© 2016. Published by The Company of Biologists Ltd | Development (2016) 143, 483-491 doi:10.1242/dev.130914

RESEARCH ARTICLE

The function and regulation of the GATA factor ELT-2 in the *C. elegans* endoderm

Tobias Wiesenfahrt[1], Janette Y. Berg[1], Erin Osborne Nishimura[2,3,*], Adam G. Robinson[3], Barbara Goszczynski[1], Jason D. Lieb[3] and James D. McGhee[1,‡]

ABSTRACT

ELT-2 is the major regulator of genes involved in differentiation, maintenance and function of *C. elegans* intestine from the early embryo to mature adult. *elt-2* responds to overexpression of the GATA transcription factors END-1 and END-3, which specify the intestine, as well as to overexpression of the two GATA factors that are normally involved in intestinal differentiation, ELT-7 and ELT-2 itself. Little is known about the molecular mechanisms underlying these interactions, how ELT-2 levels are maintained throughout development or how such systems respond to developmental perturbations. Here, we analyse *elt-2* gene regulation through transgenic reporter assays, ELT-2 ChIP and characterisation of *in vitro* DNA-protein interactions. Our results indicate that *elt-2* is controlled by three discrete regulatory regions conserved between *C. elegans* and *C. briggsae* that span >4 kb of 5′ flanking sequence. These regions are superficially interchangeable but have quantitatively different enhancer properties, and their combined activities indicate inter-region synergies. Their regulatory activity is mediated by a small number of conserved TGATAA sites that are largely interchangeable and interact with different endodermal GATA factors with only modest differences in affinity. The redundant molecular mechanism that forms the *elt-2* regulatory network is robust and flexible, as loss of *end-3* halves ELT-2 levels in the early embryo but levels fully recover by the time of hatching. When ELT-2 is expressed under the control of *end-1* regulatory elements, in addition to its own endogenous promoter, it can replace the complete set of endoderm-specific GATA factors: END-1, END-3, ELT-7 and (the probably non-functional) ELT-4. Thus, in addition to controlling gene expression during differentiation, ELT-2 is capable of specifying the entire *C. elegans* endoderm.

KEY WORDS: *Caenorhabditis elegans*, Endoderm development, Transcription, GATA factor, ELT-2, ChIP-Seq

INTRODUCTION

The *C. elegans* endoderm provides an experimentally accessible and relatively simple example of a transcriptional network that drives the development of an entire tissue, namely the intestine (reviewed by McGhee, 2013). The core transcription factors have been identified and their functional roles are understood at the level of genetics and cell biology. Current investigations focus on understanding the network at a biochemical level: (1) to define direct interactions between transcription factors and their target genes and (2) to understand how this regulatory network functions quantitatively (Maduro et al., 2015; Nair et al., 2013; Raj et al., 2010).

The entire *C. elegans* intestine (endoderm or E lineage) is produced from cells that descend from the single E blastomere of the eight-cell embryo (Fig. 1) (Sulston et al., 1983). Endoderm specification occurs when the genes encoding two small GATA-type transcription factors, END-1 and END-3, are transcriptionally activated only in the E blastomere (Maduro et al., 2005; Owraghi et al., 2009). In the current model of the endoderm regulatory network (Fig. 1), END-1 and END-3 activate transcription of the gene encoding the next factor in the endoderm regulatory cascade, the GATA factor ELT-7, at the 2E cell stage (Nair et al., 2013; Sommermann et al., 2010). The gene encoding the final factor in the cascade, the GATA factor ELT-2, is activated slightly later (at the 4E cell stage for most embryos) (Fukushige et al., 1998; Nair et al., 2013; Raj et al., 2010) and remains active into adulthood. It has been proposed that ELT-2 participates in the transcription of every gene expressed in the differentiating and mature intestine (with the likely exception of ribosomal protein genes), binding directly to TGATAA sites in intestinal gene promoters (McGhee et al., 2009, 2007). Loss of *elt-2* is completely lethal, whereas loss of *elt-7* has no obvious phenotype (Fukushige et al., 1998; McGhee et al., 2007; Sommermann et al., 2010), implying that END-1 and END-3 might interact directly with the *elt-2* promoter. In the absence of *elt-2*, the intestine is malformed but clearly specified and quite well differentiated (Fukushige et al., 1998). Loss of *elt-7* exacerbates the *elt-2* loss-of-function phenotype (Sommermann et al., 2010), but even the *elt-7; elt-2* double-mutant intestine is reasonably well formed, suggesting that END-1 and/or END-3 might also be able to activate early genes of intestinal differentiation. However, most of the direct interactions implied by this network are yet to be demonstrated.

