cis-Regulatory Control of Cell Fate-Specific Genes in *Caenorhabditis elegans* Vulval Organogenesis

Martha Kirouac and Paul W. Sternberg

(submitted for publication)

ABSTRACT

The great-grandprogeny of the *Caenorhabditis elegans* vulval precursor cells (VPCs) adopt one of the final vulA, B1, B2, C, D, E, and F cell types in a precise spatial pattern; *egl-17*, *zmp-1*, and *cdh-3* are differentially expressed in the developing vulva lineages and provide a potential readout for different signaling pathways. We have identified upstream *cis*-regulatory regions of these three genes sufficient for their ability to confer vulval cell type specific regulation. A 143-bp region of *egl-17* is sufficient to drive vulC and vulD expression, while a separate 102-bp region drives the early expression in presumptive vulE and vulF cells. A 300-bp region of *zmp-1* is sufficient to drive expression in vulE, vulA, and the anchor cell. A 689-bp region of *cdh-3* is sufficient to drive expression in the anchor cell and vulE, vulF, vulD and vulC; a 155-bp region is sufficient to drive anchor cell expression; and a separate 563-bp region is also sufficient to drive expression in these vulval cells. We have found no evidence of repressor elements in any of these genes with respect to vulval and anchor cell expression.

INTRODUCTION

In the *C. elegans* vulval ectoderm, three intercellular signaling pathways, EGF, NOTCH, and WNT, 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; Sternberg and Horvitz, 1986; Sulston and Horvitz, 1977). Once the VPCs terminally differentiate into one of the final vulval fates, vulF, E, D, C, B2, B1, and A (Figure 1), their morphogenetic interactions lead to the development of seven toroidal cells that connect the endothelium of the uterus to the external epithelium (Figure 1; Sharma-Kishore *et al*., 1999). Little is known about the individual roles of these vulval cells following their terminal differentiation, and what cell-specific functions they possess. However, the differentiation 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. While a number of transcription factors are known to be involved in vulval development (e.g., *lin-1*, *lin-29*, *egl-38*, *lin-31*, *lin-39*, *lin-11*), their targets are not known (Beitel *et al*., 1995; Bettinger *et al*., 1997; Chang *et al*., 1999; Clark *et al*., 1993; Euling *et al*., 1999; Freyd *et al*., 1990; Tan *et al*., 1998).

The gonadal anchor cell (AC) serves as the source of the inductive signal, LIN-3, which promotes vulval fates in the VPCs (Hill and Sternberg, 1992; Katz *et al*., 1995; Kimble, 1981). The anchor cell also helps establish a functional connection between the

vulva and the uterus (Newman and Sternberg, 1996; Newman *et al*., 1996). 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 identification of major transcriptional factors that control cell-specific gene expression.

Here we focus on three genes that are differentially regulated in these vulva cell types and in the anchor cell: *egl-17*, which encodes a fibroblast growth factor family member (Burdine *et al*., 1997, 1998); *cdh-3*, which encodes a FAT-like cadherin (Pettitt *et al*., 1996); and *zmp*-1, which encodes a zinc metalloproteinase, (J. Butler and J. Kramer, personal communication; Wada *et al*., 1998). These genes offer the opportunity to find *cis*-regulatory elements for multiple vulval cell types as well as the anchor cell. The identification of sequences that direct expression in these cell types will lead to a deeper understanding of the regulatory networks that pattern the vulva. We have analyzed the *cis*-regulatory sequences of these genes in *C. elegans* and report here on the different regulatory regions that drive expression of these three genes in the vulva cells and the anchor cell.

MATERIALS AND METHODS

Generation of *C. elegans* promoter GFP constructs

Using PCR (supplemental material, Table 1), the regions of interest were amplified, with TaKaRa LA Taq (Takara Shuzo), and cloned into the minimal promoter *pes-10*, pPD107.94, (a gift from the Fire lab) using restriction sites engineered into the primers. The PCR protocol used was: 94.0 °C for 4 minutes; 30 cycles 94.0 °C for 30 seconds, 60.0°C for 30 seconds, 68.0°C for 45 seconds; and 68.0 °C for 7 minutes.

As a template for PCR, the following constructs were used: the *egl-17* promoter NH#293 (Burdine *et al*., 1998); the *zmp-1* promoter pJB100 (J. Butler and J. Kramer, personal communication); and the *cdh-3* promoter jp#38 (Pettitt *et al*., 1996).

The nomenclature of the constructs generated in this study is derived from the primers used to amplify the region. In all cases, the first one to three digits represent the 5' primer, and the digits after the hyphen represent the 3' primer. Although we performed a systematic dissection of these three upstream sequences, not all constructs made are shown in this paper because of space limitations. For a comprehensive list, see the supplemental material in figures 1, 2 and 3.

The *egl-17* genomic region of NH#293 contains 3819 bp of sequence upstream of the translational start site. The first exon of the transcript starts at nucleotide 4610, and translation starts at nucleotide 4708. Nucleotide 790 of the *egl-17* upstream region corresponds with nucleotide 17648 in Genbank cosmid F38G1 (Accession # AC006635). The *zmp-1* genomic region in pJB100 contains 3472 bp of sequence upstream of the translational start site. The translational start site of ZMP-1 is at nucleotide 3473. Nucleotide 1 of this *zmp-1* upstream region corresponds with nucleotide 7630 in Genbank cosmid EGAP1 (Accession # U41266). The jp#38 genomic region of *cdh-3* contains 5928 bp of sequence upstream of the translational start site, whose start codon occurs at nucleotide 6041. Nucleotide 113 of the *cdh-3* upstream region corresponds with nucleotide 37343 in Genbank cosmid ZK112 (Accession # L14324).

Generation of *C. elegans* promoter deletion GFP constructs

An internal deletion was made using PCR primers (Supplemental Material, Table 1) that are homologous to 20 bp on either side of the region of the deletion. In primary PCR reaction, the deletion was generated using internal primers that span the deletion region, with outside primers mk151/mk50 and mk152/mk51 for construct Δ3/4. This generated two fragments with homologous ends containing the deletion. In a second round of amplification, just the outside primers mk50 and mk51 were used on the combined gel-purified products from the first PCR reaction that served as the template. The PCR protocol used for both the primary and secondary PCR reactions was: 94.0 °C 4 minutes; 30 cycles 94.0 °C for 30 seconds, 58.0°C for 30 seconds, 65.0 °C for 40 seconds; and 65.0 °C for 7 minutes.

Sequencing of constructs

The following constructs were sequenced to confirm these sequences: mk158-159, mk66-156, mk155-67, mk64-65, mk66-67, mk96-63, mk135-119, mk96-145, mk146-144, mk135-134, mk135-147, mk96-143, mk135-143, mk102-56, mk102-104, mk80-104, mk103-148, mk50-111, mk50-115, mk52-51, mk52-74, mk36-74, mk76-51, mk107-51, mk121-51, mk50-124, mk50-74, mk50-123, mkΔ3/4, mk153-148, mk153-154, mk103-56, mk36-51, mk106-51, mk50-75, mk118-143, mk135-147, mk125-132, mk96-143.

Microinjection of promoter GFP constructs into *C. elegans*

The constructs were microinjected into the gonads of animals of genotype *pha-1*(e*2123ts*); *him-5*(e*1490*) line using a standard protocol (Mello *et al*., 1991). The constructs were injected at a concentration of 100 ng/µl, with 20 ng/µl pBluescript SKII (Stratagene), and 82

ng/µl *pha-1*(+), pBX. Transgenic animals that stably transmitted the extrachromosomal arrays were isolated by selecting viable F1 animals at 22 °C to new plates, and examining their progeny for GFP expression in the anchor cell and the vulval cells.

Microscopy of transgenic animals

Animals were mounted on 5% noble agar pads and scored at 20°C for GFP expression under Nomarski optics using a Zeiss Axioplan microscope with a 200-watt HBO UV source, and a Chroma High Q GFP LP filter set (450 nm excitation/505 nm emission). At least two lines for each construct were examined.

egl-17 early expression in the grandaughters of P6.p, the precursor to vulE and vulF cells, was scored at the four-cell stage. *egl-17* vulC and vulD GFP expression was scored between the late L4 to young adult stages (Burdine *et al*., 1998). *zmp-1* anchor cell GFP expression was scored between the L3 and the early L4 stage. VulE and vulD expression was scored between the late L4 and young adult stages. *zmp-1* vulA expression was scored between the young adult and adult stages (Wang and Sternberg, 2000). *cdh-3* AC GFP expression was scored between the L3 and the early L4 stage. *cdh-3* vulE, vulF, vulC, and vulD expression was scored between the late L3 stage through late L4 stages (Figure 1; Pettitt *et al*., 1996).

Prediction of binding sites using Transfac database

Possible binding sites for known transcription factors in the regions defined by deletion analysis in the *egl-17*, *zmp-1* and *cdh-3* upstream regions were determined using the MatInspector program (http://www.genomatix.de/mat_fam; Quandt *et al*., 1995).

AlignACE predictions of over-represented sequences

AlignACE is based on a Gibbs sampling algorithm that computes a series of motifs that are over-represented in the input sequence(s) (http://atlas.med.harvard.edu/cgi-bin/alignace.pl; Roth *et al.,* 1998). This algorithm assigns a score to each motif; the MAP score (maximum a priori log likelihood) is the functional readout of the degree to which a motif is over-represented relative to the expectation for the random occurrence of such a motif in the sequence under consideration (Roth *et al.,* 1998). In this analysis, we chose a MAP cut-off of 10. When this cut-off was applied in a search of motifs in the genome of *Saccharomyces cerevisiae*, Hughes and colleagues found that this threshold did not lead to the rejection of any of the best examples of known *cis*-regulatory elements (Hughes *et al.*, 2000). We used a GC content setting of 0.35, and searched for motifs of eight and 10 nucleotides; these nucleotides do not have to be contiguous, but will receive higher MAP score if they are. A higher MAP score reflects: (1) a greater number of aligned sites; (2) a more tightly conserved motif; (3) less total input sequence; (4) more tightly packed information-rich positions; and (5) enrichment of the motif with nucleotides that are less prevalent in the genome (Hughes *et al.*, 2000).

RESULTS

To identify regulatory sequences sufficient to drive cell-specific expression, genomic fragments were tested for their ability to drive GFP expression from the heterologous *pes-10* basal promoter. This promoter does not drive expression of GFP in any of these tissues on its own. However, it contains the basic sites for the transcriptional machinery, which when

combined with an enhancer region, can drive GFP expression in a cell type-specific manner (Seydoux and Fire, 1994; G. Seydoux, personal communication).

Vulval specificity in the *egl-17 cis*-regulatory region in *C. elegans*

The *egl-17::*GFP translational fusion NH#293 is detectable in P6.p, P6.p daughters and granddaughters (the presumptive vulE and vulF cells), turns off in early L4, and turns on again in vulC and vulD cells in mid L4. This *egl-17::*GFP construct contains 3.9 kb of sequence upstream of the translational start (Burdine *et al*., 1998). We divided the 3.9 kb upstream region into four sub-fragments. Of the initial constructs, mk27-49 (3502-4586) showed expression in vulC and vulD cells, and mk15-20 (1716-3690) showed weak variable expression in the presumptive vulE and vulF cells, while mk153-154 (4565-4667) showed weak expression in the presumptive vulE and vulF cells (Figure 2).

We next sub-divided the mk27-49 (3502-4586) region to identify the minimal region sufficient for vulC and vulD specificity. A 143-bp region, mk125-132 (4331-4474), is sufficient to drive strong expression in vulC and vulD (Figure 3A). A comparison of mk125-132 (4331-4474), which drives expression clearly in both vulC and vulD cells, to mk102-56 (4359-4516), which drives expression weakly in vulD cells and not at all in vulC cells, suggests that the 5' end of this region is involved in vulC expression. Likewise, when we compare mk80-104 (4316-4466), which drives expression weakly in vulC cells but not at all in vulD cells, to mk80-132 (4316-4484), which drives expression in both vulC and vulD cells, it appears likely that the 3' end of this region is necessary for vulD expression (Figure 3A). However, the expression levels of both constructs are severely compromised when compared to the full-length construct. When both of these sites are removed, mk102-104

(4354-4466), no GFP expression is seen in either vulC or vulD. The nucleotide sequence of this entire region is shown in Figure 4A.

In addition to defining the region sufficient to drive GFP expression in vulC and vulD cells, we examined the regions defined by the initial constructs mk15-20 (4565-4667) and mk153-154 (4565-4667) for the minimal region sufficient to confer specificity of expression in the presumptive vulE and vulF cells. We have defined two regions that together confer strong expression in the presumptive vulE and vulF cells. One of these regions, 4565-4667, (mk153-154, Figure 3B), is located in the 5' UTR of the *egl-17* gene; 4565-4667 is sufficient to confer expression in these cells, but the expression is slightly variable and weaker than the full-length reporter construct. Additionally, a second element plays a role in conferring specificity to these cells. While constructs mk82-100 (2888-3611) and mk84-20 (3182-3640) show faint, inconsistent expression in the presumptive vulE and vulF cells, constructs mk82-85 (2088-3203) and mk27-20 (3502-3690) do not. This observation suggests that either the element responsible for this weak expression lies within the region 3203-3502, or that multiple sites are required that are dispersed throughout the larger region 2888-3690. When both of these regions are present, as in the much larger constructs mk15-148 (1716-4732) or mk84-148 (3182-4732), the expression is comparable to the level of early expression seen in the full-length reporter construct. The nucleotide sequence for mk84-148 (3182-4732) is shown in Figure 4A. In mk153-148 (4565-4732), despite containing the sequence that is sufficient to drive GFP expression in line mk153-154 (4565-4667), we see no GFP expression. This observation suggests the presence of a repressor of early expression in the presumptive vulE and F cells between 4667-4732. Another possibility is that the variability

of these lines (including mk103-148, 4427-4732) might be due to differences in transgene copy number.

Vulva and anchor cell specificity in the *zmp-1 cis*-regulatory region in *C. elegans*

The *zmp-1*::GFP marker strain containing pJB100 has 3.8 kb of *zmp-1* upstream regulatory sequence; the GFP in this strain is expressed in the anchor cells of L3 larvae, and in late L4 and young adult animals it is also expressed in vulE, D and vulA cells (J. Butler & J. Kramer, personal communication; Wang and Sternberg, 2000). We divided this 3.8 kb upstream region into four fragments (Figure 2). Of the first constructs made, mk29-32 (791-1618) showed expression in the anchor cell, and in vulE and vulA cells. No construct was found to drive the expression in vulD cells, which is seen in the full-length reporter construct.

