cell is terminally differentiated in the worm is by use of lineage analysis and observation of morphological changes. The advent of reporter constructs that reflect a particular cell type or fate is invaluable in figuring out cell fate specification, as well as cell termination mechanisms. We have several vulval cell fate-specific markers, that allow us to determine the identity of the vulva cells (Figure 2) (Burdine *et al.*, 1998; Struhl *et al.*, 1993; Williams-Masson *et al.*, 1998). Little is known about the individual roles of these vulva cells following their terminal differentiation, and what cell-specific functions they possess. Formation of the pattern of vulval cell types is likely to depend upon the *cis*-regulatory regions of the transcriptional targets of these intercellular signals in vulval development. The outcome of such differential activation will result in individual cell types. As in vulval development, we know few of the transcriptional regulators that control anchor cell gene expression. The isolation of response elements used by the anchor cell will facilitate biochemical and bioinformatic identification of major transcriptional factors that control cell-specific gene expression.

Genomic regulatory network analysis

It is not known how the inductive signal, lateral signal, and inhibitory signal are integrated on downstream targets resulting in an invariant pattern of cell-fate specification. However, because these signaling pathways are used elsewhere in the animal’s development, there must be a vulva-specific response mechanism. Additionally, since the same pathway appears to be used to specify multiple vulval cell fates, there may

be some branch in the pathway, or there may be key regulators that play a role in distinguishing these distinct fate specifications.

While a number of transcription factors are known to be involved in vulval development (e.g. *lin-1*, *lin-29, egl-38, lin-31*), little is known of their targets or interactions (Beitel *et al.*, 1995; Bettinger *et al.*, 1997; Chang *et al.*, 1999; Euling *et al.*, 1999; Tan *et al.*, 1998). The identification of *cis*-regulatory regions that confer cell specificity and respond to the inductive EGF pathway would be very helpful in determining such relationships. Three such target genes are: a fibroblast growth factor family member, *egl-17* (Figure 3; Burdine *et al.*, 1998); a FAT-like cadherin gene, *cdh-3* (Figure 4; Burdine *et al.*, 1998); and a zinc metalloproteinase gene, *zmp-1* (Figure 5; J. Butler and J. Kramer personal communication). These genes offer the opportunity to find response regions for multiple vulval cell types: vulE, F, C, D, and A, as well as the anchor cell. In addition, *egl-17* is an early cell-fate marker for the response to the inductive signal; the isolation of a *cis*–regulatory element that drives this early expression, and the identification of genes that regulate this expression, would be informative in determining the hierarchy of gene activation in this pathway.

egl-17 and the FGF family

The fibroblast growth factor receptor (FGFR) family plays a major role in how cells communicate with their environment. FGFR signaling is crucial for normal development, and its misregulation in human beings is linked to developmental abnormalities, and has been implicated in tumor progression. The cell-cell communication events mediated by

the FGFRs are used for the proper organization of cells into functional units during development (reviewed in Borland *et al.*, 2001).

In *C. elegans*, there are two putative FGFs, *egl-17* and *let-756*, and there is only one putative FGFR, *egl-15*. EGL-17 has been shown to be the instructive guidance cue in the attraction of a pair of bilaterally symmetric sex myoblasts (SMs: that express the EGL-15 FGFR) from the posterior of the animal to the their final positions flanking the precise center of the developing gonad (Branda and Stern, 2000a). The SMs then divide and differentiate into the muscles required for egg laying (Sulston and Horvitz, 1977). The loss of function mutation of *egl-17, e1313*, has a severe posterior displacement of hermaphrodite sex muscles due to the improper migration of the SMs (Burdine *et al.*, 1998). This displacement of the muscles disrupts the egg laying machinery, and causes the phenotypic bloating that is seen in some animals. In the vulva, *egl-17* is expressed in vulC and vulD as well as the presumptive vulE, and vulF cells (Figure 3). Besides vulva expression, *egl-17*::GFP is expressed in a variety of other tissue types (Burdine *et al.*, 1998). More recently, a reporter construct with an expanded upstream region of 10.5 kb showed additional expression that includes the dorsal uterine (DU) cells of the somatic gonad and, on rare occasions, weak expression was seen in the anchor cell and the ventral uterine cells (Branda and Stern, 2000b). This expanded region of expression has been shown to produce the gonadal attractive cue that could not be explained fully by the expression of EGL-17 in just the vulva cells; animals that do not have vulva cells due to genetic manipulation can position the SMs correctly. The expression in the descendants of P6.p is thought to play a redundant role in the positioning of the SMs. It has been hypothesized that the later expression of EGL-17 in vulC and vulD cells may play a role

in the precise positioning of the attachment of the vulva muscles between these two cells (Branda and Stern, 2000b).

zmp-1 and the Matrix Metalloproteinases

The Matrix Metalloproteinase Family, also called the Matrixins, is a family of zinc-dependent metalloendopeptides, which collectively are capable of degrading essentially all extracellular matrix components. This family has been shown to play critical roles in embryonic development, morphogenesis, reproduction, and tissue resorption and remodeling through the degradation of specific extracellular matrix components (reviewed in Matrisian, 2000). The expression of most matrixins is tightly regulated at the transcriptional level by growth factors, hormones, cytokines and cellular transformation (reviewed in Matrisian, 2000). Three genes encoding novel matrix metalloproteinases (MMPs) were recently identified and cloned by sequence similarity searching of the *Caenorhabditis elegans* genome database (Wada *et al.*, 1998). One of these three MMPs is *zmp-1*.

In *C. elegans*, a complete dissection of the expression pattern of the zinc metalloproteinase, *zmp-1*, has not been done. However, in hermaphrodites, in addition to vulA, vulE and anchor cell expression (Figure 4), it is expressed in a variety of other cell types from multiple lineages, including uterine and tail cells. The deletion of *zmp-1*, *cg115*, has no apparent phenotype and overexpression of ZMP-1 leads to a slight general degradation of the extracellular matrix components (J. Butler and J. Kramer, personal communication). While the role of this gene is unclear, it is interesting to note that at the time of ZMP-1 expression in the anchor cell, vulE and vulA there seem to be functional

rearrangements of the ECM, which must take place such that: the anchor cell can fuse with the vulF cells; vulE cells can attach to lateral epithelial seam cells; and the vulA cells can make junctions with the syncytial hypodermal cell, *hyp7*.

cdh-3 and the Cadherins

A third family of genes, the Cadherin superfamily of cell adhesion molecules, is involved in multiple morphogenetic events in animal development. Specifically, the Cadherin family plays a role in epithelial morphogenesis that is dependent upon coordinated control of changes in cell shape, proliferation, recognition and adhesion (reviewed in Tepass, 1999). It is a large family with many sub-groups that are divided by characteristic protein domains. Cadherin superfamily genes encode variable numbers of an extracellular domain termed the cadherin domain. These domains mediate intermolecular interactions and are dependent on calcium ions, which bind at sites between adjacent cadherin domains to produce a rigid structure. The extracellular domains are linked via a transmembrane helix to a cytoplasmic domain, which is known in some cases to interact with certain classes of intracellular proteins (reviewed in Tepass, 1999).

There are twelve predicted cadherin superfamily members in *C. elegans*. Of these, only *hmr-1* and *cdh-3* have been defined by experimental work on their structure and function (Hill *et al.*, 2001). CDH-3 is a member of the FAT-like cadherin sub-group. FAT-like cadherins are very large proteins with multiple cadherin domains, EGF-like, and laminin-AG domain repeats. It remains unclear whether the FAT-like cadherins operate in adhesion, signaling or both. The FAT-like cadherin family is predominantly expressed in epithelial cells (Hill *et al.*, 2001). In hermaphrodites, *cdh-3*::GFP is

expressed in the seam cells, the buccal and rectal epithelia, the excretory cell, two hypodermal cells in the tail, the uterine epithelium closest to the invaginating vulval cells followed by the multinucleated uterine seam cell (utse), the developing vulva, and associated neurons. Specifically, in the vulva, the reporter construct is expressed in vulA, E, F, C and D, as well as the anchor cell (Figure 5; Pettitt *et al.*, 1996). In *C. elegans* it is clear that CDH-3 is required for the morphogenesis of a single cell that forms the tip of the tail in the hermaphrodite. The other cells that express the *cdh-3* reporter appear to be unaffected by a probably null allele, raising the possibility that other genes can compensate for the loss of CDH-3 (Pettitt *et al.*, 1996). The genesis of the egg-laying system requires several sets of cell-cell recognition events, all of which occur during the expression of *cdh-3*::GFP. First, the anchor cell must invaginate between the two vulF cells, an event that takes place soon after GFP expression is observed in the cells involved. Second, the vulval epidermal cells must invaginate and form a connection with the uterus, and third the utse cell must make contacts with the seam cells. In addition, during the formation of the seven toroidal rings of the vulva, the vulva cells interact with one another (Pettitt *et al.*, 1996).

Regulatory analysis in *C. elegans*

A detailed analysis of *cis*-regulatory elements has been performed for only a few *C. elegans* genes. Like other multicellular organisms it appears that there are a variety of regulatory mechanisms. Genes, such as the vitellogenin gene *vit-2* (MacMorris *et al.*, 1992), the myosin gene *myo-2* (Okkema and Fire, 1994), the cuticle gene *dpy-7* (Gilleard *et al.*, 1997), the NK-2 homeobox gene *ceh-24* (Harfe and Fire, 1998), and the

acetylcholinesterase gene *ace-1* (Culetto *et al.*, 1999), are regulated in a relatively simple fashion by a tissue-specific basal promoter whose activity is enhanced by separate activator elements that can lie in the promoter, or within an intronic sequence (see discussion below). Other genes, such as the carboxylesterase gene *ges-1* (Egan *et al.*, 1995) and *mec-3* (Wang and Way, 1996b), require both activator and repressor elements to establish proper expression (see discussion below).

Upstream sequences of *dpy-7* were characterized in *C. elegans* by comparing the entire intergenic region to *C. briggsae* using a dot-matrix comparison. A single region of homology, 147 bp, was isolated. This corresponds with the minimal functional promoter region defined by deletion analysis in *C. elegans*. When 1kb of upstream sequences, and the *C. briggsae dpy-7* homolog were injected into a *dpy-7 C. elegans* strain, rescue was observed. Additionally, when two translational fusions of the *C. elegans dpy-7* gene (one with and one without the region of homology) were injected, only the translational fusion containing this region showed expression in *C. briggsae.* Contained in this conserved region is a predicted GATA site transcription factor, but no further experiments were performed to decipher a potential role for GATA factor transcription in the regulation of the *dpy-7* gene. These results provide evidence that regulated tissue- and stage-specific expression of *dpy-7* is achieved by a compact tissue-specific promoter element close to the 5′ end of the gene, and appears to involve no repressor elements (Gilleard *et al.*, 1997).