Here, we address several questions. How is transcription of the *elt-2* gene controlled? Which of the other endodermal GATA factors participate directly? Are there a small number of crucial *cis*-acting sites in the *elt-2* promoter or are there large numbers of redundant sites, thereby providing possible insights into network behaviour? Do perturbations of the regulatory network persist or do they self-correct? Finally, what is the nature of the extensive redundancy within the endoderm network? Do individual factors have unique properties as proteins, or is it their expression timing that is important? As a partial answer to these last two questions, we show that ELT-2, if expressed under the control of the *end-1* promoter in addition to its own promoter, is able to replace the entire set of endoderm-specific GATA factors: END-1, END-3, ELT-7 and (the probably non-functional) ELT-4. Thus, ELT-2 alone can both specify the endoderm and regulate intestine differentiation and maintenance.

[1]Department of Biochemistry and Molecular Biology, Alberta Children's Hospital Research Institute, Cumming School of Medicine, University of Calgary, Calgary, Alberta, Canada T2N 4N1. [2]Lineberger Comprehensive Cancer Center, University of North Carolina, Chapel Hill, NC 27599, USA. [3]Department of Human Genetics, University of Chicago, Chicago, IL 60637, USA.
*Present address: Department of Biochemistry, Colorado State University, Fort Collins, CO 80523, USA.

‡Author for correspondence (jmcghee@ucalgary.ca)

Received 21 September 2015; Accepted 12 December 2015

Development (2016) 143, 483-491 doi:10.1242/dev.130914

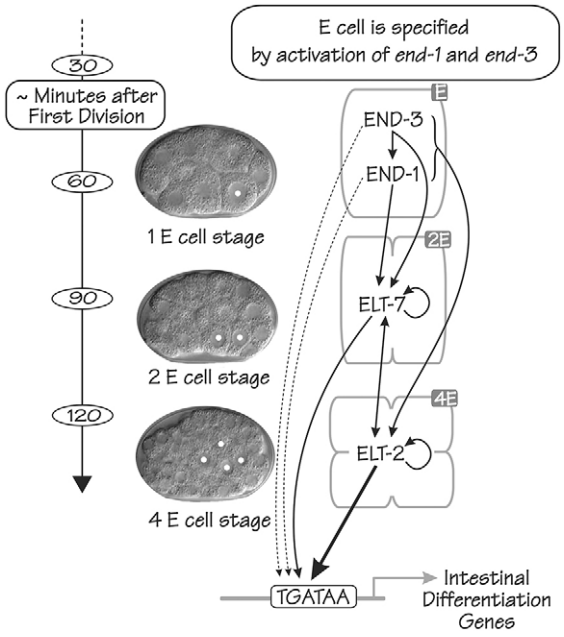


Fig. 1. Regulatory network consisting of the four zygotically expressed endoderm-specific GATA-type transcription factors that specify and differentiate the *C. elegans* early endoderm (E lineage). Time scale (minutes after first cell division at 20°C) is shown on the left. In the centre are images of three early stages of embryogenesis: the 1E, 2E and 4E cells are indicated by white dots. The current model for the roles and regulatory relations between the various transcription factors is shown on the right.

RESULTS

Defining the *elt-2* upstream regulatory region

Fig. 2A (top) shows anti-ELT-2 antibody staining in staged wild-type embryos: ELT-2 protein is never detected in 1E cell stages, is rarely (<1%) detected in 2E cell stages but is invariably detected by the 4E cell stage. Strong intestine-specific expression continues in later embryonic stages as well as in larvae and adults (8E and 1.5-fold stages are shown in Fig. 2A; adult staining is shown in Fig. S1A) (see also Fukushige et al., 1998). Fig. S1B provides evidence for the specificity of the antibody. As also shown in Fig. 2A (bottom), the native ELT-2 expression pattern is adequately reproduced by a transgenic reporter, in which 5048 bp of *elt-2* 5′ flanking region is used to drive expression of a nuclear-localised GFP reporter. GFP fluorescence is first detected weakly at the 4E cell stage and much more strongly at the 8E cell stage. This short time lag in reporter expression is consistent with the expected 30 min delay introduced by GFP folding and/or maturation (Iizuka et al., 2011) and, as expected, stronger 4E cell expression of the reporter can be detected by an anti-GFP antibody (data not shown).