We next sub-divided the sequences defined by the initial construct mk29-32 (791-1618) to define a 380 bp region (mk50-51; 1052-1438) that is sufficient to confer anchor cell, vulE and vulA cell specificity on the *pes-10* promoter (Figure 5). This region also confers uterine cell expression; however, we chose not to analyze this expression pattern further. When this minimal region was further sub-divided, we were able to identify regions of the *zmp-1* promoter that drive expression in just the anchor cell, for example mk36-51 (1180-1438), mk50-124 (1052-1268), and mk76-74 (1147-1378), but we were not able to identify fragments active only in vulA or vulE. This failure is in spite of the fact that when successive deletions are made on either end, the expression pattern is lost in a reproducible manner, so that first vulA expression is lost, then vulE expression, and finally anchor cell expression (Figure 5). For instance, consider successive 5' deletions in which the 3' end is maintained at nucleotide 1438. In mk76-51 (1147-1438), vulA expression is lost first. Then, in mk107-51

(1165-1438), vulE expression becomes variable and, finally, in mk121-51 (1191-1438), anchor cell GFP expression is lost (Figure 5, constructs mk50-51 through mk121-51). Similarly, successive 3' deletions, in which the 5' end is maintained at nucleotide 1052, show the same pattern of expression loss (Figure 5, constructs mk50-75 through mk50-11). We did not observe vulA expression without expression in both vulE and the anchor cell, nor did we observe vulE expression without anchor cell expression. This hierarchy suggests that it is the number of binding sites, rather than just the qualitative aspects of these sites, that determines the expression pattern, and hence that the different cell types have different levels of the factor that binds these sites. The nucleotide sequence of mk50-51 (1052-1438) is shown in Figure 4B.

Further support for the hypothesis that the quantity of sites dictates the *zmp-1* expression pattern is our observation that when the end points are changed, the regions that were necessary for expression in a given cell type become important for other expression patterns. For instance, as seen in Figure 5, in constructs mk76-51 (1147-1438), mk107-51 (1165-1438), and mk36-51 (1180-1438), which have the same 3' end, 1438, but are successively smaller on the 5' end, the region 1147-1165 appears to play an important role in the expression in vulE cells, and the region 1165-1180 seems to play a role in both vulA and vulE expression. However, when we tested mk76-74 (1147-1378) and mk36-74 (1180-1378), which end at 1378 instead of 1438 on the 3' end, but whose 5' end is either at 1147 or 1180, respectively, we observe that the region 1147-1180 appears to be necessary for anchor cell expression. When we compare constructs mk52-74 (1119-1378) and mk76-74 (1147-1378) to mk50-51 (1052-1438) and mk76-51 (1147-1438), we find that the region 1119-1147 can be important for either vulE expression in the first constructs, or for vulA expression in the

latter two constructs. Also, when we compare mk52-51 (1119-1378) and mk50-74 (1052-1375) to mk52-74 (1119-1375), we see that either 1052-1119 or 1378-1438 is sufficient to drive GFP in vulA cells.

To test whether the level of expression is determined by the quantity of the sites alone, or whether qualitative aspects of the sites are also crucial, we made an internal deletion, 1262-1269 (mkΔ3/4) (Figure 5). This region was shown to be important for anchor cell expression in our deletion analysis. If only anchor cell expression were lost, then the experiment would suggest that it is qualitative aspects of this site that are most important in determining the expression pattern. If vulA expression, or both vulA and vulE expression, is lost instead, it would suggest that it is the number of sites bound that determines the expression pattern. The resulting deletion had a more complex effect. Construct mkΔ3/4 (Δ1262-1269) showed expression in the anchor cell, and in vulA cells, but showed no expression in vulE cells. In this case, anchor cell expression was not lost, indicating that indeed there seem to be multiple sites that can drive expression in a given cell type. We also saw the loss of vulE expression preferentially over the loss of vulA expression. This pattern was not observed until the internal deletion was made, and it suggests that, while the quantity of bound sites is an important determining event in the expression pattern, there are also qualitative aspects of sites important for expression in a given cell type.

Vulva and anchor cell specificity in the *cdh-3 cis*-regulatory region in *C. elegans*

A *cdh-3*::GFP fusion containing 6.0 kb of upstream sequence is expressed from the L2 stage in the anchor cell; during the L3 stage it is also expressed in vulE and vulF cells, and it is also expressed in the vulC and vulD cells of L4 larvae (Pettitt *et al*., 1996). We divided this

6.0 kb upstream region into seven subfragments (Figure 2). Of the initial constructs, mk62-63 (1478-3008) showed anchor cell expression; mk66-67 (4434-4997) showed vulva expression; and mk135-134 (2412-3419) showed both anchor cell and vulva expression.

Since both mk62-63 (1478-3008) and mk135-134 (2412-3419) were sufficient to drive expression in the anchor cell, we focused our search for the minimal anchor cell element in the overlapping region (~2300-3200). This region is also sufficient to confer uterine cell expression, however, we chose to focus our attention on the vulva and anchor cell elements. The minimal *cis*-regulatory region we observed to drive anchor cell expression is 155 bp (mk146-144; 2367-2522) (Figure 6A). This construct displays variable expression, but the 232 bp construct mk96-144 (2290-2522) expresses in all animals observed. The details of the specific sequences driving anchor cell expression are more complicated. There appear to be at least three regions (α, 2290-2431; β, 2431-2522; and γ, 2989-4363) that play a role in anchor cell expression. While any one of these regions is insufficient to drive anchor cell GFP expression on its own, as demonstrated by mk96-145 (2290-2431) (α), mk135-119 (2412-2713) (β), and mk64-65 (2989-4363) (γ), any two of these regions are sufficient to drive this expression, as demonstrated by mk146-144 (2367-2522) (α and β), mk135-143 (2412-3164) (β and γ), and the co-injection of mk64-65 (2989-4363) with mk96-145 (2290-2431) (α and γ). However, in the case of the co-injection of mk64-65 and mk96-145, less than 10% of the animals show GFP expression in the anchor cell. The nucleotide sequence of mk96-134 (2290-3419), which contains the α, β, and γ sites, is shown in Figure 4C.

This anchor cell expression pattern has at least one additional layer of complexity. The expression from some constructs comes on at the VPC 2-cell stage, while from other constructs it does not express until the VPC 4-cell stage. When mk135-143 (2412-3164) and

mk135-147 (2412-3101) are compared, it appears that the region 3101-3164 can confer early anchor cell expression. This region is encompassed in the γ region. There appears to be another region, 1478-2290, that is sufficient to drive early anchor cell expression when mk62-63 (1478-3008), which expresses GFP at the VPC- 2-cell stage, is compared to mk96-63 (2290-3008), which does not express GFP at the VPC 2-cell stage (Figure 6B). This region is separate from the α, β and γ regions. It is possible that this is a separate temporal element that drives expression at the VPC 2-cell stage, or that there are general enhancers in these regions, and that without these enhancers the expression is not bright enough to see at the VPC two-cell stage. We chose not to analyse the early expression, but rather to focus on elements that drive expression at the VPC four-cell stage.

In the initial set of constructs, two separate regions were sufficient to confer vulval cell expression on the *pes-10* promoter: mk66-67 (4434-4997) and mk135-134 (2412-3419). We examined both of these regions to define the minimal sequence sufficient to confer vulval cell specificity. Vulval expression appears to be independent of anchor cell expression; the 689 bp region 2412-3101 (mk135-147), which is insufficient for AC expression, is sufficient to drive expression in vulE, C, D, and occasional expression in vulF (Figure 6C). Since mk96-143 (2290-31664) shows vulval expression while mk96-63 (2290-3008) does not, the 156 bp region 3008-3164 must be necessary for vulval expression. However, since construct mk64-65 (2989-4363) shows no expression and contains this region, it cannot be sufficient to confer vulval expression. The region 2412-2692 also appears to play a critical role in vulval expression, as demonstrated by comparing mk135-143 (2412-3164) to mk118-143 (2692-3164). Construct mk118-143 (2692-3146) shows weak expression in vulE and D, with occasional vulC expression; but there is no discernable expression in vulF, while mk135-143

(2412-3164) shows expression in vulE, F, C and D. There appear to be multiple sites involved in conferring vulva cell expression in the 2290-3164 region. The nucleotide sequence of mk96-134 (2290-3419) is shown in Figure 4C.

A second region, 4434-4997 (mk66-67, 563 bp) also appears to be sufficient to drive vulval expression (Figure 6C). There are some qualitative differences in the expression pattern when this region is compared to the other region (2412-3101) sufficient to drive vulval expression. The region 4434-4997 confers very bright vulF expression, while vulF expression in the region 2412-3101 is much weaker relative to other cell types. In addition, the vulC and vulD expression in the region 4434-4997 is weaker than the vulval expression in the 2412-3101 region. Thus, while multiple regions are sufficient to drive GFP expression in the vulva, they qualitatively differ in their detailed activity. The second region, 4434-4997,was subdivided into three overlapping regions: mk66-156 (4434-4729), mk155-67 (4719-4997), and mk158-159 (4680-4883) (Figure 6C). Of these constructs, mk66-156 (4434-4729), in a single line, showed very weak sporadic expression in vulC and vulD cells, and mk158-159 (4680-4883) drove very weak expression on rare occasion in vulE, or F. As with the other vulva cell sufficiency region, multiple sites important to all vulva cell expression must lie in this region. The nucleotide sequence of mk66-67 is shown in Figure 4D.

Transfac putative binding site predictions in upstream sequences

To find potential binding sites for transcription factors, we used the MatInspector program (http: www.genomatix.de/mat_fam; Quandt *et al*., 1995). We set the core matrix similarity to a minimum of 0.90 to maximize the specificity of the binding sites. We compared binding

sites in mk50-51 (1052-1438; *zmp-1* upstream region sufficient to drive expression in the anchor cell, vulE and vulA) to those in mk96-134 (2290-3419), the *cdh-3* region that is sufficient to drive expression in the anchor cell, as well as vulE, F, C and D) (Table 1). In this comparison, 21 shared binding sites are predicted. When we analyzed the two regions sufficient to drive *cdh-3* vulva expression, mk66-67 (4434-4962) and mk96-134 (2290-3419), we found 39 distinct binding sites that are shared between these regions. Finally, this process was utilized to compare mk84-148 (3182-4732; the *egl-17* region that is sufficient to drive vulC and D as well as early expression in the presumptive vulE and vulF cells), to mk50-51, mk96-134, and mk66-67. In this case, any putative binding site that is shared in three of these might indicate a factor involved in conferring cell specificity, since these genes express in overlapping cell types. In 12 cases, the same binding site showed up in all four regions. These might be candidates for a more general factor that drives tissue specific expression in all vulva cells. Some families are well represented in these analyses: the homeodomain family, the forkhead family, the cAMP-responsive element family, the octamer family, and the zinc finger family. However, the candidates are numerous and we have chosen to define the regions further using phylogenetic footprinting (Kirouac and Sternberg, in prep.) before attempting to distinguish between these candidates.

AlignACE predictions of over-represented sequences

The Transfac database (Quandt *et al*., 1995) is used to identify binding sites of known transcription factors, but it is likely that motifs might exist that are uncharacterized, or that have altered binding specificities in *C. elegans* from the binding sites of known transcription factors in other systems. To determine if the apparent coordinate regulation of these genes

might indicate common DNA sequences, we used the AlignACE program (Roth *et al*, 1998), which computes motifs based on sequences that are over-represented in the input sequence. We were able to identify motifs over-represented in the sufficiency regions of *egl-17* mk84-148, *zmp-1* mk50-51, and *cdh-3* mk96-134 and mk66-67 individually. In addition, we identified motifs common to mk96-134 to mk50-51, which are each sufficient to confer expression in the anchor cell, and mk84-148, mk50-51, mk96-134, and mk66-67, which all drive expression in the vulva. These motifs may represent candidate transcription factor binding sequences that are critical for either anchor cell or general vulval expression respectively. We also compared mk96-134 to mk66-67 for reasons similar to those of the other vulval expression comparison, with the additional benefit that these two regions are located in the same upstream sequence, and might identify candidate motifs that are specific to *cdh-3* vulval expression. Analysis of these motifs should help us identify candidate sequences, known or unknown, for which to search for in upstream regulatory regions of genomic sequences for potentially co-regulated genes. One caveat of this approach to keep in mind is that some motifs that occur ubiquitously in a genome may be given a high MAP score, but have little relevance to the particular set of genes being examined.

In our analysis of *egl-17* region mk84-148, we found 14 eight-bp motifs, of which only five scored above our MAP score cut-off limit of 10, and we found an additional three motifs of six 10-bp motifs that were also above this threshold MAP score (Table 2A). Several of these candidate motifs showed multiple overlapping motifs (e.g. motif 4.8 and 2.8 share five sites); this overlap is indicative of the fact that either these motifs are really the same, or that these sites co-localize, which may identify binding sites of trans-acting factors that bind cooperatively. Some of these sites, as seen in Table 2B, are in regions that we defined in our

sufficiency analysis as being important for the fidelity of *egl-17* expression; for the early expression in the presumptive vulE and vulF cells, as well as later expression in the terminally differentiated vulC and vulD cells. For instance, motif 1.8 site 1158 is located between primers mk125 and mk102; this region (4331-4359, Figure 3A) is an important one for conferring expression in vulC and vulD. One motif, 3.8, is notable for its location in between primers mk154 and mk148. This region, 4667-4732 (Figure 3) may have repressor elements that play a role in controlling the early expression of *egl-17* seen in the presumptive vulE and vulF cells.

The analysis of the *zmp-1* sufficiency region mk50-51 yielded only one 8-bp motif (Table 2A). However, this motif has multiple sites, of which two are in regions important for the fidelity of *zmp-1* expression in vulA and vulE cells (Table 2B). Site 316 lies between the boundaries of primers mk74 and mk115. This region (1367-1378; Figure 5) is critical for conferring vulA expression. Site 100 is located between primers mk107 and mk36; this region (1165-1180, Figure 5) plays an important role in driving expression in both vulE and vulA.

We analyzed two *cdh-3* regions. The first region, mk96-134, is sufficient to drive expression both in the anchor cell and in the vulva cells. The second region, mk66-67, is able to confer vulval expression in vulE, F, C and D. In our analysis of mk96-134, we identified two 8-bp motifs and one 10-bp motif (Table 2A). The 10-bp motif, 3.10, overlaps almost entirely with motif 1.8 (Table 2B). All three of these motifs are located in multiple sites throughout the alpha, beta and gamma regions (Figure 6), which play a role in conferring anchor cell expression (Table 2B). Motif 1.8 shows a paucity of sites between primers mk136 and mk164 (Figure 4). This region may be important for conferring vulval expression.