The myosin heavy chain *myo-2* gene contains at least two independent tissue-specific regulatory elements: a promoter sufficient for low-level expression in the pharyngeal muscle-specific expression is located near the transcriptional start site, and a

separable pharyngeal muscle-specific enhancer, 395 bp, located 300 bp upstream of the start site. This enhancer, which can induce pharyngeal muscle expression from a *myo-3::lacZ* fusion, involves at least three sub-elements that cooperate to activate transcription, two of which display distinct cell-type specificity (one for the whole pharynx, and two for a subset of pharyngeal cells). While individually, each of these subelements is inactive, any combination of two can drive transcription. Additionally, duplication of any of these elements is also sufficient to drive pharyngeal expression. Therefore, each of the subelements contains sufficient information to confer tissue-specific expression. Each subelement appears to contain multiple sites, as demonstrated by mutational analysis of each of these regions. Using a cDNA library, a *ceh-22* cDNA, which specifically binds one of the subelements, was identified (Okkema and Fire, 1994). Again, in this analysis, the transcriptional regulation of this gene appears to be regulated by multiple, discrete positive-acting elements. Subsequent studies have revealed that the organ-specific enhancer region contains a binding site for PHA-4 (Kalb *et al.*, 1998), a forkhead factor essential for pharyngeal development (Horner *et al.*, 1998; Kalb *et al.*, 1998; Mango *et al.*, 1994), and a binding site for DAF-3, which is a SMAD factor (Thatcher *et al.*, 1999). DAF-3, a negative regulator, is unlikely to modulate the organ specificity of this enhancer since a *daf-3* mutation does not affect the pharyngeal-specific expression pattern, or result in any pharyngeal defects, and may act to downregulate *myo-2* expression under as yet undescribed circumstances (Thatcher *et al.* 1999).

Similar experiments on the *ceh-24* upstream sequence revealed three distinct, separable tissue-specific enhancers for head neurons (57 bp), vulva muscles (48 bp) and the pharyngeal m8 cell (117 bp; Harfe and Fire, 1998).

The three previous examples demonstrate the relative simplicity of a handful of upstream *cis*-regulatory elements, which all act in a positive fashion to confer tissue-specific regulation. The following examples will show that not all promoters are as straightforward, and that, indeed, regulatory regions in *C. elegans* may contain both activator and repressor elements. Upon analysis of the carboxyesterase gene *ges-1*, it was shown that in particular deletions, it was expressed not in the gut (the E lineage, where normal expression is seen), but rather in muscle cells of the pharynx (which belong to a sister lineage of the gut, the MS lineage) and in body wall muscle and hypodermal cells (which belong to a cousin lineage of the gut). This 200-bp region responsible for the switch of expression from the E lineage to other lineages contains two binding sites for GATA factors, which have been subsequently shown to bind this sequence. Interestingly, when either of the two GATA sites or an adjacent sequence is eliminated, expression remains in the E lineage, but is restricted to a subset of cells, indicating that both of these sites are required for full expression in the gut. When any two of these three regions are eliminated, the switch to the MS lineage occurs and, when all three are eliminated, the vast majority of expression in all tissues is lost. These observations suggest that gut-specific gene expression in *C. elegans* involves not only gut-specific activators, but also multiple repressors that are present in particular non-gut lineages (Aamodt *et al.*, 1991; Egan *et al.*, 1995; Kennedy *et al.*, 1993). Subsequent studies have proposed a model in which the normal E lineage gut expression of *ges-1* is controlled by the gut-specific GATA factor such as ELT-2, while the pharynx and rectum (MS lineage) expression is controlled by PHA-4, which is normally bound to the *ges-1* 3′ enhancer sequences. The

activation of PHA-4 is kept repressed by an unknown factor binding in the vicinity of the GATA factor binding sites (Marshall and McGhee, 2001).

The 10 neurons involved in mechanosensation in *C. elegans* express *mec-3*. The expression is maintained by autoregulation. Four conserved regions, each of 24-70 bp, were identified by intraspecies comparisons to *C. vulgarensis*. The downstream region (528 bp), which includes conserved blocks I, II and III, appear to mediate establishment of the expression pattern. An additional, more distal element (917 bp), also appears sufficient to establish *mec-3* expression. Mutations in region I, III and IV can all cause transient ectopic expression of the *mec-3::lacZ* fusions in some sister cells of the normal *mec-3* expressing cells. UNC-86 binding sites have been identified in conserved regions I, II and III of the 5' flanking sequence. (In an *unc-86* background, the cells that normally express *mec-3* are not specified to the correct terminal fate). However, it seems unlikely that the binding sites for UNC-86 are the sole players in this very complex upstream region (Wang and Way, 1996a,b; Way and Chalfie, 1988; Way *et al.*, 1991; Xue *et al.*, 1992, 1993).

Although *cis*-regulatory analysis has been preformed on only a handful of upstream regions in *C. elegans*, it has been suggested that the complex regulation, particularly involving repressor elements, might be a general feature of transcriptional control in those genes expressed prior to cellular differentiation (Krause *et al.*, 1994). Genes that encode abundant structural proteins may be regulated in a simpler manner (Gilleard *et al.*, 1997). This simplicity may be an important feature of the transcription of large multigene families, or of genes that are transcribed following cellular differentiation. However, the DAF-3 binding studies on the *myo-2* enhancer serve as a

cautionary reminder that expression studies examine only one set of conditions. Under different conditions, repressor or activator activity may be utilized. They also demonstrated that in *C. elegans*, there are enhancers that function in all cell types of a tissue, and that these elements are not mutually exclusive from those that act in a distinct subtype of cells in this same tissue (Thatcher *et al.*, 1999). The *ceh-24* studies delineate that multiple modules, all apparently positive acting, may regulate tissue specificity in a variety of tissues that are not related by lineage (Harfe and Fire, 1998). These are just some of the complexities of transcriptional regulation in *C. elegans* that have been revealed to us so far. In other model organisms, such simplicity is almost unheard of. Which begs the question, "Is transcriptional regulation in *C. elegans* just that much simpler, or are we just not in deep enough to reveal all the layers of complexity that are seen in these other systems?"

An example of the complexity seen in other systems is the regulation of CD4 gene silencing expression during T-cell development. When three copies of the murine silencer were linked to a CAT reporter vector regulated by one of the CD4 enhancers and the CD4 promoter, expression of CAT was specifically repressed in CD4-CD8+, but not in CD4+CD8+ T cells. Using this system as an assay, a core 134 bp fragment was defined, which in triplicate reduced transcription 10-to 20-fold. This core silencer worked better than the larger fragment defined in transfection studies, but it had no silencing activity in transgenic mice. When flanking 5' or 3' sequences were added back to this core fragment, silencer activity was restored in the transgenic constructs. This functional redundancy of the flanking sequences in animals, and their dispensability in transient transfection studies, suggest that these flanking sequences contain elements needed for organizing the

chromatin structure to allow access of *trans*-acting factors to the silencing elements. When internal deletions were made in the core region, one of three outcomes was observed: (1) silencing, (2) no silencing, or (3) a variegation of silencing. The variegation suggested that, in many cases, the loss of a single nuclear factor binding site would not completely inactivate the silencer, but would decrease the probability of the establishment of silencing. A conclusion from these studies is that what may appear to be crucial, the 134-bp core fragment, may not be the whole story of elements involved in a gene's native transcriptional regulation. In addition, this is just one region that plays a role in CD4 gene transcription: two enhancers, a core promoter, and at least one other element in an intron have been implicated in the fidelity of the expression pattern (review in Ellmeier *et al.*, 1999).

Dissection of co-regulated genes

A common assumption in the modeling of genetic regulatory networks is that the cell-specific genes expressed in a given terminally differentiated cell type are likely to be subject to coordinate control, and hence possess similar upstream *cis*-acting sequences (Davidson, 2001). While some attempts to validate this assumption in *C. elegans* have failed, other studies have succeeded. A comparison of the cuticle gene *dpy-7*'s 5' flanking sequences with other *C. elegans* cuticle genes did not reveal any striking regions of similarity (Gilleard *et al.*, 1997). A dot-matrix comparison of two acetylcholinesterase genes, *ace-1* and *ace-2*, failed to show any similarities between the two promoters (Culetto *et al.*, 1999). And the comparison of *C. elegans* MyoD family member *hlh-1* to

mouse myogenic regulatory factors presented no striking similarities between these promoters (Krause *et al.*, 1994).

One success story is that of the vitellogenin genes. There are six *C. elegans* vitellogenin genes that are subject to sex-, stage-, and tissue-specific regulation: they are expressed solely in the adult hermaphrodite intestine. Comparative sequence analysis of upstream sequences of these genes and their *C. briggsae* homologs revealed the presence of two repeated heptameric elements, vit promoter element 1 (VPE1) and VPE2. A functional analysis of the VPEs within the 5'-flanking region of the vit-2 gene revealed that a 247 bp element containing the VPEs was sufficient for high-level, regulated expression. Furthermore, none of the four deletion mutations resulted in inappropriate expression (Blumenthal *et al.*, 1984; Spieth *et al.*, 1985, 1991a; Zucker-Aprison and Blumenthal, 1989).

Since every cell in the worm may have a unique identity at the molecular level, the use of a battery of cell type-specific markers might allow the identification of any common upstream element(s) responsible for driving expression in a specific cell or cell type. Indications that this type of analysis might work in *C. elegans* have started to appear. A comparison of the minimal promoters of *mtl-1* and *mtl-2* to other *C. elegans* intestinal cell-specific genes identified repeats of GATA transcription factor-binding sites. Mutation analyses determined that GATA elements are required for transcription, while electrophoretic mobility shift assays showed that ELT-2, a *C. elegans* GATA transcription factor, specifically binds these element. Furthermore, when *elt-2* is disrupted in *C. elegans*, *mtl-2* is not expressed. It was also shown that ectopic expression of ELT-2 can activate transcription of *mlt-2* in non-intestinal cells of *C. elegans*. These

results suggest that the binding of ELT-2 to GATA elements in these promoters regulates tissue-specific transcription of the *C. elegans* metallothionein genes (Moilanen *et al.*, 1999).