To define *cis*-acting influences on *elt-2* transcription, genomic sequences were compared for the 6 kb upstream of the *elt-2* genes from *C. elegans* and the related nematode *C. briggsae*. A dot matrix comparison (Fig. 2B) detects three conserved regions (CRs): CR I (~ −0.6 to 0 kb), which contains the transcriptional start site at −499 bp (Kruesi et al., 2013); CR II (~ −2.2 to −1.5 kb); and CR III (~ −4.4 to −3.3 kb). Pairwise sequence alignments are shown in Fig. S2. Roughly similar regions of conservation can be detected when the *C. elegans elt-2* promoter is aligned with those from *C. brenneri* and *C. remanei*, but only CR III can be detected in the more distantly related nematode *C. japonicum* (data not shown).

Previous experiments and existing chromosomal deletions define limits to functional regions within the *elt-2* promoter. A 4.3 kb promoter fragment driving a C-terminal ELT-2::GFP fusion (with an *unc-54* 3′UTR) rescues the otherwise 100% lethal *elt-2* (*ca15*) null mutation [construct pJM86 in Fukushige et al. (1999)]. Deletion *ca16*, which removes *elt-4*, and deletion *gk153*, which removes the distal half of CR III and much of the open reading frame (ORF) C39B10.7 (Fig. 2B), have no measureable effect on brood size, defecation rate or growth rate [Fukushige et al. (2003) and Fig. S3, respectively]. We conclude that: (1) all necessary regulatory information lies within 5 kb upstream of the *elt-2* gene and (2) sequences removed by deletion *ca16* or *gk153* are not required for adequate *elt-2* expression.

The three conserved promoter regions contribute synergistically to the initiation and maintenance of *elt-2* expression

The results of two opposing deletion series (5′ and 3′) of an *elt-2prom::*GFP-*lacZ* reporter construct, assayed in transgenic embryos, are collected in Fig. S4. Reporter expression decreases abruptly as the proximal region of CR III is removed from either

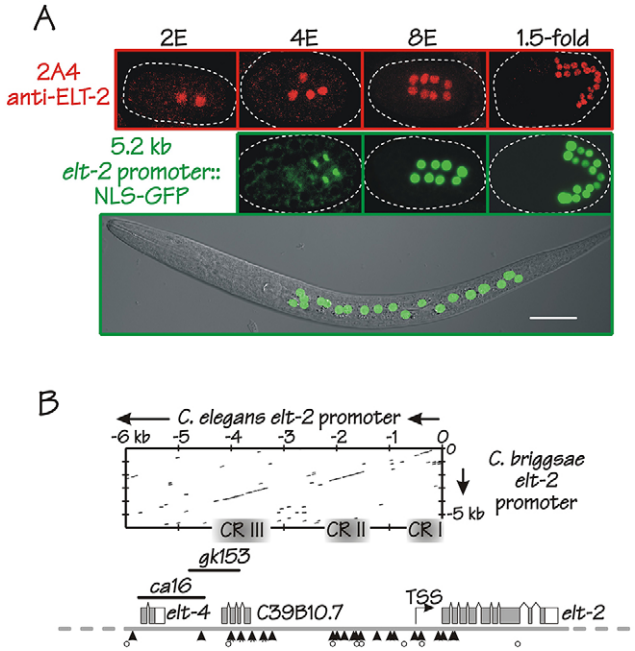


Fig. 2. Transcriptional regulation of the *elt-2* gene. (A) The expression of transgenic reporters accurately reflects the *in vivo* expression of ELT-2. The top row shows the normal endogenous expression patterns of ELT-2 in early *C. elegans* embryos, as detected by immunofluorescence using the anti-ELT-2 monoclonal antibody 455-2A4. The middle row shows expression patterns in early to mid-stage embryos of a transgenic nuclear-localised GFP reporter construct driven by the 5048 bp 5′ flanking region of the *elt-2* gene. Egg shells are outlined (dashed line). Beneath is a differential interference contrast (DIC) image of an L1 larva, with the fluorescence from the transgenic *elt-2*::GFP reporter superimposed. Scale bar: 20 µm. In each image, fluorescence signal is adjusted to high contrast to emphasize expression patterns and the lack of non-intestinal expression. (B) Identifying conserved regions in the *elt-2* 5′ control region by sequence alignments. The dot matrix plot (EMBOSS/dotmatcher, www.ebi.ac.uk/tools/emboss) compares 6 kb upstream of the ATG initiation codon for *C. elegans* (horizontal axis) and *C. briggsae* (vertical axis), revealing three blocks of conserved sequences (CR I, CR II and CR III). These conserved blocks are aligned with the genomic locus of the *C. elegans elt-2* gene, showing (to scale and from left to right) the upstream *elt-4* gene, the apparent ORF C39B10.7 and the *elt-2* coding region with the transcriptional start site (TSS) indicated. Also shown are two genomic deletions (*ca16* and *gk153*), TGATAA sites (filled triangles) and WGATAR sites that are not TGATAA (open circles).