Construct mk96-143 is able to confer vulval expression, but constructs mk96-63 and mk64-65, which divide this region in two cannot drive this expression. These results indicate that either the sites are located toward the center of this region where the break between constructs mk96-63 and mk64-65 occurs, or that sites from either end of the larger region 2290-3164 (mk96-143) are required for this expression pattern. If it is the latter case of multiple sites on either end of this region, then motif 1.8 may be a good candidate sequence, since it has few sites in the center, but multiple sites on either end of this region. We found no motifs above a MAP threshold of 10 in the region mk66-67.

To identify motifs that may be important in conferring anchor cell specificity, we compared the region mk96-134 to mk50-51; both of these regions are sufficient to drive anchor cell expression from a naïve promoter. We identified six candidate motifs (8-bp motifs and four 10-bp motifs, Table 2A). All of these motifs had sites that were located in regions important for conferring expression in the anchor cell of*zmp-1*, and all were present in multiple copies in the alpha, beta and gamma regions that are critical for anchor cell expression in *cdh-3* (Table 2B).

We identified no candidate motifs present in both mk96-134 and mk66-67. However, in our analysis of all the sufficiency regions that express in the vulva we found 13 candidates for motifs that might bind trans-acting factors that play a more general role in conferring vulva tissue specificity (Table 2A). Of these 13 candidates, all but one, motif 4.8, had at least one, and usually multiple, sites in all four regions, mk84-148, mk50-51, mk96-134 and mk66-67. Motif 4.8 was not found in mk66-67 (Table 2B).

DISCUSSION

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 5' flanking sequences of the cuticle gene *dpy-7*' 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). The success stories lie in the studies of the inducible expression of the *C. elegans* metallothionein genes, *mtl-1* and *mtl-2*, which occur in intestinal cells (Moilanen *et al*., 1999), and in the study of *daf-19* -regulated expression of genes expressed broadly in the sensory cilia (Swoboda *et al*., 2000).

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, and 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 and results in sensory defects. Twenty *C. elegans* promoters of genes 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).

We have attempted to address this assumption by analyzing the *cis*-regulatory sequences of three genes that have overlapping expression patterns in particular cell types within the *C. elegans* vulva and anchor cell. We chose three genes, *egl-17*, *zmp-1,* and *cdh-3*, whose function is not required for the normal development of the vulva and anchor cells, and hence lie downstream of the cell-fate specification pathways (Branda and Stern, 2000; Burdine *et al*., 1997, 1998; Pettitt *et al*., 1996; Wada *et al*., 1998; J. Butler and J. Kramer, personal communication). While the roles of these three genes in the vulva and anchor cell

have yet to be determined, all three are members of families that have been shown to be involved in morphogenesis and extracellular matrix remodeling.

From our analysis of the upstream sequences of these genes, we were able to identify a number of cell-specific regulatory elements. Within these sufficiency pieces we have been able to identify multiple motifs, using the program AlignACE (Roth *et al.*, 1998), that may play a role in conferring anchor cell-specific expression, and general vulva tissue-specific expression.

A variety of regulatory mechanisms are used in *C. elegans* to control gene expression throughout development. Some genes, such as the acetylcholinesterase gene, *ace-1* (Culetto *et al*., 1999), and the cuticle gene *dpy-7* (Gilleard *et al*., 1997), are regulated in a relatively simple fashion by tissue-specific promoters. Other genes such as *ges-1* (Egan *et al*., 1995), and *mec-3* (Wang and Way, 1996), are regulated in a more complex manner and require both activator and repressor elements to establish proper expression. Krause *et al.* (1994) has suggested that a more complex mechanism of control may be used in *C. elegans* to regulate genes that are expressed prior to terminal differentiation. Genes that are involved in controlling differentiation and cell fate would most likely be responsive to multiple inputs at many stages and cell interactions, as well as possessing the ability to regulate multiple gene regulatory networks, to dictate and shape these cell-fate decisions. In terminally differentiated cells, tissue identity is already established, and the need for such complex response mechanisms may be logistically unfavorable, especially in large families of genes that are likely to have partially redundant functions. In the analysis of the upstream sequences of *cdh-3, zmp-1*, and the late expression in vulC and vulD cells in *egl-17*, we indeed find discrete regions that direct tissue-specific expression, although each of these

regions appears to have multiple sub-elements. Moreover, we have found no evidence of repressor elements in these regions. Although, when this analysis is conducted in the context of the native promoter, negative regulatory elements may yet be found to play a role in regulating and establishing the cell specificity of these genes. We discuss below the regions directing expression of each of the cell markers that we examined in this study.

egl-17

The *egl-17* expression pattern is unique in our analysis: while it is expressed in the terminally differentiated cells vulC and vulD, it also shows expression in the presumptive vulE and vulF cells (Burdine et al., 1998). The early *egl-17* GFP expression in the presumptive vulE and vulF cells appears to be separable from the later expression in vulC and vulD cells. This early expression of *egl-17* in the presumptive vulE and vulF cells is the first marker indicating that the progeny of P6.p are specified to become primary cells (Ambros, 1999). Therefore, this expression may respond directly to the RAS-signaling pathway involved in the specification of these cells. There appear to be at least two regions directing this early expression pattern. One element that lies within 281 bp of the transcriptional start is sufficient for this pattern, but is not as strong as the full-length reporter. This expression is enhanced by an element that lies 1 kb further upstream. There also appears to be a region, 4667-4732, that may be involved in the negative regulation of early expression, as exhibited by mk153-148. This region inhibits early expression but its removal does not drive ectopic expression in *C. elegans*.

The minimal region that is sufficient for vulC and vulD cell-specific expression is 143 bp (mk125-132). There is some separability of the regions that drive vulC expression

from those that drive expression in vulD. The 5' end of this region plays a critical role for vulC expression. Likewise, the 3' end of this region is important for vulD expression. However, removal of either the 5' or 3' ends of this region substantially reduces expression levels when compared to the full-length reporter. While it is clear that there are sites that are required for expression that reside on either end of this region, it is unclear what role the remaining portion of this sufficient fragment plays in controlling the expression. Further systematic dissection of this region may elucidate other sites required for the fidelity of the expression pattern in vulC and vulD. *egl-17* expression conferred in the terminally differentiated vulC and vulD cells does not appear to be under negative regulation. Using the Transfac database (Quandt *et al*., 1995), we found no evidence of convergence of signaling pathways, in particular ETS or WNT target sites in conserved regions at the level of this promoter. However, the alignACE program (Roth *et al.,* 1998) did identify several candidate motifs that might identify binding sites of components of these pathways, either new or as yet uncharacterized.

zmp-1

Using deletion analysis, we have defined a 380 bp region, mk50-51 (1052-1438), that is sufficient to confer vulE, vulA, and anchor cell specificity on the *pes-10* promoter. In our analysis of the *zmp-1* expression pattern, we did not identify any region that drives the weak vulD expression found in the full length reporter marker.

Multiple sites within the small 386 bp region confer expression in a reproducible, predictable fashion. When successive deletions are made on either end of this region, the expression pattern is lost in a reproducible manner: vulA expression is lost first, then vulE

expression, and finally anchor cell expression. We found segments of the *zmp-1* 5' region that drive expression in only the anchor cell, but we were not able to identify regions that confer expression only in vulA or vulE. The AlignACE program (Roth *et al.,* 1998) was able to identify a motif, 1.8, that is present in multiple copies throughout this region, and may serve as binding sites for such a factor. Yet, when sites necessary for the expression in the anchor cell are internally deleted, vulE expression, rather than anchor cell expression is lost. This observation suggests that while sequential deletions generate a reproducible pattern of expression loss, there are also some sites that are cell-type specific. Although this region appears to be part of a more complex regulatory mechanism, we saw no ectopic expression, suggesting that there are no repressor elements involved in the coordinated expression of this gene.

cdh-3

The complex *cdh-3* 5' regulatory region contains discernable tissue-specific *cis*-regulatory elements. *cdh-3* expression was examined in two tissues: the vulva (vulE, F, C, and D) and the gonad (anchor cell). The DNA elements that are sufficient to drive anchor cell expression are separable from the elements that drive expression in the vulva.

There appear to be at least three regions (α, β, and γ) that are important in anchor cell expression; two of these three elements must be present for expression. This mechanism of transcriptional control is reminiscent of the regulation of the *myo-2* gene in *C. elegans*, where three separable elements with pharyngeal enhancer activity were identified. Any combination of two of these elements, or duplication of a single element, was sufficient to confer expression in the pharynx, while singly they were inadequate to drive expression

(Okkema and Fire, 1994). When these regions are compared using the AlignACE program (Roth *et al.,* 1998), three motifs, 1.8, 2.8 and 3.10, where identified that each had multiple sites in the alpha, beta and Gamma regions. There are also two separable regions that are each sufficient to drive expression in the vulva cells. When these regions were compared to identify over-represented sequences with the AlignACE program (Roth *et al*, 1998), no common motifs were found. Despite the fact that both regions were sufficient for vulval expression, there were qualitative differences in the strength of expression in the individual cell types. In the second region, the expression in vulF was stronger than in the first vulval region. GFP expression in vulC is weaker when driven by the second region than by the first. Although we found limited evidence that there are individual elements responsible for expression in each of the vulva cell types, we did find evidence that multiple binding sites within both of these regions are responsible for the fidelity of the expression pattern. It is possible that the loss of expression in all the vulva cells is the result of a more general regulatory mechanism on all vulva cells. This all-or-none mechanism of conferring tissue specificity is also reminiscent of the *C. elegans myo-2*, analysis in which one of the enhancer elements described above was responsible for conferring expression in all pharyngeal cells, while the other two elements identified conferred specificity for specific pharyngeal sub-types (Okkema and Fire, 1994). In the analysis of *cdh-3*::GFP expression, we never saw expansion or ectopic expression of this marker, suggesting that there are no repressor elements.

Distance of elements from translational start sites in *egl-17*, *zmp-1*, and *cdh-3*

The distances of elements that confer cell specificity do not seem to lie within a fixed distance from the translational start sites of their respective genes. In the case of *egl-17*, the transcriptional start site of *egl-17* is less than 400 bp from the elements that are sufficient to drive expression in vulC and vulD cells, and is less than 281 bp from the element that is sufficient to drive early expression in the presumptive vulE and vulF cells. However, the 386 bp *zmp-1* regulatory region that confers tissue specificity lies over 2.0 kb upstream of the translational start site of ZMP-1. Finally, the *cdh-3* regulatory regions that are responsible for anchor cell expression lie almost 3.8 kb upstream of the translational start site of the gene, while the elements that control vulva expression lie 3.6 kb and 1.6 kb from the translational start of the *cdh-3* gene.

Analysis of putative *trans*-acting factors

While the focus of this project was to isolate cell-specific *cis*-regulatory response elements rather than identifying *trans*-acting factors, we were also looking forward to the more distant goal of determining the integration of the signaling pathways in the downstream genes of these pathways. Our deletion analysis defined small regions that are critical for the fidelity of the expression pattern of these three genes; however, these regions are still broad enough to obscure the resolution of distinct binding sites. We used the Transfac database (Quandt *et al*., 1995) to look for common putative *trans*-acting factor binding sites, as well as to indicate the integration of the known signaling pathways in the *cis*-acting regions. We found no obvious candidate sites based on location within analyzed regions to known transcription factors involved in *C. elegans* vulval patterning.

While a number of transcription factors (for example *egl-38*, *lin-26*, *lin-29*, cog-*1* and *lin-11*) (Bettinger *et al*., 1997; Chamberlin *et al*., 1997; Freyd *et al*., 1990; Labouesse *et al*., 1994; Rougvie and Ambros, 1995; Palmer *et al*., in prep) are known to affect the marker-gene expression patterns in the vulva, it is unclear at this time whether they are acting directly in the regulation of the genes, or more proximally in the specification of these cell types (M. Wang and P. Sternberg, unpublished data). Biochemical studies using the sufficiency pieces defined in this study will determine which of these factors has a direct effect on the transcriptional regulation of these genes.

An anchor cell element that drives transcription of LIN-3 has been identified, and involves two trans-acting factors (B. Hwang and P. Sternberg, in prep). Removal of these factors does not disrupt the expression of *cdh-3* or *zmp-1::*GFP in the anchor cell (B. Hwang and P. Sternberg, in prep). This observation suggests that there are at least two different mechanisms and/or factors that are used to establish the anchor cell expression.

Analysis of overrepresented sequences in regions of sufficiency

While the Transfac database (Quandt *et al*., 1995) identifies binding sites of known transcription factors, AlignACE (Roth *et al*, 1998) identifies sequences that are over-represented in a given sequence. This approach allows the isolation of candidate motifs either within a gene, or between genes. As discussed above, we were able to identify a number of candidate motifs that bound in mk84-148, mk50-51, and mk96-134. Additionally, we compared sufficiency regions that are expressed in the same tissue to identify common motifs that may play a role in conferring cell-type-specific expression in co-regulated genes. Through these inter-regulatory region comparisons, we have identified candidate motifs that

may play a role in anchor cell expression, and a more general vulva tissue-specific expression.

Conclusions

By analyzing the functional anatomy of tissue-specific and cell-specific patterns of three reporter genes, *zmp-1, egl-17* and *cdh-3*, we have narrowed a 3.9 kb upstream region of *egl-17* to a 143 bp region of *egl-17* that confers vulC and vulD expression, and a separate 102 bp region sufficient to drive the early expression in presumptive vulE and vulF cells. A 3.5 kb *zmp-1* upstream region has been narrowed to a 300 bp region that is sufficient to confer expression in vulE, vulA, and the anchor cell. Moreover, a 6.0 kb upstream region of *cdh-3* upstream sequence has been delimited to: a 689 bp region sufficient to drive expression in the anchor cell and vulE, vulF, vulD and vulC; a 155 bp region sufficient to drive anchor cell expression; and a separate 563 bp region also sufficient to drive expression in these vulval cells. One theme that remained the same in all the analyses is that we failed to identify any repressor elements involved in conferring expression in terminally differentiated cell types. However, we identified regions of similarity between these three *cis*-regulatory sequences, and provided evidence for several different mechanisms through which *C. elegans* regulates transcription. These mechanisms include the use of discrete, separable elements that confer cell-type specific expression (*cdh-3* in the anchor cell and *egl-17* in sister cells), the use of complex patterns of binding sites that act combinatorially to establish the fidelity of expression in a variety of cell types from different lineages (*zmp-1*), and tissue-specific elements responsible for driving expression in an entire tissue rather than in sub-domains of its constituent cells (*cdh-3* vulval expression). Finally, we have been able to isolate candidate

motifs for each of these regions, and between anchor cell specific, and vulva specific regions that may be important for the fidelity of these expression patterns. We can now use these candidate motifs as targets for mutational analysis and for searching the genome for other candidate genes that have these motifs, to test if they are expressed in a similar fashion.