Another success story was the *C. elegans* gene *daf-19*, which encodes an RFX-type transcription factor that is expressed specifically in all ciliated sensory neurons (Swoboda *et al.*, 2000). Loss of *daf-19* function causes the absence of cilia, resulting in sensory defects. Twenty *C. elegans* promoters of genes that are expressed in ciliated sensory neurons were searched for X boxes. (X boxes are the mammalian targets for RFX-type transcription factors.) Target sites were found within the promoters of four of these genes, *che-2*, *daf-19*, *osm-1* and *osm-6*, which are expressed in most or all ciliated sensory neurons. Target sites were not found in the promoter regions of any of the genes that are expressed in only a subset of ciliated sensory neurons, e.g., *gcy-5*, *gcy-8* and *gcy-32*. Using an in vivo assay, it was shown that expression of the X box-containing genes was dependent on both *daf-19* function and the presence of the promoter X box. In a genome-wide search for X-box-containing genes, a novel gene was examined and found to be expressed in ciliated sensory neurons in a *daf-19*-dependent manner. These data suggest that *daf-19* is a transcriptional regulator of gene products that function broadly in sensory cilia (Swoboda *et al.*, 2000). To date, there are no studies that have looked at the co-regulation of genes at the cell-specific, rather than tissue-specific, level.

One of the fallbacks of this type of analysis is that assumptions have to be made on what genes may constitute a group of co-regulated genes. Groupings of co-regulated genes based on family function are not necessarily going to lead to the identification of a common element(s). The advent of microarray analysis and SAGE techniques will make

the determination of cohorts of co-regulated genes easier to identify. In a recent study, the expression pattern of 11,917 genes from *C. elegans* were monitored using microarrays to determine which of these genes was upregulated in response to heat-shock treatment. The upstream regions of the 28 genes that appeared to be upregulated by greater than four fold in response to heat-shock were examined using several computational and statistical methods. The resulting two heat-shock elements (HSE) were conserved in the upstream regions of the *C. briggsae* orthologs of the *C. elegans* genes. Upon mutational analysis of the *hsp-16-2*::GFP, these elements were found to be neither necessary nor sufficient, but did have an effect on the strength of the GFP expression, indicating that this type of element may be hard to isolate using the traditional experimental methods such as systematic deletion (GuhaThakurta *et al.*, 2002). In another recent study, *C. elegans* touch-receptor cells were cultured and used for microarray analysis. The culturing of these cells enabled the sensitivity of the microarray data to be increased, so that *mec-3*-dependent genes could be identified (there are only six touch-receptor cells in the worm). Using the 5' regions of genes that were significantly enriched in this analysis, Zhang *et al.* were able to determine that a heptanucleotide element was over-represented in this population (Zhang *et al.*, 2002). However, the functional significance of this element has not been shown. These are the first steps in a very promising future of experiments. The isolation of subpopulations of cells and microarray analysis will allow the identification of overrepresented upstream elements that are specific to a cellular function, or a specific cell type. However, what this technology does not ensure is the identification of all the important sequences involved in the fidelity of the expression pattern.

Phylogenetic footprinting

With whole genome sequences becoming readily available, and with the failure of de novo computational programs to recognize functional motifs in *cis*-regulatory regions (Loots *et al.*, 2000; Pennacchio and Rubin, 2001), there is a growing interest in comparing genome sequences to identify regulatory regions (Stojanovic *et al.*, 1999). Phylogenetic footprinting is a method for the identification of regulatory elements in a set of orthologous regulatory regions from multiple species; it does so by identifying the best-conserved motifs in those orthologous regions (Tagle *et al.*, 1988).

To see the real power of this technique, examine the studies performed on the human *epsilon-globin* gene, which undergoes dramatic changes in transcriptional activity during development. Elucidation of the mechanisms that govern these interactions could suggest strategies to reactivate fetal (*gamma*) or embryonic (*epsilon*) genes in individuals with severe hemoglobinopathies. The expression pattern of the *epsilon-globin* gene is conserved in all placental mammals. The *epsilon-globin* sequences from seven mammalian species- human, orangutan, gibbon, capuchin, monkey, galago, and rabbit- were used to compare the upstream regulatory regions of this gene. The total number of evolutionary years included in such an alignment is additive. Since the evolutionary time of these species is greater than 270 million years, nucleotide sequences have had ample time to accumulate changes. Twenty-one conserved elements were identified in the 2 kb of sequence immediately upstream of the coding region of the epsilon gene. Probes spanning each of these footprints bound proteins in gel-shift assays. Among the 47 binding interactions characterized were: eight sites for the yin and yang 1 (a protein

shown to have both activator and repressor properties); five binding sites for a putative stage-selective protein SSP; and seven sites for an as-yet-unidentified protein (Gumucio *et al.*, 1993). Such studies allow for an unbiased selection of factors involved in the transcriptional regulation of this gene, which speaks neither to the sufficiency nor the necessity of the individual factors, but rather to a more global picture of the milieu of the elements and factors involved.

For this type of analysis to be fruitful, the genomes that are used must be selected carefully. Comparison with too-closely related genome will reveal shared conservation in non-functional areas. However, if the comparison is performed on a species that is too-distantly related, the genomes will likely lack the conservation needed to be informative. Studies in bacteria and animals have suggested that a slightly less-diverged species is a better choice when looking for the conservation of *cis*-regulatory elements (Cargill *et al.*, 1999; Huynen and Bork, 1998).

Despite having diverged from each other an estimated 50-120 million years ago (Coghlan, 2002), both *C. elegans* and *C. briggsae* share almost identical development and morphology (Nigon and Dougherty, 1949). Cross-species rescue of mutant phenotypes has demonstrated that there is functional conservation between the two species (Culetto *et al.*, 1999; de Bono and Hodgkin, 1996; Kennedy *et al.*, 1993; Krause *et al.*, 1994; Kuwabara, 1996; Maduro and Pilgrim, 1996). This should not be taken to mean that all homologs will function and be expressed in a similar fashion between the two species. For instance, at least one aspect of the *hlh-1* gene's regulation, a homolog of the MyoD family of myogenic regulatory factors, differs between the two species. The *C. elegans hlh-1* is expressed in the MS-granddaughter cells during embryogenesis, while this

expression is not detected by *lacZ* reporter constructs and antibody staining in *C. briggsae* (Krause *et al.*, 1994). Despite this, the two almost completely sequenced genomes make *C. briggsae* an obvious choice for genome comparisons to *C. elegans*. The analysis of similarity within 142 pairs of orthologous intergenic regions shows regions of high similarity interspersed with non-alignable sequence (Webb *et al.*, 2002). The high degree of similarity in some of these regions suggests that they have undergone selective pressure. Such intergenic conservation between *C. elegans* and *C. briggsae* has been utilized in a handful of studies to isolate putative binding sites for *trans*-acting regulatory factors.

Upstream sequences from *ace-1* were compared to the orthologous *C. briggsae* gene by dot -matrix comparison. This analysis revealed four blocks (35, 58, 140 and 409 bp) of conserved sequence. These blocks were between 70-80% identical between species. The first block contained splicing site sequences and alternative splice-sites, indicating that this region was probably part of the minimal promoter. (Interestingly, it is devoid of TATA and CAAT boxes.) To test whether the other conserved sequences could qualitatively modulate the basal activity of the promoter, a CAT reporter gene expression system in mammalian cell lines was used. Two of the conserved blocks did not affect transcriptional activity, whereas one block in this system acted as a transcriptional repressor. However, in expression studies, the block that was found to repress CAT reporter gene expression was involved in driving expression in the body wall and anal muscle cells, and the two blocks that did not effect expression levels were also required for expression in other areas of the animal. Additionally, the conserved region that appeared to be a repressor in the CAT system, when combined with the minimal

promoter element, was sufficient to drive expression in body wall and anal muscles. These data suggest that *cis*-regulatory sequences of *C. elegans* are not recognized in the same way as in the transcriptional apparatus of the mouse cells. Intra-species comparisons with *C. briggsae* were able to identify the important *cis*-regulatory regions of this gene, but were unable to isolate distinct factor binding sites (Culetto *et al.*, 1999)

In *ceh-24* upstream sequences, *C. briggsae* was used in a species comparison to confirm the importance of a pair of NdE-boxes and the m8 pharyngeal cell enhancer. Intra-species comparison did not reveal any additional binding sites (Harfe and Fire, 1998).

Studies of the gut esterase gene, *ges-1* (discussed above), illuminate the benefits and risks of intra-species comparison studies between *C. elegans* and *C. briggsae*. A 17-bp region of conservation between the *C. elegans* and *C. briggsae* 5' flanking sequences was found, but deletion of this element had no effect on the expression pattern of the reporter transgene (Egan *et al.*, 1995). It is likely that not all conserved sequences between these two species will have a functional significance. On the other hand, an important binding site located in the 3' flanking regions of the coding sequence of this gene was identified using the comparison between these two species. This binding site, critical to the regulation of the *ges-1* gene in the pharynx and rectum, had not been found by conventional deletion analysis (Marshall and McGhee, 2001).

Thesis overview

In chapter one of this thesis, I analyze the *cis*-regulatory sequence regions sufficient to confer vulva cell- and anchor cell- specific expression of three putatively co-regulated

genes: *zmp-1*, *egl-17* and *cdh-3*. These genes are expressed in a restricted and overlapping expression pattern in specific vulva cell types and the uterine anchor cell within *C. elegans*. We chose these genes because their function is not required for the normal development of the cells in which they are expressed, and hence they lie downstream of the cell-fate-specification pathways.

In chapter two, I used an orthogonal approach to isolate vulva- and anchor cell-specific elements. I have identified the *C. briggsae* homologs of these three genes and used phylogenetic footprinting to identify the predicted control regions corresponding to the sufficiency regions identified in *C. elegans*. Together, these two approaches elucidate similar elements that are sufficient to confer expression to a subset of vulval cells and the uterine anchor cell.

REFERENCES

Aamodt, E., Chung, M., and McGhee, J. (1991). Spatial control of gut-specific gene expression during *Caenorhabditis elegans* development. *Science* **252,** 579-582.

Ambros, V. (1999). Cell cycle-dependent sequencing of cell fate decisions in *Caenorhabditis elegans* vulva precursor cells. *Development* **126,** 1947-1956.

Aroian, R., Koga, M., Mendel, J., Ohshima, Y., and Sternberg, P. (1990). The *let-23* gene necessary for *Caenorhabditis elegans* vulval induction encodes a tyrosine kinase of the EGF receptor subfamily. *Nature* **348,** 693-699.

Baron, M., Aslam, H., Flasza, M., Fostier, M., Higgs, J., Mazaleyrat, S., and Wilkin, M. (2002). Multiple levels of Notch signal regulation (review). *Mol Membr Biol* **19,** 27-38.

Beitel, G., Clark, S., and Horvitz, H. (1990). *Caenorhabditis elegans* ras gene *let-60* acts as a switch in the pathway of vulval induction. *Nature* **348,** 503-509.

Beitel, G., Tuck, S., Greenwald, I., and Horvitz, H. (1995). The *Caenorhabditis elegans* gene *lin-1* encodes an ETS-domain protein and defines a branch of the vulval induction pathway. *Genes & Development* **9,** 3149-3162.

Berset, T., Hoier, E., Battu, G., Canevascini, S., and Hajnal, A. (2001). Notch inhibition of Ras signaling through MAP kinase phosphatase LIP-1 during *C. elegans* vulval development. *Science* **291,** 1055-1058.

Bettinger, J., Euling, S., and Rougvie, A. (1997). The terminal differentiation factor LIN-29 is required for proper vulval morphogenesis and egg laying in *Caenorhabditis elegans*. *Development* **124,** 4333-4342.

Bird, D., and Riddle, D. (1989). Molecular cloning and sequencing of *ama-1*, the gene encoding the largest subunit of *Caenorhabditis elegans* RNA polymerase II. *Molecular and Cellular Biology* **9,** 4119-4130.

Blumenthal, T., Squire, M., Kirtland, S., Cane, J., Donegan, M., Spieth, J., and Sharrock, W. (1984). Cloning of a yolk protein gene family from *C. elegans*. *Journal of Molecular Biology* **174,** 1-18.

Borland, C., Schutzman, J., and Stern, M. (2001). Fibroblast growth factor signaling in *Caenorhabditis elegans*. *Bioessays* **23,** 1120-1130.

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

Branda, C., and Stern, M. (2000). Mechanisms controlling sex myoblast migration in *Caenorhabditis elegans* hermaphrodites. *Developmental Biology* **226,** 137-151.

Burdine, R., Branda, C., and Stern, M. (1998). EGL-17(FGF) expression coordinates the attraction of the migrating sex myoblasts with vulval induction in *C. elegans*. *Development* **125,** 1083-1093.

Burdine, R., Chen, E., Kwok, S., and Stern, M. (1997). *egl-17* encodes an invertebrate fibroblast growth factor family member required specifically for sex myoblast migration in *Caenorhabditis elegans*. *Proceedings of the National Academy of Sciences USA* **94,** 2433-2437.

Cadigan, K., and Nusse, R. (1997). Wnt signaling: a common theme in animal development. *Genes & Development* **11,** 3286-3305.

Camps, M., Nichols, A., and Arkinstall, S. (2000). Dual specificity phosphatases: a gene family for control of MAP kinase function. *FASEB J* **14,** 6-16.

Carey, M. (1998). The enhanceosome and transcriptional synergy. *Cell* **92,** 5-8.

Cargill, M., Altshuler, D., Ireland, J., Sklar, P., Ardlie, K., Patil, N., Shaw, N., Lane, C., Lim, E., Kalyanaraman, N., Nemesh, J., Ziaugra, L., Friedland, L., Rolfe, A., Warrington, J., Lipshutz, R., Daley, G., and Lander, E. (1999). Characterization of single-nucleotide polymorphisms in coding regions of human genes. *Nat Genet* **22,** 231-238.

Chang, C., Newman, A., and Sternberg, P. (1999). Reciprocal EGF signaling back to the uterus from the induced *C. elegans* vulva coordinates morphogenesis of epithelia. *Current Biology* **9,** 237-246.

Chuang, P.-T., Albertson, D., and Meyer, B. (1994). DPY-27: A chromosome condensation protein homolog that regulates *C. elegans* dosage compensation through association with the X chromosome. *Cell* **79,** 459-474.

Church, D., Guan, K., and Lambie, E. (1995). Three genes of the MAP kinase cascade, *mek-2*, *mpk-1/sur-1* and *let-60* ras, are required for meiotic cell cycle progression in *Caenorhabditis elegans*. *Development* **121,** 2525-2535.

Clandinin, T., Katz, W., and Sternberg, P. (1997). *Caenorhabditis elegans* HOM-C genes regulate the response of vulval precursor cells to inductive signal. *Developmental Biology* **182,** 150-161.

Clark, S., Chisholm, A., and Horvitz, H. (1993). Control of cell fates in the central body region of *C. elegans* by the homeobox gene *lin-39*. *Cell* **74,** 43-55.

Clark, S., Lu, W., and Horvitz, H. (1994). The *Caenorhabditis elegans* locus *lin-15*, a negative regulator of a tyrosine kinase signaling pathway, encodes two different proteins. *Genetics* **137,** 987-997.

Coghlan, A. W. K. (2002). Fourfold faster rate of genome rearrangement in nematodes than in *Drosophila*. *Genomic Research* **12,** 857-867.

Culetto, E., Combes, D., Fedon, Y., Roig, A., Toutant, J., 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*. *Journal of Molecular Biology* **290,** 951-966.

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

Davidson, E. H. (2001). "*Genomic Regulatory Systems: Development and Evolution*." Academic Press,

de Bono, M., and Hodgkin, J. (1996). Evolution of sex determination in *Caenorhabditis*: Unusually high divergence of *tra-1* and its functional consequences. *Genetics* **144,** 587-595.

Egan, C., Chung, M., Allen, F., Heschl, M., Vanbuski, C., and McGhee, J. (1995). A gut-to-pharynx/tail switch in embryonic expression of the *Caenorhabditis elegans ges-1* gene centers on two GATA sequences. *Developmental Biology* **170,** 397-419.

Eisenmann, D., and Kim, S. (2000). Protruding vulva mutants identify novel loci and Wnt signaling factors that function during *Caenorhabditis elegans* vulva development. *Genetics* **156,** 1097-1116.

Eisenmann, D., Maloof, J., Simske, J., Kenyon, C., and Kim, S. (1998). The B-catenin homolog BAR-1 and LET-60 Ras coordinately regulate the Hox gene *lin-39* during *Caenorhabditis elegans* vulval development. *Development* **125,** 3667-3680.

Ellmeier, W., Sawada, S., and Littman, D. (1999). The regulation of CD4 and CD8 coreceptor gene expression during T cell development. *Annual Review of Immunology* **17,** 523-554.

Euling, S., Bettinger, J., and Rougvie, A. (1999). The LIN-29 transcription factor is required for proper morphogenesis of the *Caenorhabditis elegans* male tail. *Developmental Biology* **206,** 142-156.

Felsenfeld, G. (1996). Chromatin unfolds. *Cell* **86,** 13-19.

Ferguson, E., and Horvitz, H. (1985). Identification and characterization of 22 genes that affect the vulval cell lineages of the nematode *C. elegans*. *Genetics* **110,** 17-72.

Ferguson, E., and Horvitz, H. (1989). The multivulva phenotype of certain *Caenorhabditis elegans* mutants results from defects in two functionally redundant pathways. *Genetics* **123,** 109-121.

Ferguson, E., Sternberg, P., and Horvitz, H. (1987). A genetic pathway for the specification of the vulval cell lineages of *C. elegans*. *Nature* **326,** 259-267.

Gilleard, J., Barry, J., and Johnstone, I. (1997). *cis* regulatory requirements for hypodermal cell-specific expression of the *Caenorhabditis elegans* cuticle collagen gene *dpy-7*. *Molecular and Cellular Biology* **17,** 2301-2311.

Gleason, J., Korswagen, H., and Eisenmann, D. (2002). Activation of Wnt signaling bypasses the requirement for RTK/Ras signaling during *C. elegans* vulval induction. *Genes Dev* **16,** 1281-1290.

Grant, B., and Greenwald, I. (1996). The *Caenorhabditis elegans sel-1* gene, a negative regulator of *lin-12* and *glp-1*, encodes a predicted extracellular protein. *Genetics* **143,** 237-247.

Greenwald, I. (1997). Development of the Vulva in: *"C. elegans II." DL Riddle, T Blumenthal, BJ Meyer and JR Priess (eds), Cold Spring Harbor Laboratory Press.* **II,** 519-541.

Grosschedl, R., and Birnstiel, M. (1980). Identification of regulatory sequences in the prelude sequences of an H2A histone gene by the study of specific deletion mutants in vivo. *Proc Natl Acad Sci U S A* **77,** 1432-1436.

Gu, T., Orita, S., and Han, M. (1998). *Caenorhabditis elegans* SUR-5, a novel but conserved protein, negatively regulates LET-60 ras activity during vulval induction. *Molecular and Cellular Biology* **18,** 4556-4564.

GuhaThakurta, D., Palomar, L., Stormo, G., Tedesco, P., Johnson, T., Walker, D., Lithgow, G., Kim, S., and Link, C. (2002). Identification of a novel *cis*-regulatory element involved in the heat shock response in *Caenorhabditis elegans* using microarray gene expression and computational methods. *Genome Res* **12,** 701-712.

Gumucio, D., Shelton, D., Bailey, W., Slightom, J., and Goodman, M. (1993). Phylogenetic footprinting reveals unexpected complexity in *trans* factor binding upstream from the *epsilon-globin* gene. *Proc Natl Acad Sci U S A* **90,** 6018-6022.

Gutch, M., Flint, A., Keller, J., Tonks, N., and Hengartner, M. (1998). The *Caenorhabditis elegans* SH2 domain-containing protein tyrosine phosphatase

PTP-2 participates in signal transduction during oogenesis and vulval development. *Genes & Development* **12,** 571-585.

Hajnal, A., Whitfield, C., and Kim, S. (1997). Inhibition of *Caenorhabditis elegans* vulval induction by *gap-1* and by *let-23* receptor tyrosine kinase. *Genes & Development* **11,** 2715-2728.

Han, M., Golden, A., Han, Y., and Sternberg, P. (1993). *C. elegans lin-45* raf gene participates in *let-60* RAS-stimulated vulval differentiation. *Nature* **363,** 133-140.

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

Harris, J., Honigberg, L., Robinson, N., and Kenyon, C. (1996). Neuronal cell migration in *C. elegans*: regulation of Hox gene expression and cell position. *Development* **122,** 3117-3131.

Hill, E., Broadbent, I., Chothia, C., and Pettitt, J. (2001). Cadherin superfamily proteins in *Caenorhabditis elegans* and *Drosophila melanogaster*. *Journal of Molecular Biology* **305,** 1011-1024.

Hill, R., and Sternberg, P. (1992). The gene *lin-3* encodes an inductive signal for vulval development in *C. elegans*. *Nature* **358,** 470-476.

Hoier, E., Mohler, W., Kim, S., and Hajnal, A. (2000). The *Caenorhabditis elegans* APC-related gene *apr-1* is required for epithelial cell migration and Hox gene expression. *Genes & Development* **14,** 874-886.

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

Horvitz, H., and Sternberg, P. (1991). Multiple intercellular signaling systems control the development of the *Caenorhabditis elegans* vulva. *Nature* **351,** 535-541.

Horvitz, H., and Sulston, J. (1980). Isolation and genetic characterization of cell-lineage mutants of the nematode *C. elegans*. *Genetics* **96,** 435-454.