Acknowledgments

We are grateful to Jim Butler and Jim Kramer for *zmp-1*::GFP and the sharing of unpublished data; Rebecca Burdine and Micheal Stern for *egl-17*::GFP; and Jonathan Pettitt, William B. Wood and Ronald Plasterk for the *cdh-3*::GFP. We would like to thank Cheryl Van Buskirk, Takao Inoue, Erich Schwarz, David Sherwood, Bhagwati Gupta and Rene Garcia for the critical reading of this manuscript. P.W.S is an investigator with the Howard Hughes Medical Institute, which supported this work.

REFERENCES

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

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.

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.

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.

Chamberlin, H., Palmer, R., Newman, A., Sternberg, P., Baillie, D., and Thomas, J. (1997). The PAX gene *egl-38* mediates developmental patterning in *Caenorhabditis elegans*. *Development* **124,** 3919-3928.

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.

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.

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.

Davidson, E. H. (2001). Regulatory Hardwiring: A Brief Overview of the Genomic Control Apparatus and its Causal Role in Development and Evolution in: *"Genomic Regulatory Systems: Development and Evolution." Academic Press,* 1-24.

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.

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.

Freyd, G., Kim, S., and Horvitz, H. (1990). Novel cysteine-rich motif and homeodomain in the product of the *Caenorhabditis elegans* cell lineage gene *lin-11*. *Nature* **344,** 876-879.

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.

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.

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

Hughes, JD., Estep, PW., Tavazoie, S., and Church, GM. (2000). Computational identification of *cis*-regulatory elements associated with groups of functionally related genes in *Saccharomyces cerevisiae*. *Journal of Molecular Biology* **296**, 1205-1214.

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.

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

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.

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.

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

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.

Newman, A., and Sternberg, P. (1996). Coordinated morphogenesis of epithelia during development of the *Caenorhabditis elegans* uterine-vulva connection. *Proceedings of the National Academy of Sciences USA* **93,** 9329-9333.

Newman, A., White, J., and Sternberg, P. (1996). Morphogenesis of the *C. elegans* hermaphrodite uterus. *Development* **122,** 3617-3626.

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.

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.

Quandt, K., Frech, K., Karas, H., Wingender, E., and Werner, T. (1995). MatInd and MatInspector: new fast and versatile tools for detection of consensus matches in nucleotide sequence data. *Nucleic Acids Res* **23,** 4878-4884.

Roth, FP., Hughes, JD., Estep, PW., and Church, GM. (1998). Finding DNA regulatory motifs within unaligned noncoding sequences clustered by whole-genome mRNA quantitation. *Nature Biotechnology* **10**, 939-945.

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

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

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.

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.

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

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.

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.

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.

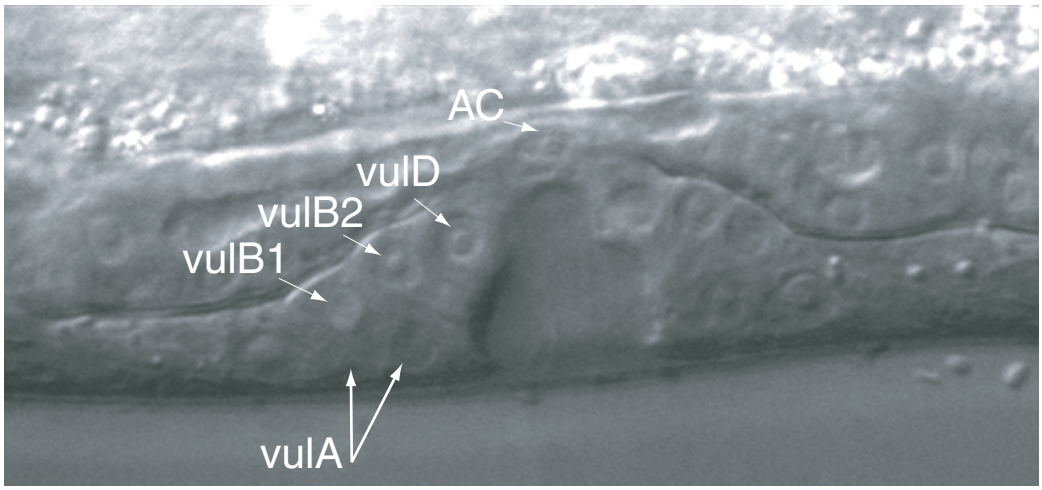
Wang, L., and Way, J. (1996). Promoter sequences for the establishment of *mec-3* expression in the nematode *Caenorhabditis elegans*. *Mechanisms of Development* **56,** 183-196.

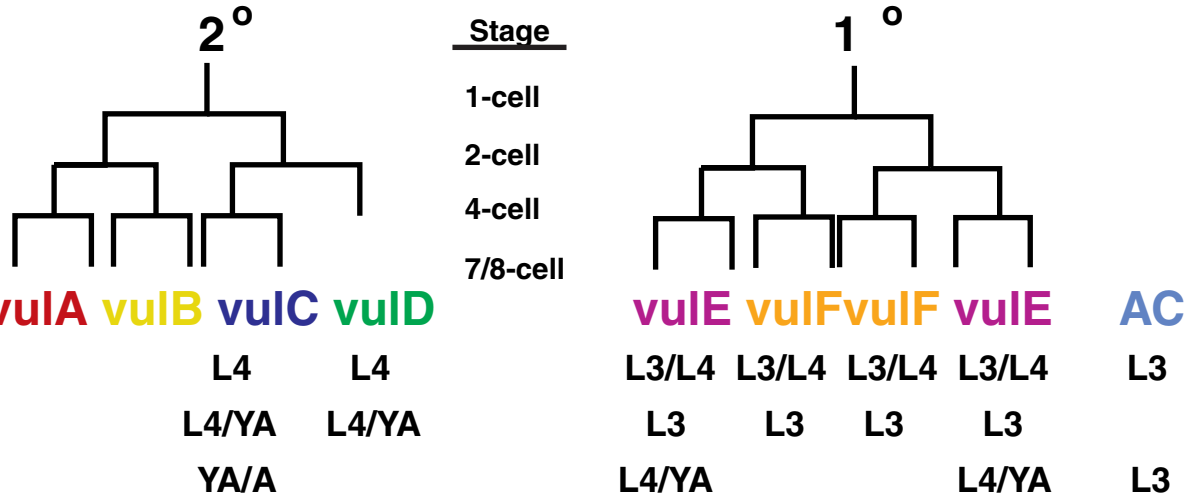
Wang, M., and Sternberg, P. (2000). Patterning of the *C. elegans* primary vulval lineage by RAS and Wnt pathways. *Development* **127,** 5047-5058.

Figure 1: Marker gene expression summary

Nomarski image of the vulva of an L4 animal showing the anchor cell as well as vulA, B1, B2 and D cells (vulC, E, and F are not in this focal plane). The three markers *egl-17*::GFP, *zmp-1*::GFP and *cdh-3*::GFP are expressed in different cell types of the vulva at various developmental stages. This figure shows the stages at which the expression of three marker genes was scored.

Figure 1: Marker gene expression summary





	vulA	vulB	vulC	vulD		vulE	vulF	vulF	vulE	AC
cdh-3			L4	L4		L3/L4	L3/L4	L3/L4	L3/L4	L3
egl-17			L4/YA	L4/YA		L3	L3	L3	L3	
zmp-1			YA/A			L4/YA			L4/YA	L3

Figure 2: Initial dissection of *egl-17, zmp-1* and *cdh-3* regulatory regions

For each of the three genes, the genomic region was divided into large sub-pieces of approximately 1kb each. **(A)** The genomic region of *egl-17* contains 3819 bp of upstream sequence. The first exon of the transcript starts at 4609, with the translational start at 4708. Nucleotide(nt) 790 of the *egl-17* upstream region corresponds with nt 17,648 in Genbank cosmid F38G1 (Accession # AC006635). **(B)** The *zmp-1* genomic region contains 3472 bp of upstream sequence. The start site of ZMP-1 is at nt 3,473. Nucleotide 1 of the *zmp-1* upstream region corresponds with nt 7630 in Genbank cosmid EGAP1 (Accession # U41266). AC stands for anchor cell. **(C)** The genomic region of *cdh-3* contains 5,928 nucleotides of upstream sequence. The translational start site occurs at nt 6041. Nucleotide 113 of the *cdh-3* upstream region corresponds with nt 37,343 in Genbank cosmid ZK112 (Accession # L14324).

Figure 2: Initial dissection of *egl-17*, *zmp-1,* and *cdh-3* regulatory regions

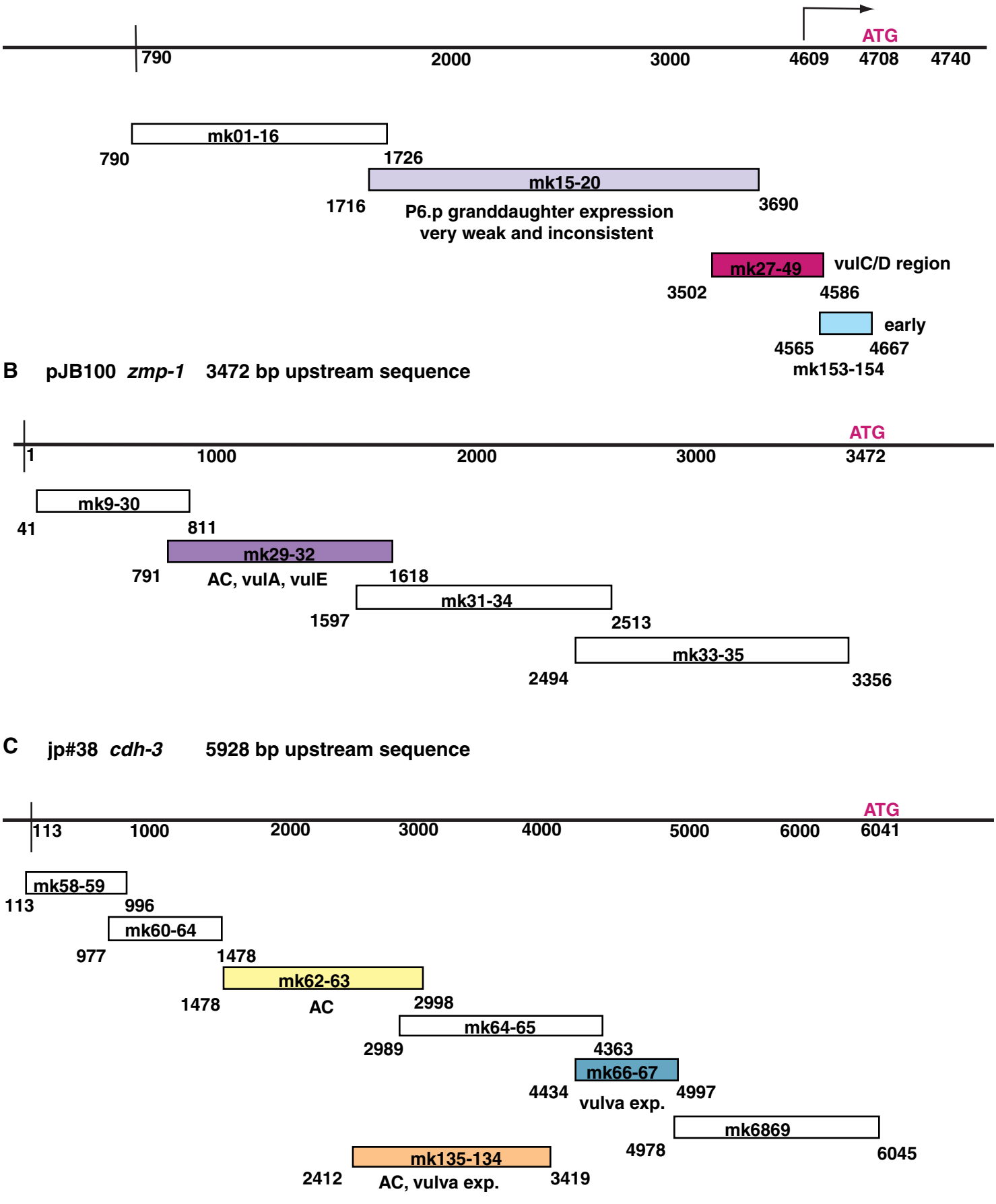


Figure 3: Upstream regions that direct *egl-17* expression

The list of constructs does not encompass all constructs that were made. (For a comprehensive list see Supplemental Figure 2.) (A) The constructs that were informative in determining two regions, 4331-4359 and 4466-4474, shown in pink, important in driving vulC and vulD expression. (B) The constructs that were most informative in determining the regions that drive the early expression in the presumptive vulE and vulF cells are depicted. The first region, 3182-3611, highlighted in orange, shows that, while not sufficient alone to drive the early *egl-17*:: GFP expression, when combined with the region shown in blue, 4565-4667, drives GFP expression at a level comparable to the full-length reporter construct. The region highlighted in blue depicts the region that alone is sufficient to drive expression in the presumptive vulE and vulF cells. A +/- indicates that either the expression level was reduced with respect to other constructs, or that not all animals showed consistent expression in the cell. Mk80-104 showed very weak vulC expression in 1/2 lines. 102-56 showed weak expression in vulD in 3/3 lines. On rare occasion, expression in vulC and vulD was seen in mk103-148. The early expression for this construct was variable from line to line. mk153-154 shows variable expression in the presumptive vulE and vulA cells, although this expression is neither as weak nor as variable as that seen in mk103-148, mk84-20, mk82-100 and mk15-20.