Hsieh, J., Liu, J., Kostas, S., Chang, C., Sternberg, P., and Fire, A. (1999). The RING finger/B-box factor TAM-1 and a retinoblastoma-like protein LIN-35 modulate context-dependent gene silencing in *Caenorhabditis elegans*. *Genes & Development* **13,** 2958-2970.

Huang, L., Tzou, P., and Sternberg, P. (1994). The *lin-15* locus encodes two negative regulators of *Caenorhabditis elegans* vulval development. *Molecular Biology of the Cell* **5,** 395-411.

Hubbard, E., Wu, G., Kitajewski, J., and Greenwald, I. (1997). *sel-10*, a negative regulator of *lin-12* activity in *Caenorhabditis elegans*, encodes a member of the CDC4 family of proteins. *Genes & Development* **11,** 3182-3193.

Huynen, M., and Bork, P. (1998). Measuring genome evolution. *Proc Natl Acad Sci USA* **95,** 5849-5856.

Jongeward, G., Clandinin, T., and Sternberg, P. (1995). *sli-1*, a negative regulator of *let-23*-mediated signaling in *C. elegans*. *Genetics* **139,** 1553-1566.

Kalb, J., Lau, K., Goszczynski, B., Fukushige, T., Moons, D., Okkema, P., and McGhee, J. (1998). *pha-4* is Ce-fkh-1, a fork head/HNF-3a,B,y homolog that functions in organogenesis of the *C. elegans* pharynx. *Development* **125,** 2171-2180.

Katz, W., Hill, R., Clandinin, T., and Sternberg, P. (1995). Different levels of the *C. elegans* growth factor LIN-3 promote distinct vulval precursor fates. *Cell* **82,** 297-307.

Kennedy, B., Aamodt, E., Allen, F., Chung, M., Heschl, M., and McGhee, J. (1993). The gut esterase gene (ges-1) from the nematodes *Caenorhabditis elegans* and *Caenorhabditis briggsae*. *Journal of Molecular Biology* **229,** 890-908.

Kimble, J. (1981). Alterations in cell lineage following laser ablation of cells in the somatic gonad of *C. elegans*. *Developmental Biology* **87,** 286-300.

Kimble, J., Sulston, J., and White, J. (1979). Regulative development in the post-embryonic lineages of *C. elegans* in: *"Cell Lineage, Stem Cells and Cell Determinations." Le Douarin N (ed), Elsevier, NY.***,** 59-68.

Kornfeld, K., Hom, D., and Horvitz, H. (1995). The *ksr-1* gene encodes a novel protein kinase involved in Ras-mediated signaling in *C. elegans*. *Cell* **83,** 903-913.

Korswagen, H., Coudreuse, D., Betist, M., van de Water, S., Zivkovic, D., and Clevers, H. (2002). The Axin-like protein PRY-1 is a negative regulator of a canonical Wnt pathway in *C. elegans*. *Genes Dev* **16,** 1291-12302.

Kostrouch, Z., Kostrouchova, M., and Rall, J. (1995). Steroid/thyroid hormone receptor genes in *Caenorhabditis elegans*. *Proceedings of the National Academy of Sciences USA* **92,** 156-159.

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

Krause, M., Harrison, S., Xu, S., Chen, L., and Fire, A. (1994). Elements regulating cell- and stage-specific expression of the *C. elegans* myoD family homolog *hlh-1*. *Developmental Biology* **166,** 133-148.

Krause, M., and Hirsh, D. (1987). A *trans*-spliced leader sequence on actin mRNA in *C. elegans*. *Cell* **49,** 753-761.

Kuwabara, P. (1996). Interspecies comparison reveals evolution of control regions in the nematode sex-determining gene *tra-2*. *Genetics* **144,** 597-607.

Kwon, J., and Lee, J. (2001). Biological significance of a universally conserved transcription mediator in metazoan developmental signaling pathways. *Development* **128,** 3095-3104.

Labouesse, M., Sookhare, S., and Horvitz, H. (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.

Lackner, M., Kornfeld, K., Miller, L., Horvitz, H., and Kim, S. (1994). A MAP kinase homolog, *mpk-1*, is involved in ras-mediated induction of vulval cell fates in *Caenorhabditis elegans*. *Genes & Development* **8,** 160-173.

Latchman, D. S. (1998). "*Eukaryotic Transcription Factors*." Academic Press.

Lee, J., Jongeward, G., and Sternberg, P. (1994). *unc-101*, a gene required for many aspects of *Caenorhabditis elegans* development and behavior, encodes a clathrin-associated protein. *Genes & Development* **8,** 60-73.

Levitan, D., and Greenwald, I. (1995). Facilitation of *lin-12*-mediated signalling by *sel-12*, a *Caenorhabditis elegans* S182 Alzheimer's disease gene. *Nature* **377,** 351-354.

Matrisian, LM. (2000). Quick guide Matrix Metalloproteinases. *Curr Biol* **10,** R692.

Loots, G., Locksley, R., Blankespoor, C., Wang, Z., Miller, W., Rubin, E., and Frazer, K. (2000). Identification of a coordinate regulator of interleukins 4, 13, and 5 by cross-species sequence comparisons. *Science* **288,** 136-140.

Lu, X., and Horvitz, H. (1998). *lin-35* and *lin-53*, Two genes that antagonize a *C. elegans* Ras pathway, encode proteins similar to Rb and its binding protein RbAp48. *Cell* **95,** 981-991.

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. *Molecular and Cellular Biology* **12,** 1652-1662.

Maduro, M., and Pilgrim, D. (1996). Conservation of function and expression of *unc-119* from two *Caenorhabditis* species despite divergence of non-coding DNA. *Gene* **183,** 77-85.

Maloof, J., and Kenyon, C. (1998). The Hox gene *lin-39* is required during *C. elegans* vulval induction to select the outcome of Ras signaling. *Development* **125,** 181-190.

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

Marshall, S., and McGhee, J. (2001). Coordination of *ges-1* expression between the *Caenorhabditis* pharynx and intestine. *Developmental Biology* **239,** 350-363.

McGhee, J., and Krause, M. (1997). Transcription Factors and Transcriptional Regulation in: *"C. elegans II." DL Riddle, T Blumenthal, BJ Meyer and JR Priess (eds), Cold Spring Harbor Laboratory Press.* **II,** 147-184.

McKnight, S., and Tjian, R. (1986). Transcriptional selectivity of viral genes in mammalian cells. *Cell 1986* **46,** 795-805.

Meyer, B. (2000). Sex in the worm-counting and compensating X-chromosome dose. *Trends in Genetics* **16,** 247-253.

Miller, L., Gallegos, M., Morisseau, B., and Kim, S. (1993). *lin-31*, a *Caenorhabditis elegans* HNF-3/fork head transcription factor homolog, specifies three alternative cell fates in vulval development. *Genes & Development* **7,** 933-947.

Moilanen, L., Fukushige, T., and Freedman, J. (1999). Regulation of metallothionein gene transcription-Identification of upstream regulatory elements and transcription factors responsible for cell-specific expression of the metallothionein genes from *C. elegans*. *Journal of Biological Chemistry* **274,** 29655-29665.

Muller, M., Gerster, T., and Schaffner, W. (1988). Enhancer sequences and the regulation of gene transcription. *Eur J Biochem* **176,** 485-495.

Narlikar, G., Fan, H., and Kingston, R. (2002). Cooperation between complexes that regulate chromatin structure and transcription. *Cell* **108,** 475-487.

Nigon, V., and Dougherty, E. (1949). Reproduction patterns and attempts at reciprocal crossing of *Rhabditis elegans* Maupas, 1900, and *Rhabditis briggsae* Dougherty & Nigon, 1949 (Nematoda: Rhabditidae). *Journal of Experimental Zoology* **112,** 485-503.

Okkema, P., 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.

Pazin, M., and Kadonaga, J. (1997). What's up and down with histone deacetylation and transcription? *Cell* **89,** 325-328.

Pelham, H. (1982). A regulatory upstream promoter element in the *Drosophila hsp 70* heat-shock gene. *Cell* **30,** 517-528.

Pennacchio, L., and Rubin, E. (2001). Genomic strategies to identify mammalian regulatory sequences. *Nat Rev Genet* **2,** 100-109.

Pettitt, J., Wood, W., and Plasterk, R. (1996). *cdh-3*, a gene encoding a member of the cadherin superfamily, functions in epithelial cell morphogenesis in *Caenorhabditis elegans*. *Development* **122,** 4149-4157.

Rachez, C., and Freedman, L. (2001). Mediator complexes and transcription. *Curr Opin Cell Biol* **13,** 274-280.

Riddle, D., Blumenthal, T., Meyer, B., and Priess, J. (1997). Introduction to *C. elegans* in: *"C. elegans II" DL Riddle, T Blumenthal, BJ Meyer and JR Priess (eds), Cold Spring Harbor Laboratory Press.* **II,** 1-22.

Roberts, S., Emmons, S., and Childs, G. (1989). Nucleotide sequences of *Caenorhabditis elegans* core histone genes. Genes for different histone classes share common flanking sequence elements. *Journal of Molecular Biology* **206,** 567-577.

Roberts, S., Sanicola, M., Emmons, S., and Childs, G. (1987). Molecular characterization of the histone gene family of *C. elegans*. *Journal of Molecular Biology* **196,** 27-38.

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

Sanford, T., Golomb, M., and Riddle, D. (1983). RNA polymerase II from wild type and alpha-amanitin-resistant strains of *C. elegans*. *Journal of Biological Chemistry* **258,** 12804-12809.

Sanford, T., Prenger, J., and Golomb, M. (1985). Purification and immunological analysis of RNA polymerase II from *C. elegans*. *Journal of Biological Chemistry* **260,** 8064-8069.

Sanicola, M., Ward, S., Childs, G., and Emmons, S. (1990). Identification of a *Caenorhabditis elegans* histone H1 gene family. Characterization of a family member containing an intron and encoding a poly(A)+ mRNA. *Journal of Molecular Biology* **212,** 259-268.

Sen, R., and Baltimore, D. (1986). Multiple nuclear factors interact with the immunoglobulin enhancer sequences. *Cell* **46,** 705-716.

Sentenac, A. (1985). Eukaryotic RNA polymerases. *CRC Crit Rev Biochem* **18,** 31-90.

Sewack, G., Ellis, T., and Hansen, U. (2001). Binding of TATA binding protein to a naturally positioned nucleosome is facilitated by histone acetylation. *Mol Cell Biol* **21,** 1404-1415.

Sharma-Kishore, R., White, J., Southgate, E., and Podbilewicz, B. (1999). Formation of the vulva in *Caenorhabditis elegans*: a paradigm for organogenesis. *Development* **126,** 691-699.

Sieburth, D., Sun, Q., and Han, M. (1998). SUR-8, a conserved Ras-binding protein with leucine-rich repeats, positively regulates Ras-mediated signaling in *C. elegans*. *Cell* **94,** 119-130.

Sieburth, D., Sundaram, M., Howard, R., and Han, M. (1999). A PP2A regulatory subunit positively regulates Ras-mediated signaling during *Caenorhabditis elegans* vulval induction. *Genes & Development* **13,** 2562-2569.

Solari, F., and Ahringer, J. (2000). NURD-complex genes antagonise Ras-induced vulval development in *Caenorhabditis elegans*. *Current Biology* **10,** 223-226.

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 Research* **13,** 5283-5295.

Spieth, J., Nettleton, M., Zucker-Aprison, E., Lea, K., and Blumenthal, T. (1991a). Vitellogenin motifs conserved in nematodes and vertebrates. *Journal of Molecular Evolution* **32,** 429-438.

Spieth, J., Shim, Y., Lea, K., Conrad, R., and Blumenthal, T. (1991b). *elt-1*, an embryonically expressed *Caenorhabditis elegans* gene homologous to the GATA transcription factor family. *Molecular and Cellular Biology* **11,** 4651-4659.

Sternberg, P. (1988). Lateral inhibition during vulval induction in *Caenorhabditis elegans*. *Nature* **335,** 551-554.

Sternberg, P., and Alberola-Ila, J. (1998). Conspiracy theory: RAS and RAF do not act alone. *Cell* **95,** 447-450.

Sternberg, P., and Han, M. (1998). Genetics of RAS signaling in *C. elegans*. *Trends in Genetics* **14,** 466-472.

Sternberg, P., and Horvitz, H. (1986). Pattern formation during vulval development in *C. elegans*. *Cell* **44,** 761-772.

Sternberg, P., and Horvitz, H. (1989). The combined action of two intercellular signaling pathways specifies three cell fates during vulval induction in *C. elegans*. *Cell* **58,** 679-693.

Stojanovic, N., Florea, L., Riemer, C., Gumucio, D., Slightom, J., Goodman, M., Miller, W., and Hardison, R. (1999). Comparison of five methods for finding conserved sequences in multiple alignments of gene regulatory regions. *Nucleic Acids Res* **27,** 3899-38910.

Strahl, B., and Allis, C. (2000). The language of covalent histone modifications. *Nature* **403,** 41-45.

Struhl, G., Fitzgerald, K., and Greenwald, I. (1993). Intrinsic activity of the *lin-12* and notch intracellular domains in vivo. *Cell* **74,** 331-345.

Sulston, J., and Horvitz, H. (1977). Post-embryonic cell lineages of the nematode, *Caenorhabditis elegans*. *Developmental Biology* **56,** 110-156.

Sulston, J., and White, J. (1980). Regulation and cell autonomy during postembryonic development of *C. elegans*. *Developmental Biology* **78,** 577-597.

Sundaram, M., and Greenwald, I. (1993). Genetic and phenotypic studies of hypomorphic *lin-12* mutants in *Caenorhabditis elegans*. *Genetics* **135,** 755-763.

Sundaram, M., and Han, M. (1995). The *C. elegans ksr-1* gene encodes a novel Raf-related kinase involved in Ras-mediated signal transduction. *Cell* **83,** 889-901.

Swoboda, P., Adler, H., and Thomas, J. (2000). The RFX-type transcription factor DAF-19 regulates sensory neuron cilium formation in *C. elegans*. *Molecular Cell* **5,** 411-421.

Tagle, D., Koop, B., Goodman, M., Slightom, J., Hess, D., and Jones, R. (1988). Embryonic *epsilon* and *gamma globin* genes of a prosimian primate (*Galago crassicaudatus*). Nucleotide and amino acid sequences, developmental regulation and phylogenetic footprints. *Journal Molecular Biology* **3,** 439-455.

Tan, P., Lackner, M., and Kim, S. (1998). MAP kinase signaling specificity mediated by the LIN-1 Ets/LIN-31 WH transcription factor complex during *C. elegans* vulval induction. *Cell* **93,** 569-580.

Tax, F., Thomas, J., Ferguson, E., and Horvitz, H. (1997). Identification and characterization of genes that interact with *lin-12* in *Caenorhabditis elegans*. *Genetics* **147,** 1675-1695.

Tepass, U. (1999). Genetic analysis of cadherin function in animal morphogenesis. *Curr Opin Cell Biol* **11,** 540-548.

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

Thomas, J., and Horvitz, H. (1999). The *C. elegans* gene *lin-36* acts cell autonomously in the *lin-35* Rb pathway. *Development* **126,** 3449-3459.

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

Tse, C., Sera, T., Wolffe, A., and Hansen, J. (1998). Disruption of higher-order folding by core histone acetylation dramatically enhances transcription of nucleosomal arrays by RNA polymerase III. *Mol Cell Biol* **18,** 4629-4638.

Tsukiyama, T., and Wu, C. (1997). Chromatin remodeling and transcription. *Curr Opin Genet Dev* **7,** 182-191.

Tuck, S., and Greenwald, I. (1995). *lin-25*, a gene required for vulval induction in *Caenorhabditis elegans*. *Genes & Development* **9,** 341-357.

Vanfleteren, J., and Van, B. J. (1983). Nematode chromosomal proteins-III. Some structural properties of the histones of *C. elegans*. *Comparative Biochemistry & Physiology* **76B,** 179-184.

Vanfleteren, J., Van, B. S., and Van, B. J. (1989). The histones of *Caenorhabditis elegans*: no evidence of stage-specific isoforms. *FEBS Letters* **257,** 233-237.

Wada, K., Sato, H., Kinoh, H., Kajita, M., Yamamoto, H., and Seiki, M. (1998). Cloning of three *Caenorhabditis elegans* genes potentially encoding novel matrix metalloproteinases. *Gene* **211,** 57-62.

Wang, B., Muller-Immergluck, M., Austin, J., Robinson, N., Chisholm, A., and Kenyon, C. (1993). A homeotic gene-cluster patterns the anteroposterior body axis of *C. elegans*. *Cell* **74,** 29-42.

Wang, L., and Way, J. (1996a). Activation of the mec-3 promoter in two classes of stereotyped lineages in *Caenorhabditis elegans*. *Mechanisms of Development* **56,** 165-181.

Wang, L., and Way, J. (1996b). Promoter sequences for the establishment of *mec-3* expression in the nematode *Caenorhabditis elegans*. *Mechanisms of Development* **56,** 183-196.

Way, J., 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.

Way, J., Wang, L., Run, J., and Wang, A. (1991). The *mec-3* gene contains *cis*-acting elements mediating positive and negative regulation in cells produced by

asymmetric cell division in *Caenorhabditis elegans*. *Genes & Development* **5,** 2199-2211.

Webb, C., Shabalina, S., Ogurtsov, A., and Kondrashov, A. (2002). Analysis of similarity within 142 pairs of orthologous intergenic regions of *Caenorhabditis elegans* and *Caenorhabditis briggsae*. *Nucleic Acids Res* **30,** 1233-1239.

Wen, C., Metzstein, M., and Greenwald, I. (1997). SUP-17, a *Caenorhabditis elegans* ADAM protein related to *Drosophila* KUZBANIAN, and its role in LIN-12/NOTCH signaling. *Development* **124,** 4759-4767.

Williams-Masson, E., Heid, P., Lavin, C., and Hardin, J. (1998). The cellular mechanism of epithelial rearrangement during morphogenesis of the *Caenorhabditis elegans* dorsal hypodermis. *Developmental Biology* **204,** 263-276.

Wu, Y., and Han, M. (1994). Suppression of activated Let-60 Ras protein defines a role of *Caenorhabditis elegans* Sur-1 MAP kinase in vulval differentiation. *Genes & Development* **8,** 147-159.

Wu, Y., Han, M., and Guan, K. (1995). MEK-2, a *Caenorhabditis elegans* MAP kinase kinase, functions in Ras-mediated vulval induction and other developmental events. *Genes & Development* **9,** 742-755.

Xue, D., Finney, M., Ruvkun, G., and Chalfie, M. (1992). Regulation of the *mec-3* gene by the *C. elegans* homeoproteins UNC-86 and MEC-3. *EMBO Journal* **11,** 4969-4979.

Xue, D., Tu, Y., and Chalfie, M. (1993). Cooperative interactions between the *Caenorhabditis elegans* homeoproteins UNC-86 and MEC-3. *Science* **261,** 1324-1328.

Yoon, C., Lee, J., Jongeward, G., and Sternberg, P. (1995). Similarity of *sli-1*, a regulator of vulval development in *C. elegans*, to the mammalian proto-oncogene c-cbl. *Science* **269,** 1102-1105.

Zhang, H., and Emmons, S. (2000). A *C. elegans* mediator protein confers regulatory selectivity on lineage-specific expression of a transcription factor gene. *Genes & Development* **14,** 2161-2172.

Zhang, Y., Ma, C., Delohery, T., Nasipak, B., Foat, B., Bounoutas, A., Bussemaker, H., Kim, S., and Chalfie, M. (2002). Identification of genes expressed in *C. elegans* touch receptor neurons. *Nature* **418,** 331-335.

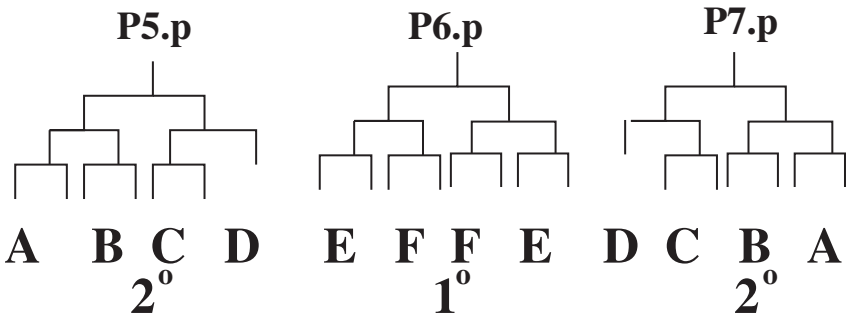
Zorio, D., Cheng, N., Blumenthal, T., and Spieth, J. (1994). Operons as a common form of chromosomal organization in C. elegans. *Nature* **372,** 270-272.

Zucker-Aprison, E., and Blumenthal, T. (1989). Potential regulatory elements of nematode vitellogenin genes revealed by interspecies sequence comparison. *Journal of Molecular Evolution* **28,** 487-496.