Figure 3: Upstream regions that direct *egl-17* expression

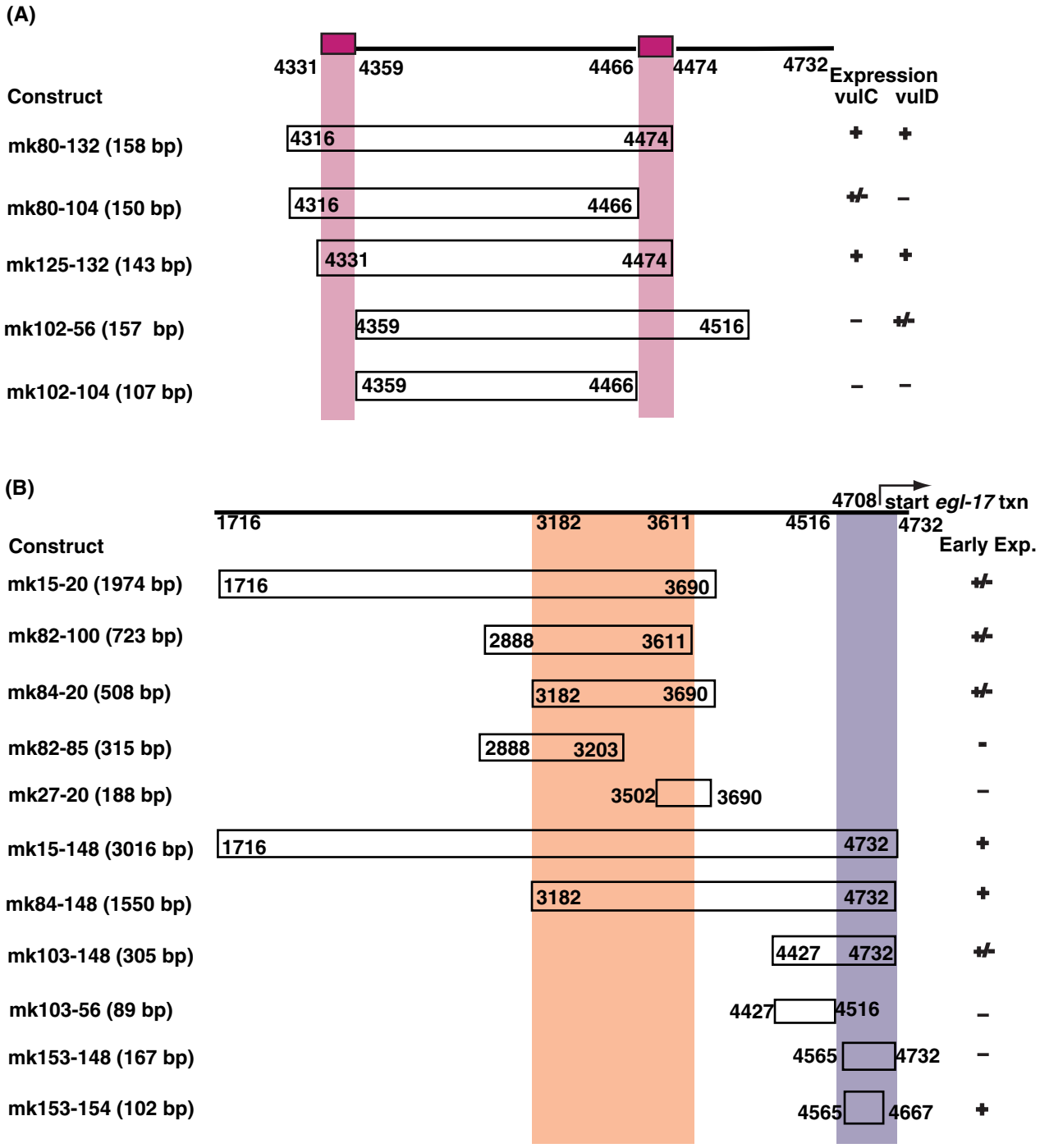
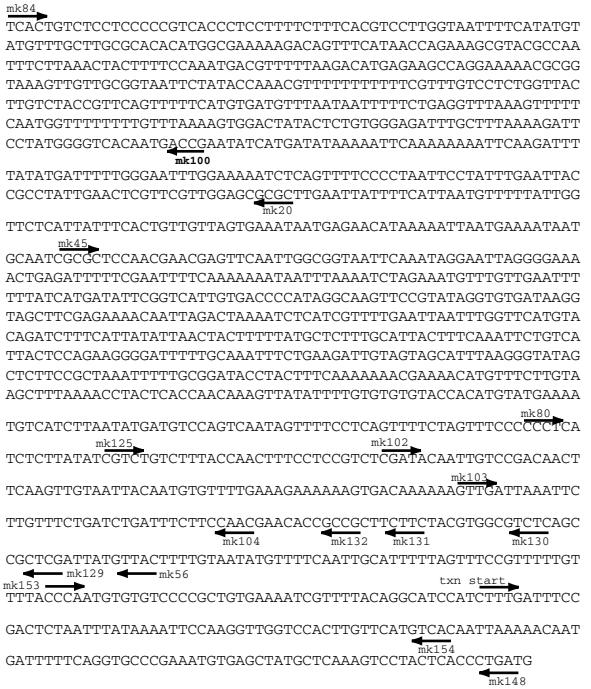


Figure 4: Upstream sequences of mk84-148, mk50-51, mk96-134 and mk66-67

Nucleotide sequences are shown for: the *egl-17* region mk84-148, with the translational start site in bold (A); *zmp-1* region mk50-51 (B); *cdh-3* region mk96-134 (C) ; and *cdh-3* region mk66-67 (D). Arrows show the end points and direction of primers in the regions.

Figure 4: Upstream sequences of mk84-148, mk50-51, mk96-134 and mk66-67

(A) *C. elegans egl-17* mk84-148



(B) *C. elegans zmp-1* mk50-51

(C) *C. elegans cdh-3* mk96-134

(D) *C. elegans cdh-3* mk66-67

Figure 5: Multiple regions direct *zmp-1* expression

The construct that confers GFP expression in vulE, vulA, and anchor cell (AC) at a similar level as the full-length reporter construct is shown at the top. The expression pattern of each construct shown is located at the right. The colored zones represent regions that confer GFP expression in a particular cell type: orange regions are those areas that contribute to vulA cell expression; yellow regions contribute to both vulE and vulA expression; blue regions important for driving expression in vulE; and purple regions are those regions that contribute to anchor cell expression. mk106-51 is the smallest construct that drives expression in the three cell types (depicted in green). The small boxes depict these regions graphically. If there is more then one box in a region, then that region is important for driving GFP expression in multiple cell types, depending on the surrounding DNA context. The sizes of each of these regions are listed in the appropriate box. A +/- indicates that either the expression level was reduced with respect to other constructs, or that not all animals showed consistent expression in the cell. A gap in the graphical depiction in construct mkΔ3/4 indicates an internal deletion.

Figure 5: Multiple Regions Direct *zmp-1* expression

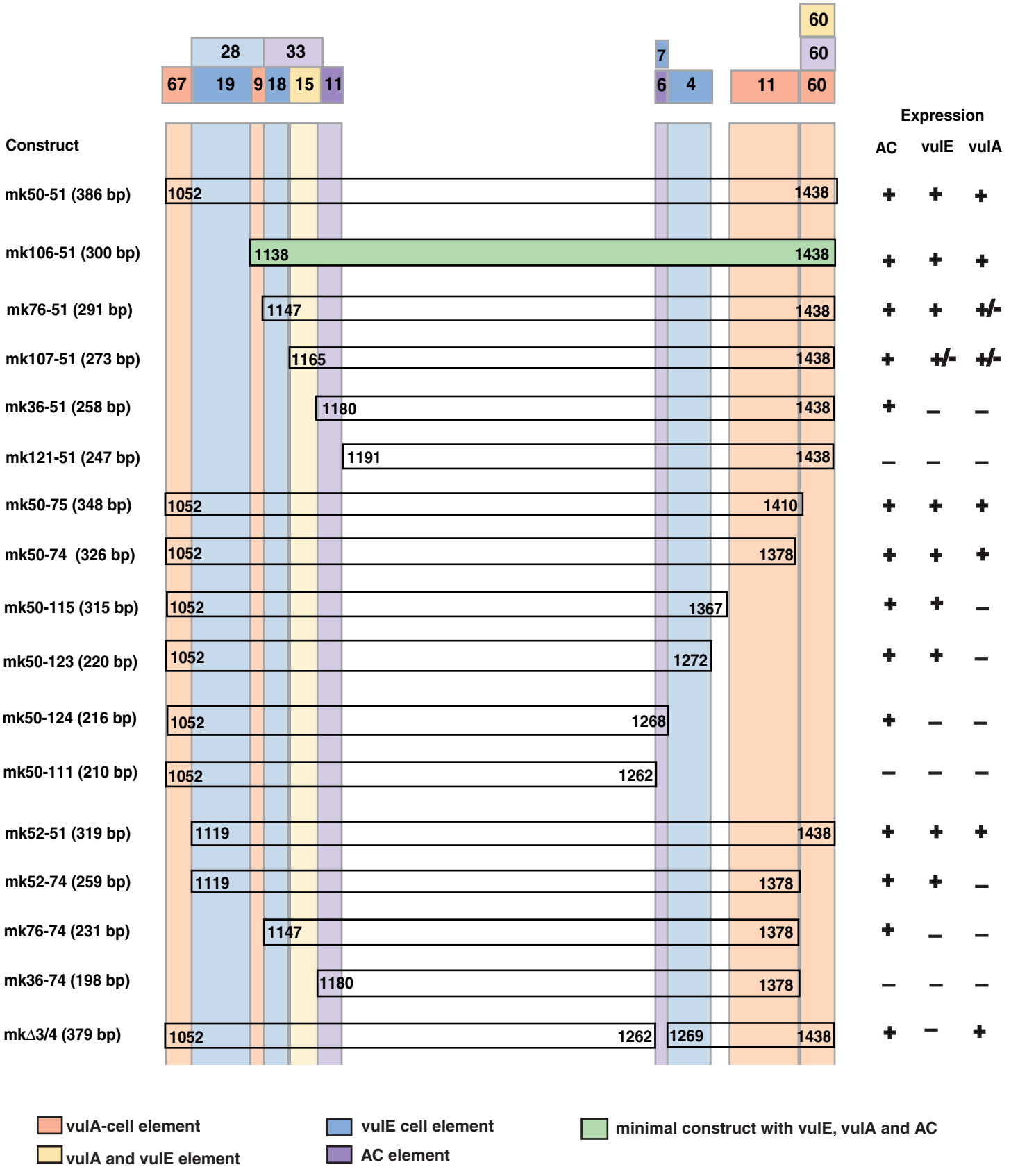


Figure 6: Regions that direct *cdh-3* expression

The name of the constructs, a graphical depiction of their location with respect to the full-length upstream region, and a summary of the expression of each construct is shown for each panel. These lists of constructs do not encompass all constructs that were made. A +/- indicates that either the expression level was reduced with respect to other constructs, or that not all animals showed consistent expression in the cell. (A) *cdh-3* constructs that illustrate the importance of two regions that direct the expression of the GFP in the anchor cell from the VPC 4-cell stage. The alpha, beta and gamma sub-regions are also shown. (B) The constructs that are listed illustrate the importance of two regions that confer expression of GFP in the anchor cell from the 2-cell stage of the VPC. (C) The constructs that are listed illustrate the importance of these two regions in directing the expression of the GFP in vulE, F, C and D. The first region, in blue, is bounded by nucleotides 2412-3101, and the second region, in yellow, is bounded by nucleotides 4434-4997. Construct mk66-156 shows variable weak expression in the occasional animal in vulC and vulD, while mk158-159 shows variable weak expression in the occasional animal in vulC, E and F, but never vulD.

Figure 6: Regions that direct *cdh-3* Expression

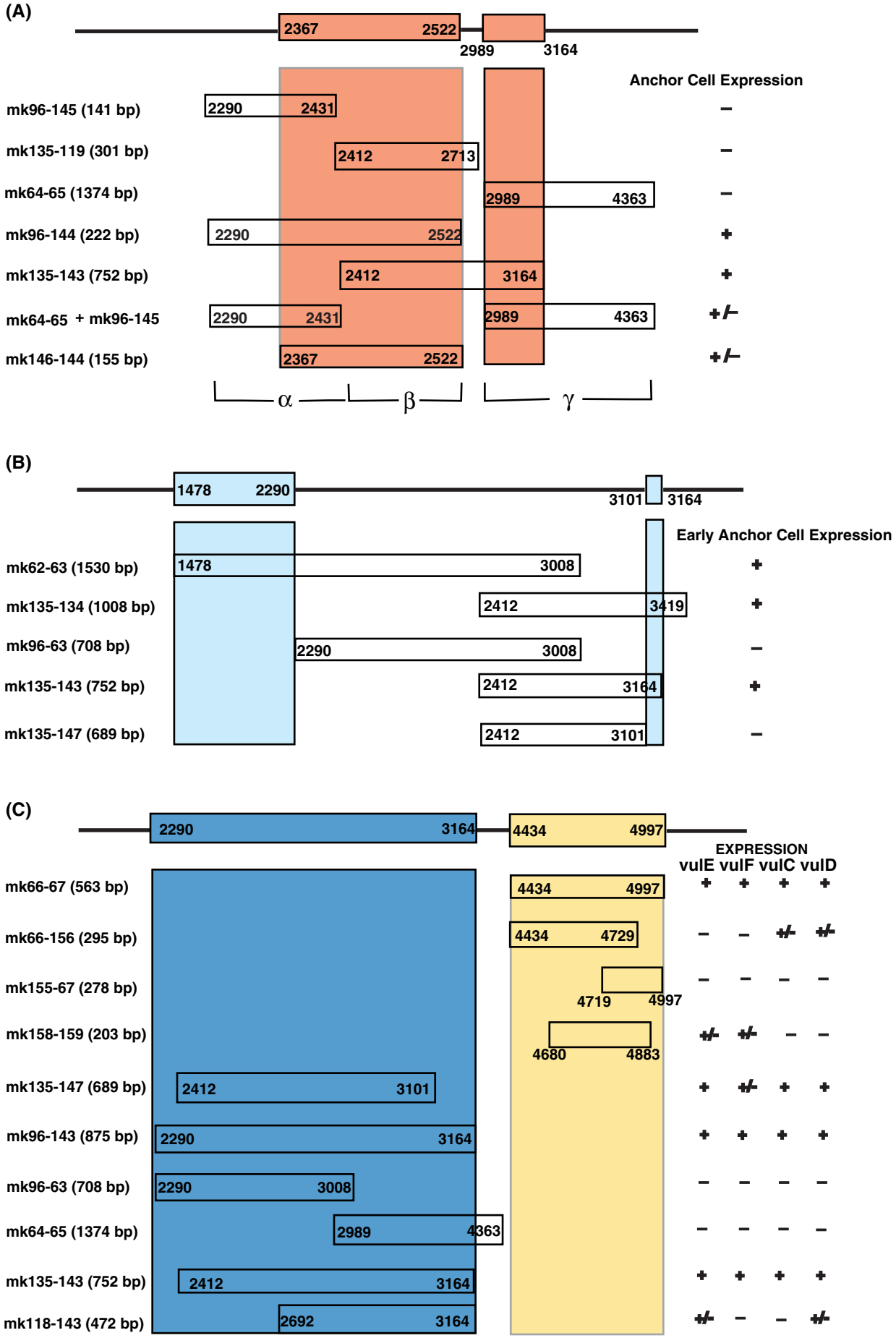


Table 1: Transfac binding site predictions for regions that confer cell-specific expression

The program MatInspector was used to make Transfac Database binding site predictions. Construct mk50-51 was compared to mk96-134, both of which can drive expression in the anchor cell. Construct mk96-134 was compared to mk66-67 since both of these *cdh-3* regions can confer vulva expression. Finally, *egl-17* mk84-148, *zmp-1* mk50-51, *cdh-3* constructs mk96-134 and mk66-67 were compared, since all of these constructs overlap in the vulva cells. Transfac prediction binding sites were listed that meet the following criteria: (1) the minimum core binding specificity had to be at least 0.90, and (2) the maximum Random Expectation Value, "re", which is the number of times this site would appear in a random 1000 bp, was not exceed 0.51. The number of sites in the first site is followed by a slash, and then the number of sites in second region. * These factors were not necessarily found in both, but were included because they are part of potentially interesting transcription families.