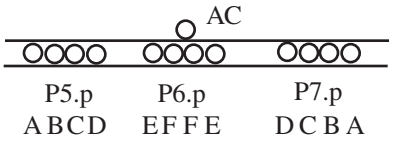
Figure 1: Vulva formation in *C. elegans*

The top panel shows the lineage relationship of P5, 6 and 7.p descendents that give rise to the vulva. In the bottom panels, nuclei are indicated by circles, and prominent cell boundaries are indicated by thin lines. The ventral surface is down in all panels, and the dark horizontal line represents the ventral cuticle. Since animals were typically observed from the side, different focal planes correspond to the midline (top panel), the sublateral plane (middle panel) and the lateral plane (bottom panel). A three-dimensional schematic is shown to the right. "A, B1, B2..." correspond to "vulA, vulB1, vulB2...".

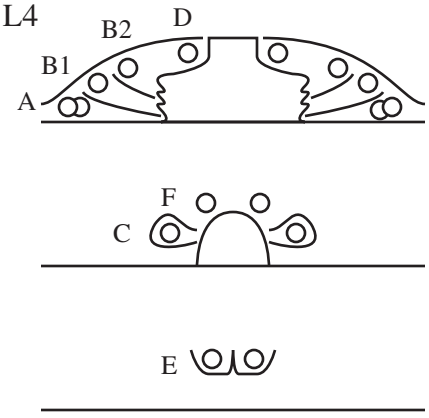
Figure 1: Vulva formation in *C. elegans*

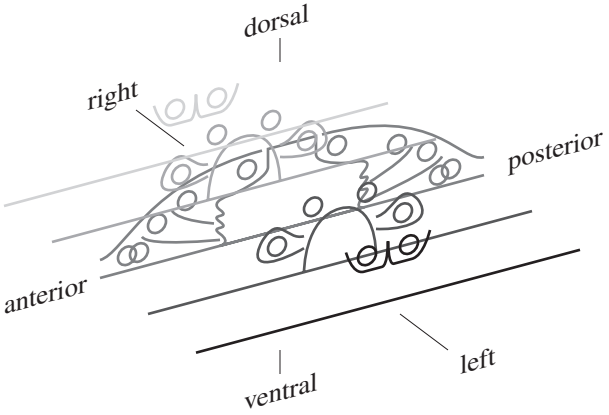


late L3

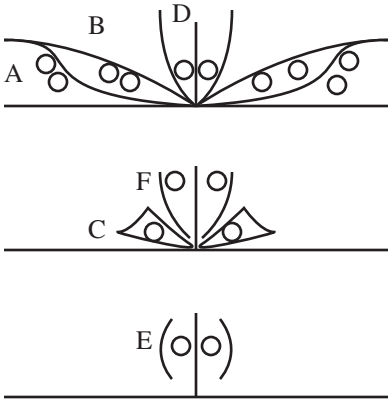


mid L4





adult



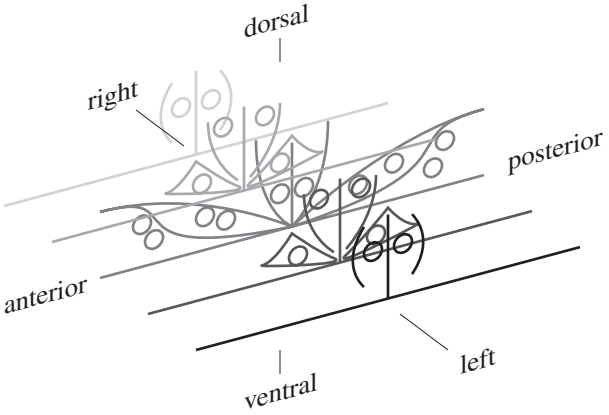
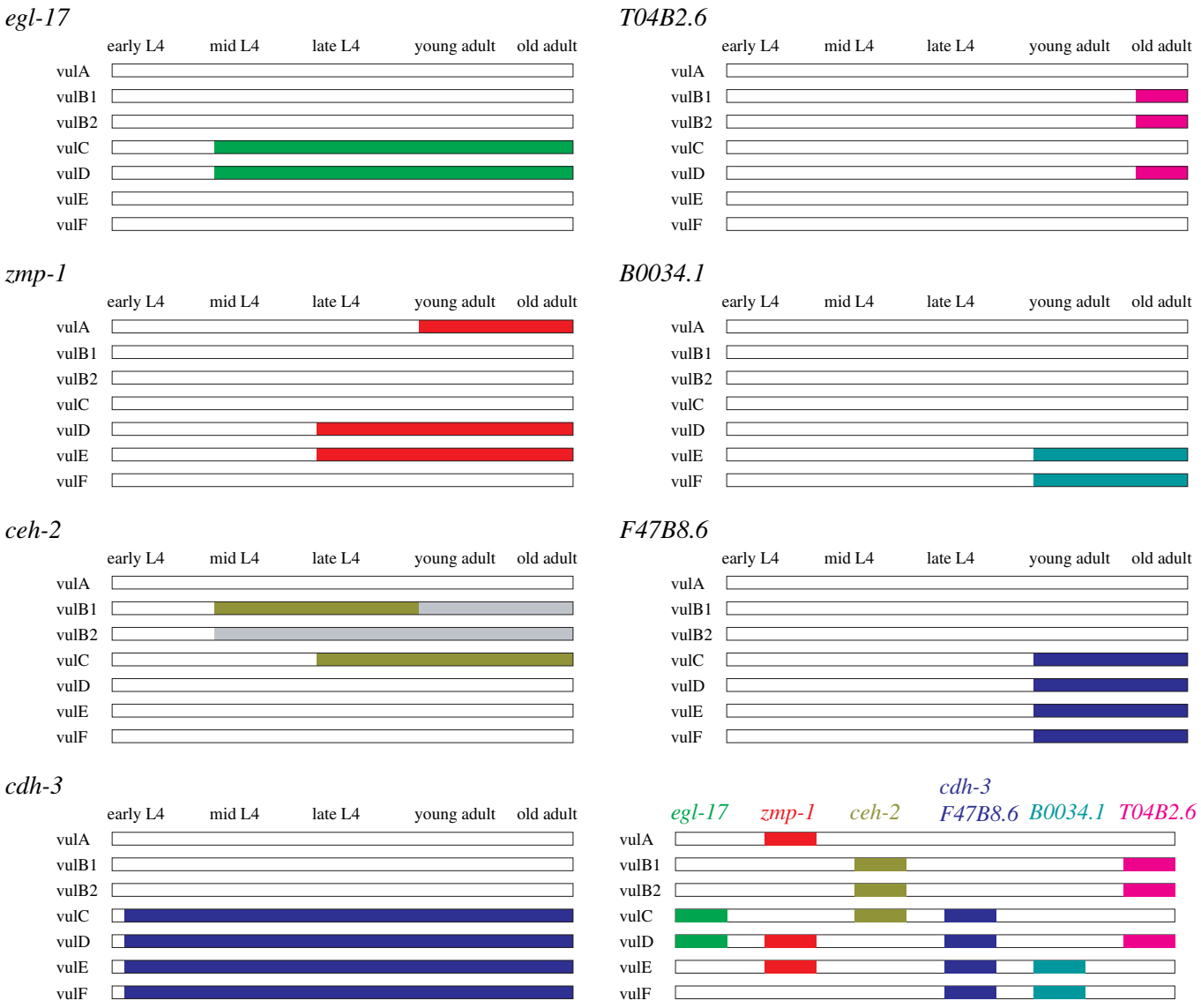


Figure 2: Available vulval marker gene's expression pattern in *C. elegans*

Filled bars indicate consistent expression observed in all animals, gray bars indicate expression observed in some but not all animals. The last round of cell division in the vulva takes place within the first one or two hours of the L4 stage. *egl-17::gfp* is also expressed earlier in the parents and grandparents of vulE and vulF cells, P6.p progeny, (Burdine *et al*., 1998) (not shown). This expression occasionally persists into the L4 stage in some lines. The expression of *T04B2.6::gfp* is observed in old adults (animals with a significant number of eggs in the gonad) but not in young adults (animals without eggs in the gonad immediately after the L4 molt). The last panel is a side-by-side comparison of markers disregarding the temporal aspect, demonstrating that six different cell types can be distinguished based on the expression pattern.