Table 1: Transfac binding site predictions for regions that confer cell-specific expression

		Anchor Cell	vulF, E, C and D	vulF, E, C and D
FACTOR FAMILY	FACTOR	mk50-51/ mk96-134	mk96134/ mk66-67	mk84-148/mk50-51/ mk96-134/mk66-67
ARS binding factor	ABF1.01	1/2		
cAMP-Responsive Element	ATF.02		1/1	1/0/1/1
cAMP -Responsive Element	CREB.01		1/1	1/0/1/1
cAMP -Responsive Element	CREB.03		1/1	
cAMP -Responsive Element	CREB.04		1/1	1/0/1/1
cAMP -Responsive Element	CREBP1.02		1/1	1/0/1/1
Cart-1 (cartilage homeoprotein)	CART1.01	1/1	1/1	4/1/1/1
CLOX FAMILY	CDP.01	1/1		1/1/1/0
CLOX FAMILY	CDPCR3.01		1/1	3/1/1/1
Enhancer-CcAaT binding factors	NFY.01			2/3/2/0
Enhancer-CcAaT binding factors	NFY.02			2/1/2/1
ETS	PU.1ETS *		2/1	2/0/2/1
EVI myeloid transforming protein	EVI1.02	1/2	2/3	0/1/3/2
Floral determination	MADSA.01	1/4		2/1/5/0
Fork Head and Related	FREAC2.01			2/1/4/0
Fork Head and Related	FREAC4.01			1/1/2/0
Fork Head and Related	HFH1.01			1/1/2/0
Fork Head and Related	HNF1.01	1/1		1/1/1/0
Glucocorticoid Responsive	GRE.01		2/1	
Glucocorticoid Responsive	PRE.01	1/1		
Homeodomain Factor	FTZ.01		3/4	4/0/3/1
Homeodomain Factor	NKX25.02	1/7	8/1	6/1/8/1
Homeodomain Factor	NKX31.01		1/3	
Homeodomain Factor	PBX1.01	1/3		
Homeodomain Factor myeloid lukemia	MEIS1.01	1/1		3/1/1/0
Homeodomain Pancreatic /Intestinal (LIM domain family)	ISLI1.01	1/2	2/2	5/1/2/2
Homeoprotein Caudel	CDX2.01	1/4	4/2	5/1/4/2
HSF family	FHSF.03	1/2		2/1/2/0
HSF family	FHSF.04			2/1/0/1
Human muscle-specific Mt binding site	MTBF.01		2/2	2/0/2/2
Interferon Regulated Factor	IRF1.01	1/2	2/1	2/1/2/1
Interferon Regulated Factor	IRF2.01		1/1	2/1/1/1
Interferon Regulated Factor	ISRE.01			1/1/1/0
MYB-Like protein (Petunia)	MYBPH3.01		1/1	
Octamer Family	OCT.01			2/1/2/2
Octamer Family	OCT1.01		3/2	3/1/1/2
Octamer Family	OCT1.02			2/2/0/2
Octamer Family	OCT1.03		3/3	7/0/3/3

Octamer Family	OCT1.04		1/2	2/1/1/2
Octamer Family	OCT1.05			4/0/1/2
Octamer Family	OCT1.06		3/3	10/0/3/3
Paired homeodomain factors	PAX6 HD.01	1/1		6/1/2/0
Phaseolus vulg. SILencer reg. of chalcone synth. prom.	SBF1.01	1/4		3/1/4/0
Plant P-Box binding sites	PBOX.01	1/1	1/1	0/1/1/1
Poly A	APOLYA.01		2/2	3/0/2/2
Poly A	POLYA.01		1/2	1/0/1/2
Promoter-CcAaT binding factors	ACAAT.01		1/2	1/0/1/2
Repr. of RXR-mediated & retinoic acid responses	COUP.01		1/1	
signal transducers and activators of txn	STAT1.01		1/1	
signal transducers and activators of txn	STAT3.01		1/1	1/0/1/1
SMAD Family TGF-B	FAST1.01	1/2		
Special AT rich binding Sequence	SATB1.01	1/1	1/1	0/1/1/1
TATA FAMILY	TATA.02		1/2	9/0/1/2
Tata-Binding Protein Factor	ATATA.01			
TCF/LEF	LEF1.01 *	2/1	2/1	2/1/2/1
Vertebrate steroidogenic factor	SF1.01		1/1	
Yeast CCAAT binding factors	HAP234.01			4/0/1/1
Zea mays Transcriptional activator OPAQue-2	O2.03		1/1	
zinc finger POZ domain B-Cells	BCL6.02	1/1	1/1	0/1/1/1
zinc finger W Box family	WRKY.01		2/1	1/0/2/1
zinc finger *Xenopus* MYT1 C2HC	MYT1.01		4/1	
zinc finger *Xenopus* MYT1 C2HC	MYT1.02			5/0/4/1
C. elegans maternal gene	SKN1.01		1/1	2/0/1/1

Table 2: AlignACE predictions of overrepresented sequences

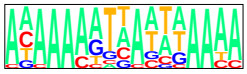
(A) A summary table of the number of motifs found in each of the listed regions. The total number of motifs identified by AlignACE is shown in parentheses, while the number of motifs that scored above the MAP score threshold of 10 is shown outside the parentheses for both the eight- and 10-bp motifs. The left-hand column identifies the cell-type expression of interest when two or more regions were compared. * Indicates that all four regions, mk84-148, mk50-51, mk96-134, and mk66-67 were compared. (B) This table summarizes the data for each of the motifs listed in Table 2A that had a MAP score over 10. The region is listed in the left-hand column. The motif numbers are consecutive and are followed by the size of the motif. The MAP score for each motif is shown under the column head MAP. The sites for each motif are listed. If more than one region was compared, the sites for the first, as indicated by the left-hand column, are in parentheses, followed by the second set of sites in parentheses, and so on. Abbreviations are as follows: expr., expression; imp., importance and; elem. element. The pictograms were generated using the Pictogram program (http://genes.mit.edu/pictogram.html).

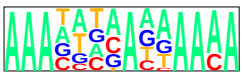
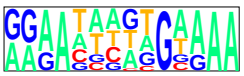
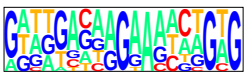
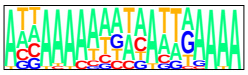
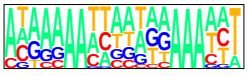
Table 2: AlignACE predictions of overrepresented sequences

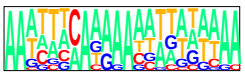
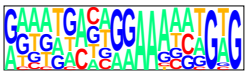
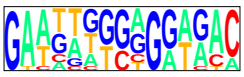
A.

Expression	Regions examined	Gene	8 bp motif	10 bp motif
	mk84-148	*egl-17*	5 (14)	3 (6)
	mk50-51	*zmp-1*	1 (5)	0 (2)
	mk96-134	*cdh-3*	2 (12)	1 (8)
	mk66-67	*cdh-3*	0 (5)	0 (7)
Vulval cdh-3	mk96-134/mk66-67	*cdh-3*	0 (8)	0 (12)
Anchor cell	mk50-51/96-134	*zmp-1/cdh-3*	2 (10)	4 (11)
Vulval general	*all	*all	5 (12)	8 (15)

B.

Region	Motif	MAP	Consensus	Sites	Comments
mk84-148 *egl-17*	1.8	17.15		124, 269, 395, 419, 433, 455, 510, 581, 646, 667, 683, 708, 876, 917, 1111, 1158, 1232, 1248, 1327	Sites 1158 & 1232 are located between mk125 & mk102 and mk102 & mk103 regions respectively imp. for vulC/D expr.
	2.8	12.38		161, 217, 277, 295, 433, 453, 647, 667,874, 917, 974, 1110, 1157,1413, 1502	Site 1502 is located between mk154 & mk148 region important fidelity early expression.
	3.8	11.88		124, 394, 405, 471, 486, 510, 594, 618, 633, 686, 712, 818, 876, 1252, 1355	Overlaps with multiple motif 1.8 sites.
	4.8	11.08		74, 160, 214, 273, 296, 392, 417, 518, 581, 683, 854, 916, 992, 1230, 1336, 1361, 1379	Site 1230 is between mk102 & mk103 region imp. for vulC/D expression. Overlaps with multiple motif 2.8 sites.
	5.8	10.79		176, 190, 365, 374, 392, 453, 475, 631, 653, 733, 742, 1127, 1187, 1380, 1434	Sites 1380, 1434,1187 betw. mk153 & mk154, & several in distal region imp. for early expr. Site 1127 betw. mk102 & mk103 region imp. for vulC/D expr.
	6.10	19.87		212, 267, 296, 392, 417, 518, 579, 681, 706, 816, 854, 1109, 1230, 1325, 1361	Multiple sites overlap motif 1.8 sites.
	7.10	19.69		49, 158, 218, 296, 430, 455, 518, 540, 562, 580, 643, 668, 776, 875, 910, 991, 1064, 1110, 1367	Multplie sites overlap with motif 2.8 sites.
	8.10	13.91		124, 289, 392, 403, 471, 486, 510, 592, 616, 631, 684, 712, 818, 876, 1252	Multiple sites overlap with motif 3.8 sites.
mk50-51 *zmp-1*	1.8	11.30		16, 31, 50, 58, 101, 110, 123, 173, 182, 239, 316, 342	Site 316 between mk74 & mk115, region imp. for vulA expression. Site 110 between mk107 & mk36, region imp. for vulA and vulE expr.

Region	Motif	MAP	Consensus	Sites	Comments
	1.8	17.17		104, 129, 148, 166, 555, 687, 710, 750, 770, 812, 842, 1112	Multiple sites in alpha, beta, and gamma. Site could be important for vuval expression (few sites between mk136 & mk164).
mk96-134 *cdh-3*	2.8	10.62		64, 91, 128, 369, 380, 416, 438, 471, 606, 764, 811, 836, 866, 1035	Multiple sites in alpha, beta, and gamma.
	3.10	16.99		103, 129, 147,166, 687, 709, 749, 770, 812, 846	All sites overlap motif1.8 sites.
	1.8	16.08		[116, 285] [319, 333, 353, 418, 610, 957] [35] [397, 669, 689, 816, 885, 1348, 1487]	
mk50-51 *zmp-1*	2.8	16.04		[105, 169, 298] [38, 388, 623, 1056, 1098] [31, 482] [213, 277, 395, 582, 706, 864, 1110, 1232]	
mk96-134 *cdh-3* mk66-67 *cdh-3*	3.8	12.14		[91] [89, 463, 600, 720, 833, 1026, 1055] [205, 354, 390, 483] [160, 218, 296, 432, 454, 582, 646, 875, 912, 1120, 1158]	
mk84-148 *egl-17*	4.8	11.75		[25, 95] [94,113, 465, 770] [ ] [1, 114, 160, 230, 448, 647, 1115, 1220, 1531]	No sites present in mk66-67
	5.8	10.98		[278, 372] [81, 156, 324, 416, 610] [171, 261, 382] [112, 169, 496, 656, 918, 1113, 1255, 1267, 1516]	
	6.10	21.39		[51, 168, 239, 285] [42, 321, 341, 381, 610, 1081] [35] [49, 212, 255, 296, 390, 518, 576, 678, 1227, 1360]	
	7.10	21.09		[88, 236, 283, 316] [40, 254, 351, 404, 512, 718, 952, 1023] [29, 202] [167, 212, 293, 392, 414, 451, 527, 579, 643, 681, 705, 862, 923, 1108, 1227]	
	8.10	16.85		[31, 173, 360] [38, 120, 180, 263, 384, 472, 619, 710, 1036] [31, 336, 548] [258, 395, 570, 674, 708, 832, 1011, 1232, 1477]	

Region	Motif	MAP	Consensus	Sites	Comments
	9.10	15.47		[91, 239, 352] [38, 374, 484, 945,1026] [60] [29, 269, 296, 395, 417, 510, 582, 674, 698, 856, 916, 1234]	
mk50-51 *zmp-1*	10.10	15.42		[25, 95, 158] [71, 113, 166, 465, 604, 770] [356, 477] [21, 159, 448, 556, 646, 1115]	
mk96-134 *cdh-3*	11.10	15.36		[6] [0, 743, 845] [454, 497, 525]	
mk66-67 *cdh-3*				[72, 367, 479, 622, 735, 1291, 1467, 1514]	
mk84-148 *egl-17*	12.10	15.29		[25, 280, 320] [81, 324, 416, 593, 612, 979] [157, 263] [114, 448, 656, 702, 920, 1115, 1516]	
	13.10	11.55		[275] [15, 63, 100, 685, 765, 840, 897] [195] [5, 452, 647, 1110, 1127, 1500]	
	1.8	11.60		[126] [65, 105, 129, 149, 166, 687, 751, 770, 812]	Site 126 between mk36 & mk121 imp. for ac expression in mk50-51. Mulitple sites in mk96-134 regions alpha, beta and gamma.
	2.8	10.76		[91, 239, 286] [51, 148, 238, 257, 302, 333, 354, 374, 948, 964]	Site 91 between mk76 & mk36 imp. for ac expression in mk50-51. Multiple sites in mk96-134 regions alpha, beta, and gamma.
mk50-51 *zmp-1*	3.10	15.10		[53, 105, 173, 286, 359] [47, 133, 180, 268, 330, 354, 383, 618, 816, 920, 1040, 1097]	Site 105 between mk76 & mk36 imp. for ac expression in mk50-51. Multiple sites in mk96-134 regions alpha, beta, and gamma.
mk96-134 *cdh-3*	4.10	10.63		[54, 112, 173] [120, 339, 370, 384, 417, 472, 619, 846, 867, 952, 1036]	Site 112 between mk76 & mk36 imp. for ac expression in mk50-51. Multiple sites in mk96-134 regions alpha, beta, and gamma.
	5.10	10.63		[126] [65, 100, 129, 166, 687, 812]	Multiple sites overlap with motif 1.8 sites. See 1.8 comments.
	6.10	9.92		[50, 101, 168, 184, 239, 286] [37, 72, 263, 333, 374, 948, 964, 1027, 1055]	Site 101 between mk76 & mk36 imp. for ac expression in mk50-51. Multiple sites in mk96-134 regions alpha, beta, and gamma.

Supplemental Table 1: PCR primers

The PCR primers used to generate the constructs that were analyzed in this study are listed.

Supplemental Table 1: PCR Primers

PRIMER	GENE	SITE	SEQUENCE OF PRIMER
mk01	egl-17	SphI	5' CCC CCG CAT GCC ACT ATA GAA TAC ATA GGA TC 3'
mk02	egl-17	SalI	5' CCC CCG TCG ACT TTT CAC AGC GGG GAC ACA CAT TGG 3'
mk09	zmp-1	SphI	5' CCC CCG CAT GCG TGT TTA ATT TTG ACC CAA AGA TGC 3'
mk15	egl-17	SphI	5' CCC CCG CAT GCC CAT CTT ACG GTT ATA TTC 3'
mk16	egl-17	StuI	5' CCC CCA GGC CTG GAA TAT AAC CGT AAG ATG G 3'
mk20	egl-17	StuI	5' CCC CCA GGC CTG CGC GCT CCA ACG AAC GAG 3'
mk27	egl-17	SphI	5' CCC CCG CAT GCG TGG ACT ATA CTC TGT GGG 3'
mk29	zmp-1	SphI	5' CCC CCG CAT GCC TTG AAT CTA GCT ATA TGT AG 3'
mk30	zmp-1	XbaI	5' CCC CCT CTA GAC TAC ATA TAG CTA GAT TCA AG 3'
mk31	zmp-1	SphI	5' CCC CCG CAT GCC AGT AAC CAA GCA CTC GTT ATC 3'
mk32	zmp-1	XbaI	5' CCC CCT CTA GAG ATA ACG AGT GCT TGG TTA CTG 3'
mk33	zmp-1	SphI	5' CCC CCG CAT GCC ATA TGC TAC CTT CAC CAG C 3'
mk34	zmp-1	XbaI	5' CCC CCT CTA GAG CTG GTG AAG GTA GCA TAT G 3'
mk35	zmp-1	XbaI	5' CCC CCT CTA GAG CTG ACT CAT TAG CAC AAG AC 3'
mk36	zmp-1	SphI	5' CCC CCG CAT GCC TGC CAG ATG CAA ACT GAT TC 3'
mk37	zmp-1	XbaI	5' CCC CCT CTA GAG AAT CAG TTT GCA TCT GGC AG 3'
mk45	egl-17	HindIII	5' CCC CCA AGC TTC GCG CTC CAA CGA ACG AGT TC 3'
mk48	egl-17	SphI	5' CCC CCG CAT GCG CTT ACA AGA AAC ATG TTT TC 3'
mk49	egl-17	SphI	5' CCC CCG CAT GCC ACA GCG GGG ACA CAC ATT GG 3'
mk50	zmp-1	SphI	5' CCC CCG CAT GCG AAG GAC GGC TCG TTG AAC AG 3'
mk51	zmp-1	XbaI	5' CCC CCT CTA GAC TAG AAA ATG GCC AAT ATG C 3'
mk52	zmp-1	SphI	5' CCC CCG CAT GCG ATC ATC AAA GTA TTC GAG 3'
mk53	zmp-1	XbaI	5' CCC CCT CTA GAC TAC AAC TGA ATA CTA CTA CTA CGA C 3'
mk54	zmp-1	XbaI	5' CCC CCT CTA GAC AAG CAA TAC GTA CAC ATG 3'
mk55	zmp-1	XbaI	5' CCC CCT CTA GAG CGA TGA CCA TCT TAA GTG GCG 3'
mk56	egl-17	XbaI	5' CCC CCT CTA GAG TAA CAT AAT CGA GCG GCT GAG 3'
mk57	egl-17	SphI	5' CCC CCG CAT GCG CAT TTA AGG GTA TAG CTC TTC CC 3'
mk58	cdh-3	SphI	5' CCC CCG CAT GCG GAG GGT ACC ATG GCC ATC CC 3'
mk59	cdh-3	XbaI	5' CCC CCT CTA GAG CGG AAC ATC GAT TCT ATG G 3'
mk60	cdh-3	SphI	5' CCC CCG CAT GCC CAT AGA ATC GAT GTT CCG C 3'
mk62	cdh-3	SphI	5' CCC CCG CAT GCC TAG AGC ATG ATG TCC TTA CC 3'
mk63	cdh-3	XbaI	5' CCC CCT CTA GAG GGA CGG TCG GTC CGT TTT G 3'
mk64	cdh-3	SphI	5' CCC CCG CAT GCC AAA ACG GAC CGA CCG TCC C 3'
mk65	cdh-3	XbaI	5' CCC CCT CTA GAC ACT AGT TAC TCC AAC TGA TC 3'
mk66	cdh-3	SphI	5' CCC CCG CAT GCG TGA AAG CTC CAG GGA GCT G 3'
mk67	cdh-3	XbaI	5' CCC CCT CTA GAC AGA TGG TCC CAA GAA ACG C 3'
mk68	cdh-3	SphI	5' CCC CCG CAT GCG CGT TTC TTG GGA CCA TCT G 3'
mk69	cdh-3	XbaI	5' CCC CCT CTA GAG TCA TCT ATT CAG CAT TGA TC 3'
mk70	zmp-1	SphI	5' CCC CCG CAT GCC GCC ACT TAA GAT GGT CAT CGC 3'
mk71	zmp-1	SphI	5' CCC CCG CAT GCC ATG TGT ACG TAT TGC TTG 3'
mk72	zmp-1	SphI	5' CCC CCG CAT GCG TAG AAG GGT ATT AGT CGT AG 3'
mk73	zmp-1	SphI	5' CCC CCG CAT GCC AGT TGT AGT AAT ATA TAT TTC 3'
mk74	zmp-1	XbaI	5' CCC CCT CTA GAC GTT TTC TCT TCT GAA AAG TCC 3'
mk75	zmp-1	XbaI	5' CCC CCT CTA GAC TGT CAC ATT TTC CAT TC 3'
mk76	zmp-1	SphI	5' CCC CCG CAT GCC ACT GGT TTC TGT TCT TTC CG 3'

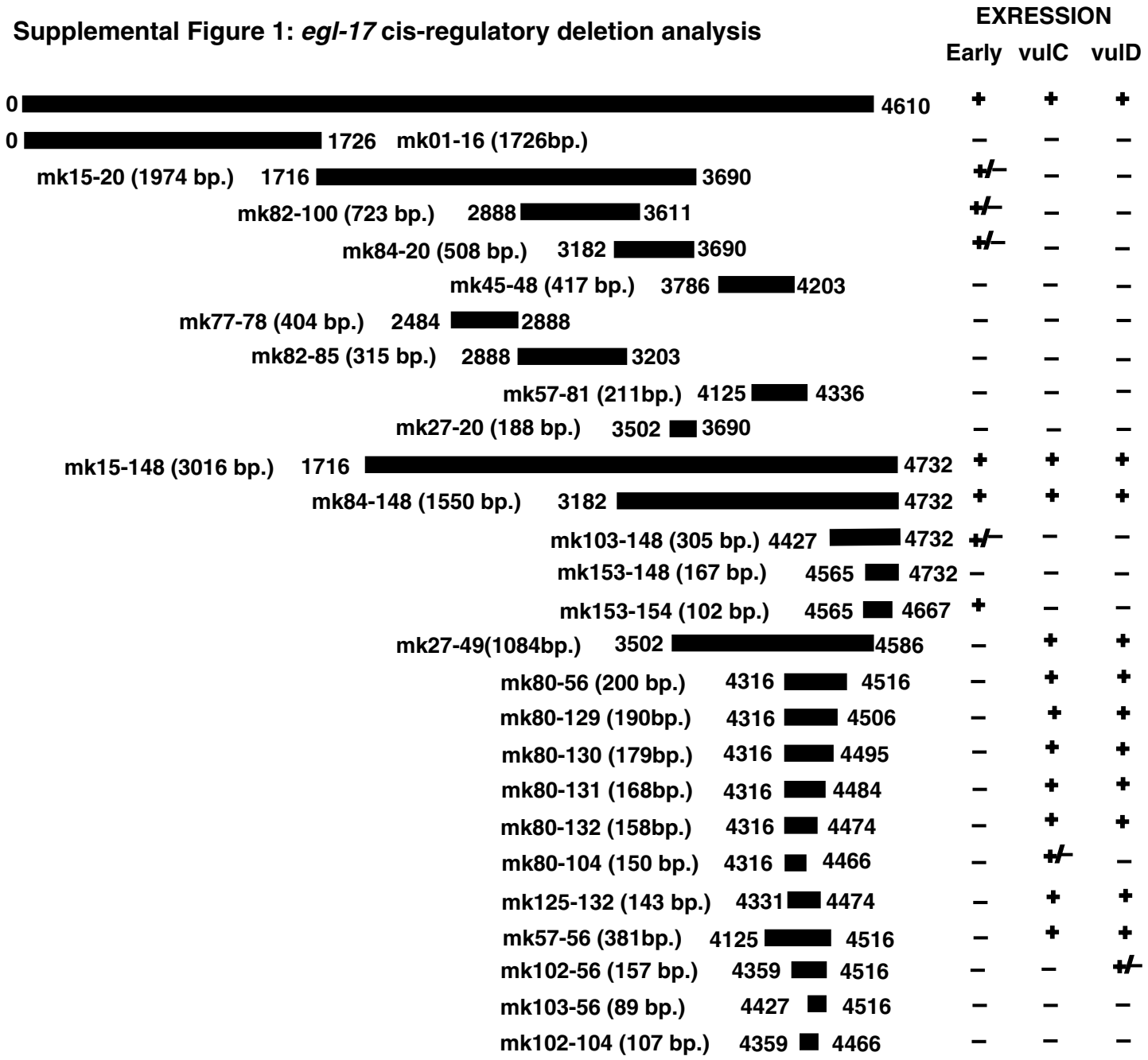
mk77	egl-17	SphI	5' CCC CCG CAT GCG TCT GCT GCC TCG CCT CAT CG 3'
mk78	egl-17	XbaI	5' CCC CCT CTA GAC TAT GTT TCT AGA GAA TTT TG 3'
mk79	zmp-1	none	5' CTA CGA CTA ATA CCC TTC TAC GAG AAA TTA AAA ACG GAA AG 3'
mk80	egl-17	SphI	5' CCC CCG CAT GCC CTC ATC TCT TAT ATC GTC TG 3'
mk81	egl-17	XbaI	5' CCC CCT CTA GAC AGA CGA TAT AAG AGA TGA GG 3'
mk82	egl-17	SphI	5' CCC CCG CAT GCG TAT TAC ATT CCC TAT CAG TC 3'
mk84	egl-17	SphI	5' CCC CCG CAT GCC ACT GTC TCC TCC CCC GTC ACC 3'
mk85	egl-17	XbaI	5' CCC CCT CTA GAG GTG ACG GGG GAG GAG ACA GTG 3'
mk87	zmp-1	none	5' CAT TCA TAA AAT ATT TCG ACC TTT TTC TTG TTA CTC TTT TTT TC 3'
mk89	zmp-1	none	5' GAA AAA AAG AGT AAC AAG AAA AAG GTC GAA ATA TTT TAT GAA TG 3'
mk92	zmp-1	none	5' GCA TGC GTA GAA GGG TAT TAG TCG TAG TAG TAG TAT TCA GTT GTA GTC TAG A 3'
mk93	zmp-1	none	5' TCT AGA CTA CAA CTG AAT ACT ACT ACT ACG ACT AAT ACC CTT CTA CGC ATG C 3'
mk96	cdh-3	SphI	5' CCC CCG CAT GCC CGC ATT TTC ATC AAG ATT CC 3'
mk97	cdh-3	XbaI	5' CCC CCT CTA GAG GAA TCT TGA TGA AAA TGC GG 3'
mk98	cdh-3	StuI	5' CCC CCA GGC CTC AGC TCC CTG GAG CTT TCA C 3'
mk100	egl-17	XbaI	5' CCC CCT CTA GAC GGT CAT TGT GAC CCC ATA GG 3'
mk102	egl-17	SphI	5' CCC CCG CAT GCC GAT ACA ATT GTC CGA CAA C 3'
mk103	egl-17	SphI	5' CCC CCG CAT GCG TTG ATT AAA TTC TTG TTT C 3'
mk104	egl-17	XbaI	5' CCC CCT CTA GAG TTG GAA GAA ATC AGA TCA G 3'
mk105	zmp-1	SphI	5' CCC CCG CAT GCC AAA GTA TTC GAG TAC GTT TAC 3'
mk106	zmp-1	SphI	5' CCC CCG CAT GCG TAC GTT TAC ACT GGT TTC TG 3'
mk107	zmp-1	SphI	5' CCC CCG CAT GCC CGT TTT TAA TTT CTC CTG CC 3'
mk108	zmp-1	SphI	5' CCC CCG CAT GCG AAA AAA AGA GTA ACA AG 3'
mk109	zmp-1	SphI	5' CCC CCG CAT GCG TAT TAG TCG TAG TAG TAG 3'
mk110	zmp-1	XbaI	5' CCC CCT CTA GAC CCT TCT ACT TTT TCT TGT TAC 3'
mk111	zmp-1	XbaI	5' CCC CCT CTA GAC TAC GAC TAA TAC CCT TCT AC 3'
mk112	zmp-1	SphI	5' CCC CCG CAT GCG TAA CAA GAA AAA GTA GAA G 3'
mk113	zmp-1	XbaI	5' CCC CCT CTA GAG CAA AAA AAA ACT GTC ACA TTT TCC 3'
mk114	zmp-1	XbaI	5' CCC CCT CTA GAG TAA GTA TTT TAT AAA GCT G 3'
mk115	zmp-1	XbaI	5' CCC CCT CTA GAC TGA AAA GTC CAC CAA AAA ATT 3'
mk116	zmp-1	XbaI	5' CCC CCT CTA GAC AAA AAA TTA ATT GAA AAT TGC G 3'
mk117	zmp-1	XbaI	5' CCC CCT CTA GAC TCT TTT TTT CAA GCA ATA C 3'
mk118	cdh-3	SphI	5' CCC CCG CAT GCG ACG ACA ACT TAA TGA AAT TTG 3'
mk119	cdh-3	XbaI	5' CCC CCT CTA GAC AAA TTT CAT TAA GTT GTC GTC 3'
mk120	zmp-1	SphI	5' CCC CCC GTA CGC TGT TCT TTC CGT TTT TTA ATT TC 3'
mk121	zmp-1	SphI	5' CCC CCC GTA CGC AAA CTG ATT CAT TGT GTA CG 3'
mk122	zmp-1	SphI	5' CCC CCC GTA CGA GTA GAA GGG TAT TAG TCG TAG 3'
mk123	zmp-1	XbaI	5' CCC CCT CTA GAA ATA CTA CTA CTA CGA CTA ATA C 3'
mk124	zmp-1	XbaI	5' CCC CCT CTA GAC TAC TAC TAC GAC TAA TAC CC 3'
mk125	egl-17	SphI	5' CCC CCC GTA CGC GTC TGT CTT TAC CAA CTT TC 3'
mk129	egl-17	XbaI	5' CCC CCT CTA GAC GAG CGG CTG AGA CGC CAC G 3'
mk130	egl-17	XbaI	5' CCC CCT CTA GAG ACG CCA CGT AGA AGA AGC GG 3'
mk131	egl-17	XbaI	5' CCC CCT CTA GAG AAG AAG CGG CGG TGT TCG TTG 3'
mk132	egl-17	XbaI	5' CCC CCT CTA GAC GGT GTT CGT TGG AAG AAA TC 3'
mk133	cdh-3	XbaI	5 'CCC CCT CTA GAG AAG CAA GAC TGT TGA CAG C3'
mk134	cdh-3	XbaI	5 'CCC CCT CTA GAC ATA AAA CTG CCC GGC TTT G3'
mk135	cdh-3	SphI	5 'CCC CCG CAT GCC CTG TCC CAC TTG CCC ATT C3'