Figure 2: Available vulva marker genes expression pattern in *C. elegans*

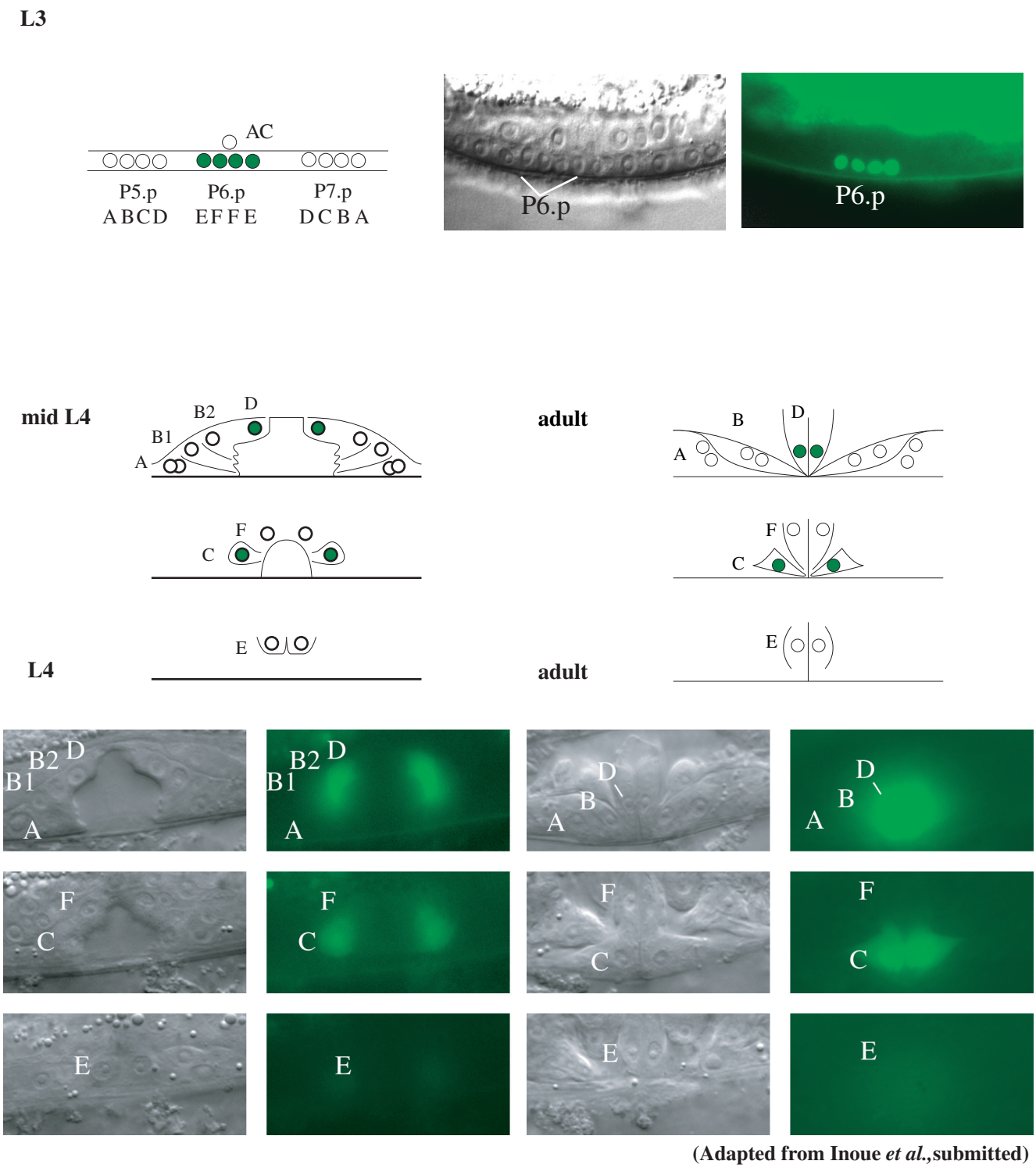


from Inoue *et al.*, submitted

Figure 3: *egl-17::*GFP

A schematic diagram of cell positions at various stages of development, late L3, mid-L4 and adult are shown (Sharma-Kishore *et al*., 1999; Sulston and Horvitz, 1977). Nuclei are indicated by circles, and prominent cell boundaries are indicated by thin lines. The green filled-in circles depict the GFP expressing cells. The ventral is down in all panels, and the dark horizontal line represents the ventral cuticle. Each set of Nomarski images to the left have corresponding epifluorescence images to the right. The top panels are from the midline, the middle panels are from the sub-lateral plane, and the bottom panels are from the lateral plane (L3 animals were only photographed in midline plane). The strain and array photographed is MT2466 *ayIs4[egl-17::gfp]*.

Figure 3: *egl-17::gfp*



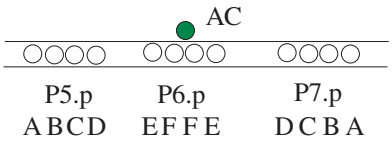
(Adapted from Inoue *et al.*,submitted)

Figure 4: *zmp-1::*GFP

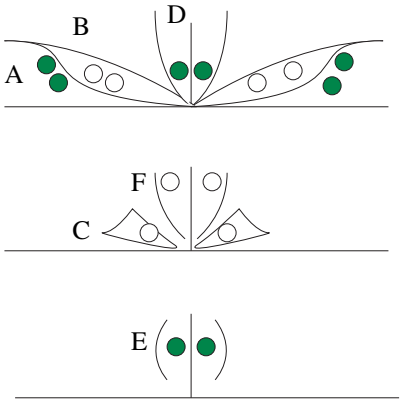
A schematic diagram of cell positions at various stages of development, late L3, and adult stages are shown (Sharma-Kishore *et al*., 1999; Sulston and Horvitz, 1977). Nuclei are indicated by circles, and prominent cell boundaries are indicated by thin lines. The green filled-in circles depict the GFP expressing cells. The ventral is down in all panels, and the dark horizontal line represents the ventral cuticle. Each set of Nomarski images to the left have corresponding epifluorescence images to the right. In the photomicrographs of the adult animals, the top panels are from the midline, the middle panels are from the sub-lateral plane, and the bottom panels are from the lateral plane. The strain and array photographed is PS3239 *syIs49[zmp-1::gfp]*.

Figure 4: *zmp-1::gfp*

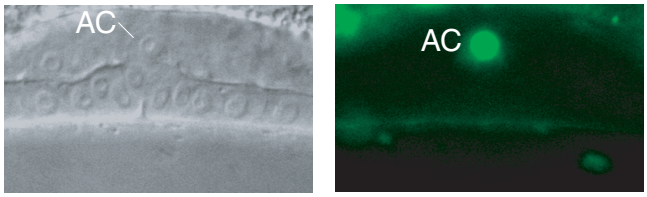
L3



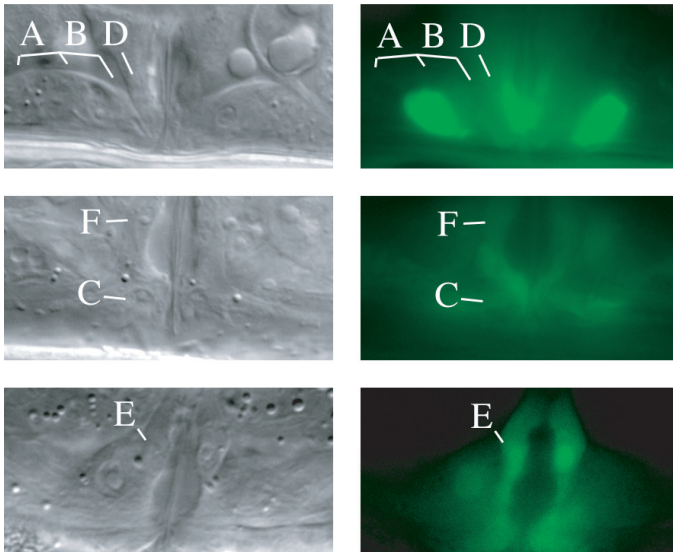
adult



L3



adult



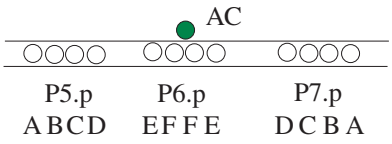
(Adapted from Inoue *et al.*,submitted)

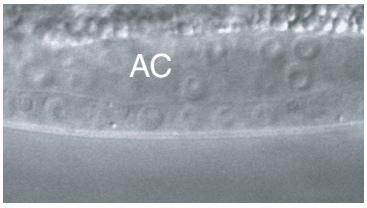
Figure 5: *cdh-3::*GFP

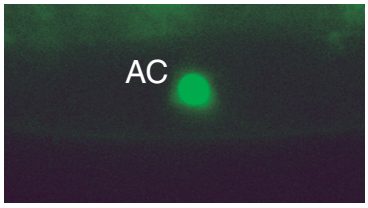
A schematic diagram of cell positions at various stages of development, late L3, mid-L4 and adult stages are shown (Sharma-Kishore *et al*., 1999; Sulston and Horvitz, 1977). Nuclei are indicated by circles, and prominent cell boundaries are indicated by thin lines. The green filled-in circles depict the GFP expressing cells. The ventral is down in all panels, and the dark horizontal line represents the ventral cuticle. Each set of Nomarski images to the left have corresponding epifluorescence images to the right. In the photomicrographs of the mid-l4 and adult animals, the top panels are from the midline, the middle panels are from the sub-lateral plane, and the bottom panels are from the lateral plane. In the photomicrographs of the mid-L4 and adult animals, the *cdh-3* is also expressed along with *ceh-3* in vulC cells (looks yellow in epifluorescence photomicrographs). The strain and array photographed is PS3528 *syIs51[cdh-3::cfp]; syIs55[ceh-2::gfp]* (for the mid-L4 and adult animals) and NL1008 *pkEx246[cdh-3::gfp]* (for anchor cell expression).

Figure 5: *cdh-3::gfp*

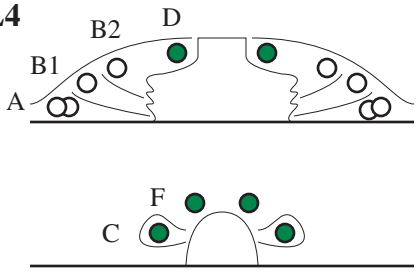
L3



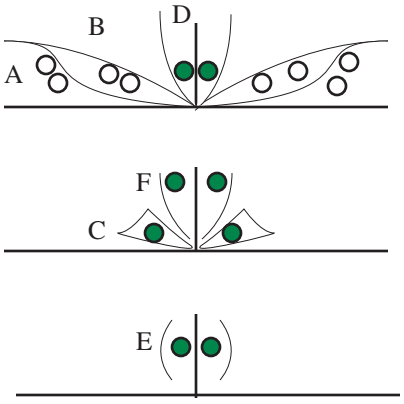




mid L4

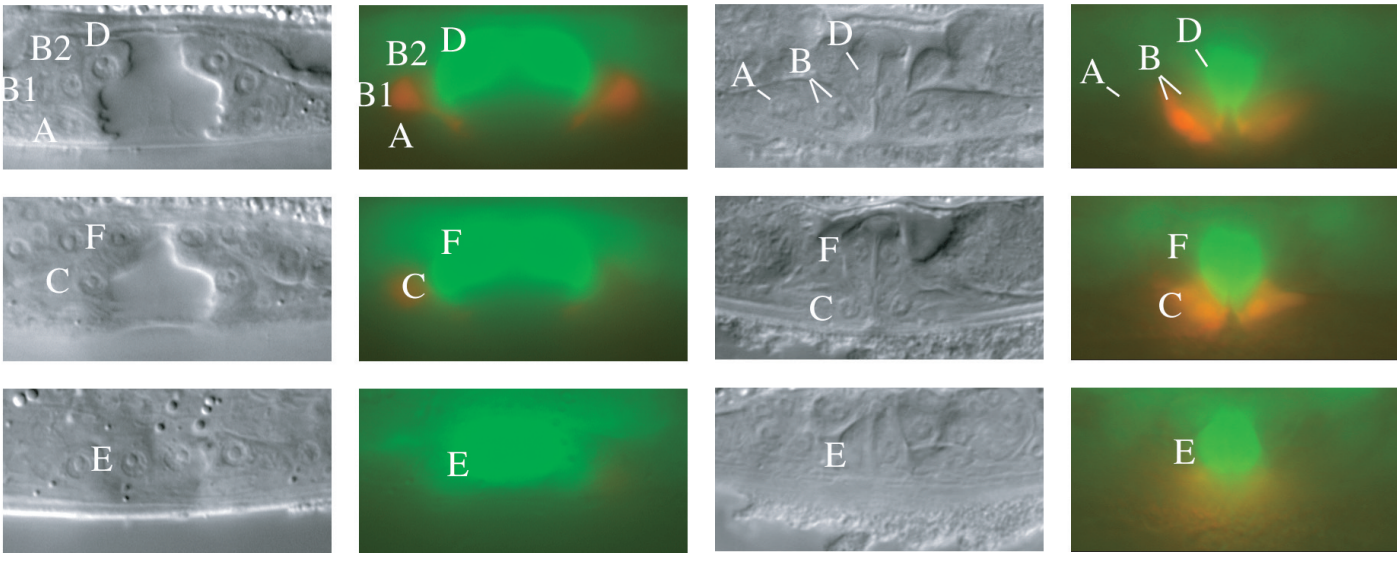


adult



L4

adult



(Adapted from Inoue *et al.*,submitted)