mk136	cdh-3	SphI	5 'CCC CCG CAT GCG TTT ATG TGT CAT GAA TAA TG3'
mk137	cdh-3	XbaI	5 'CCC CCT CTA GAC ATT TCA AAA ATT AAA AAT CAG3'
mk138	zmp-1	SphI	5' CCC CCG CAT GCA ATA TTT CGA CGT TTT CTC TTC 3'
mk141	zmp-1	none	5' GTA GAA GGG TAT TAG TCG TAT ATT CAG TTG TAG TAA TAT ATA TTT C 3'
mk142	zmp-1	none	5' GAA ATA TAT ATT ACT ACA ACT GAA TAT ACG ACT AAT ACC CTT CTA C 3'
mk143	cdh-3	XbaI	5' CCC CCT CTA GAG ACA TGC ATT GGT GCA AGT TG 3'
mk144	cdh-3	XbaI	5' CCC CCT CTA GAC ATT ATT CAT GAC ACA TAA AC 3'
mk145	cdh-3	XbaI	5' CCC CCT CTA GAG AAT GGG CAA GTG GGA CAG G 3'
mk146	cdh-3	SphI	5' CCC CCG CAT GCC TCA GAA ATC GGA GAA TGA TTG 3'
mk147	cdh-3	XbaI	5' CCC CCT CTA GAG TAG TGT AGT CAT AGA GGT CCG 3'
mk148	egl-17	StuI	5' CCC CCA GGC CTC ATC AGG GTG AGT AGG ACT TTG 3'
mk151	zmp-1	none	5' CGT ACA CAT GAA TCA GTT TGG AAA TTA AAA ACG GAA AGA AC 3'
mk152	zmp-1	none	5' GTT CTT TCC GTT TTT AAT TTC CAA ACT GAT TCA TGT GTA CG 3'
mk153	egl-17	SphI	5' CCC CCG CAT GCC CCA ATG TGT GTC CCC GCT G 3'
mk154	egl-17	XbaI	5' CCC CCT CTA GAG TGA CAT GAA CAA GTG GAC C 3'
mk155	cdh-3	SphI	5' CCC CCG CAT GCC ACA CCA CAC AGT GCC AAT TG 3'
mk156	cdh-3	XbaI	5' CCC CCT CTA GAC AAT TGG CAC TGT GTG GTG TG 3'
mk158	cdh-3	SphI	5' CCC CCG CAT GCC CTT GAT TCTCTT GTA TAT CC 3'
mk159	cdh-3	XbaI	5' CCC CCT CTA GAG GAC AGA GAATGC AAA TTA GC 3'

Supplemental Figure 1: *egl-17 cis*-regulatory deletion analysis

The upstream region of *egl-17* is depicted at the top of this figure. The translational start occurs at nucleotide 4610. A +/- indicates that either the expression level was reduced with respect to other constructs, or that not all animals showed consistent expression in the cell. Mk80-104 showed very weak vulC expression in 1/2 lines. 102-56 showed weak expression in vulD in 3/3 lines. On rare occasion expression in vulC and vulD was seen in mk103-148. The early expression for this construct was variable from line to line. mk153-154 shows variable expression in the presumptive vulE and vulA cells, although this expression is neither as weak nor as variable as that seen in mk103-148, mk84-20, mk82-100 and mk15-20.

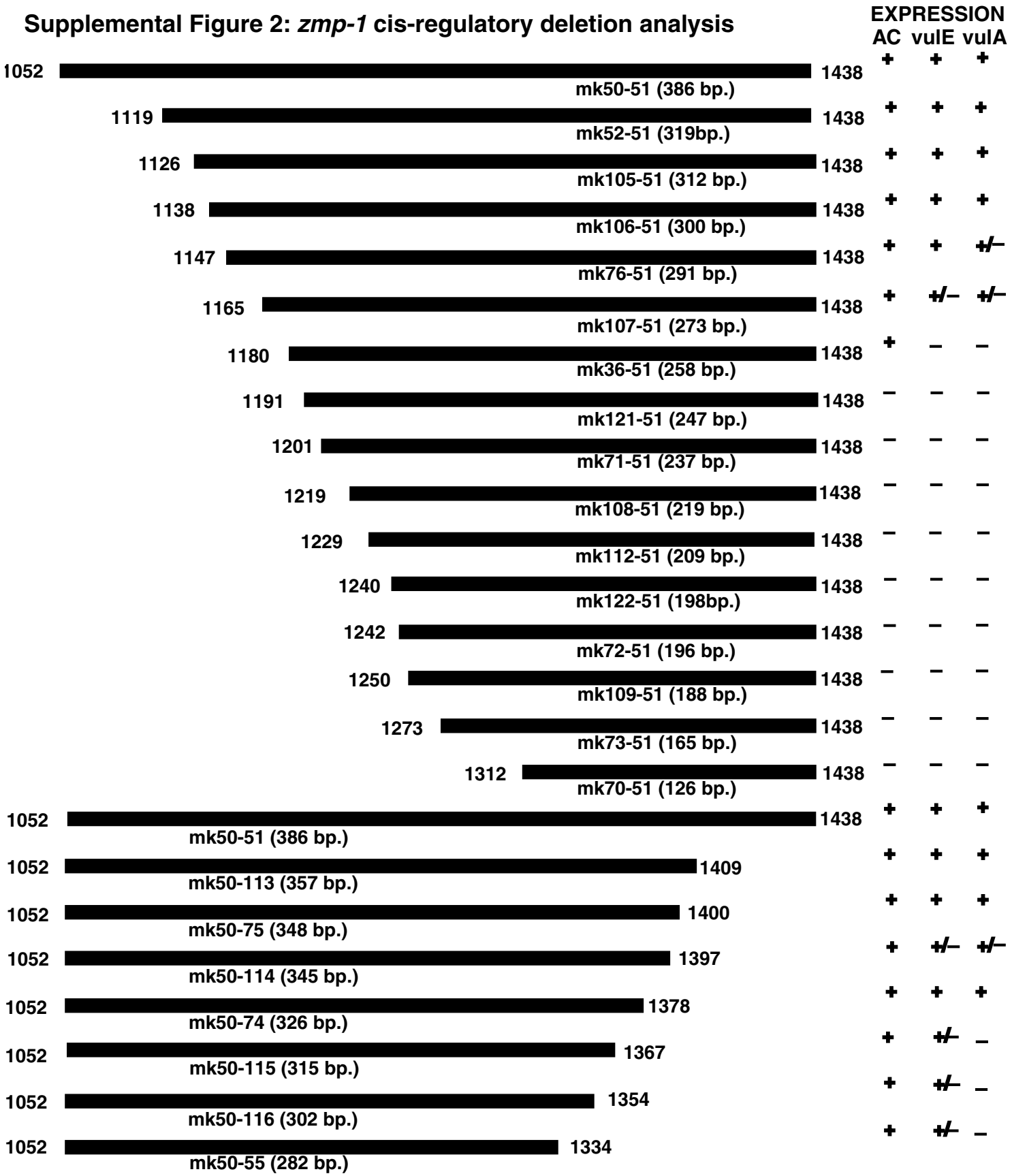
Supplemental Figure 1: *egl-17* cis-regulatory deletion analysis

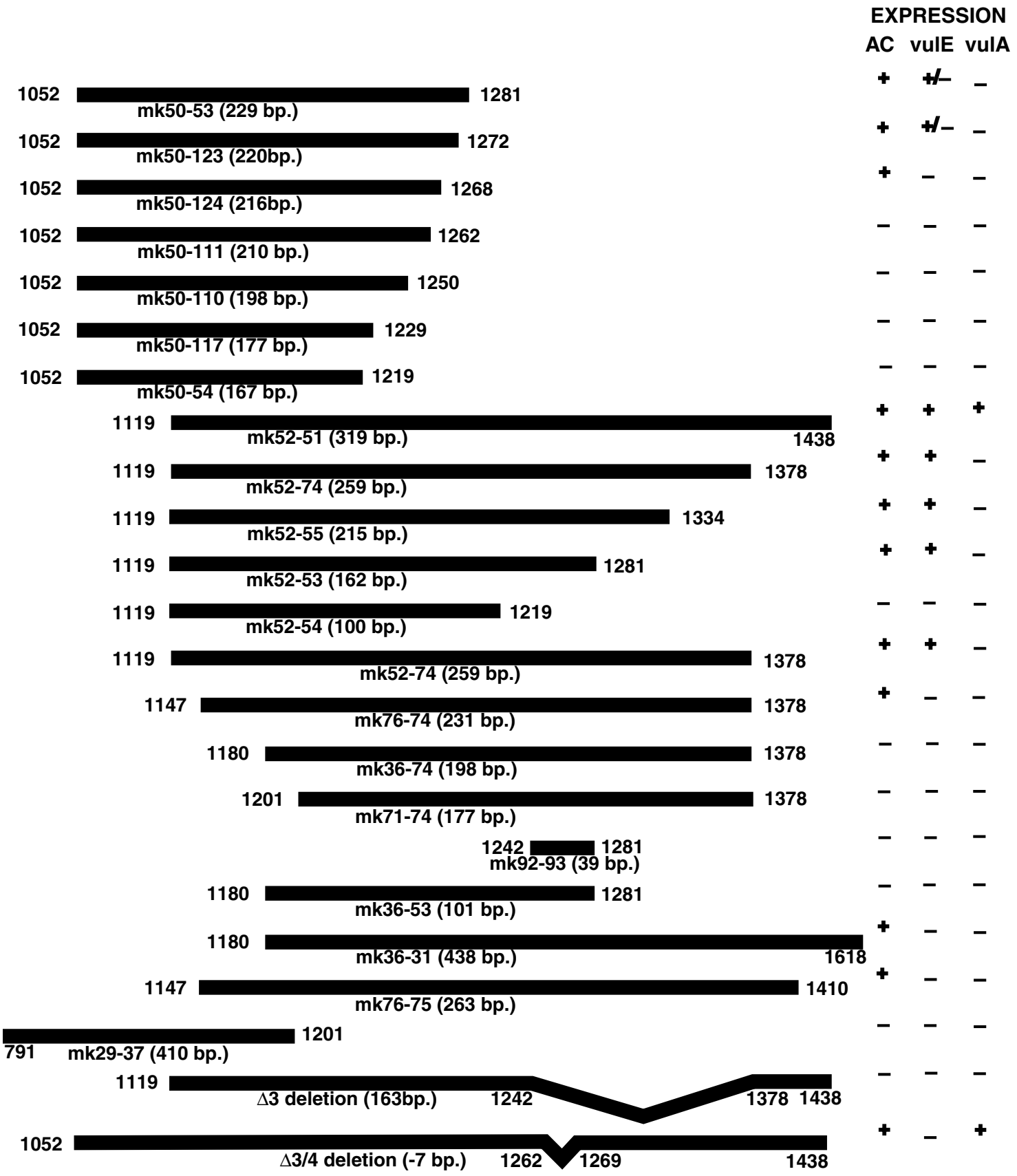


Supplemental Figure 2: *zmp-1 cis*-regulatory deletion analysis

The *zmp-1* upstream region is depicted at the top of the figure. A +/- indicates that either the expression level was reduced with respect to other constructs, or that not all animals showed consistent expression in the cell. AC stands for anchor cell.

Supplemental Figure 2: *zmp-1* cis-regulatory deletion analysis




EXPRESSION
AC vulE vulA
1052 1281
mk50-53 (229 bp.)
1052 1272
mk50-123 (220bp.)
1052 1268
mk50-124 (216bp.)
1052 1262
mk50-111 (210 bp.)
1052 1250
mk50-110 (198 bp.)
1052 1229
mk50-117 (177 bp.)
1052 1219
mk50-54 (167 bp.)
1119 1438
mk52-51 (319 bp.)
1119 1378
mk52-74 (259 bp.)
1119 1334
mk52-55 (215 bp.)
1119 1281
mk52-53 (162 bp.)
1119 1219
mk52-54 (100 bp.)
1119 1378
mk52-74 (259 bp.)
1147 1378
mk76-74 (231 bp.)
1180 1378
mk36-74 (198 bp.)
1201 1378
mk71-74 (177 bp.)
1242 1281
mk92-93 (39 bp.)
1180 1281
mk36-53 (101 bp.)
1180 1618
mk36-31 (438 bp.)
1147 1410
mk76-75 (263 bp.)
791 1201
mk29-37 (410 bp.)
1119 1242 1378 1438
Δ3 deletion (163bp.)
1052 1262 1269 1438
Δ3/4 deletion (-7 bp.)

Supplemental Figure 3: *cdh-3* cis-regulatory deletion analysis

The *cdh-3* upstream region is depicted at the top of the figure. A +/- indicates that either the expression level was reduced with respect to other constructs, or that not all animals showed consistent expression in the cell. Mk66-156 shows variable weak expression in the occasional animal in vulC and vulD, while mk158-159 shows variable weak expression in the occasional animal in vulC, E and F, but never in vulD.

.

Supplemental Figure 3: *cdh-3* cis-regulatory deletion analysis