Development (2016) 143, 483-491 doi:10.1242/dev.130914

direction. These results are consistent with two interpretations: (1) the proximal region of CR III contains a site crucial for *elt-2* expression in the embryo or (2) the *elt-2* promoter contains multiple distributed sites contributing to activity, the proximal region of CR III being the point where a critical number of these sites has been removed in either deletion series such that overall promoter activity now falls below a threshold.

To distinguish between these two models, CR I, CR II and CR III were fused individually and in combinations to a GFP reporter. Constructs were assessed for 'initiation' activity (expression at the 4E to 8E cell stage of transgenic embryos) and 'maintenance' activity (expression from the comma stage embryo through the larval and adult stages). When CR III, CR II and CR I are fused directly to each other and to the GFP reporter (Fig. 3A, Construct #1), expression is strong and robust, starting at the 4E cell stage (again, allowing for maturation time lag of GFP) and continuing to adulthood. Within experimental uncertainty, expression levels approximate to those produced by the full unmodified 5 kb promoter assayed with the same reporter, suggesting that all necessary or even influential *cis*-acting regulatory motifs are contained within the three conserved regions.

CR I in isolation (Fig. 3A, Construct #2) is neither able to initiate reporter expression at the 4E cell stage nor able to maintain expression past hatching but does significantly contribute to expression during the mid-to-late embryonic stages. CR II in isolation (Construct #3) is unable to drive detectable reporter expression at any stage. By contrast, CR III in isolation (Construct #4) is able to drive expression from the earliest initiation phase (4E) and at all subsequent stages into adulthood. CR III (which is perhaps augmented in its activity by basal promoter activity associated with the unannotated ORF C39B10.7; Fig. 2B) thus appears to provide the strongest contribution of the three individual conserved regions to *elt-2* transcriptional activity, both in initiation and in maintenance phases. However, when the conserved regions are combined pairwise (Fig. 3A, Constructs #5-7), they show clear synergy. That is, together the conserved regions are able to drive higher levels of reporter expression at more stages than the estimated sum of the activities of the two tested individual conserved regions. Part of the apparent synergy between CR III and CR I could be because CR I provides a basal promoter activity for CR III. When CR III is fused to a non-endodermal basal promoter (containing no TGATAA sites) from the *C. elegans* heat shock gene (Construct #8), reporter activity is indistinguishable from that of the CR III-CR I combination in Construct #7.

The synergies observed between the three conserved regions appear incompatible with the model in which transcription of *elt-2* is controlled by a crucial site situated in the proximal region of CR III but do appear compatible with the alternative interpretation of there being multiple *cis*-acting motifs distributed throughout the conserved regions that all contribute to promoter activity. Under this model, if the summed contributions of some subclass or combination of these *cis* acting sites lie above a threshold, *elt-2* is transcribed. The results also provide evidence against a model in which each conserved region contributes solely and uniquely to a spatial subpattern of activity (for example, to *elt-2* expression in the anterior or posterior intestine) or is solely responsible for *elt-2* expression during a restricted developmental time window (for example, only in embryos or only in L1 larvae).

Conserved TGATAA sites are individually dispensable but collectively crucial for *elt-2* control

Of the 30 potential GATA factor binding sites (defined as WGATAR) in the genomic region depicted in Fig. 2B, 22 are TGATAA, the site highly enriched in promoters of intestinal genes (McGhee et al., 2009, 2007; Pauli et al., 2006); this proportion (~73%) is more than twice that expected from base composition and applies to the overall region as well as to each of the three conserved regions. It would be an overwhelming task to mutate these individual sites combinatorially and comprehensively, especially without a precise quantitative assay. The present analysis is thus limited to mutating four conserved TGATAA sites in CR III and three conserved TGATAA sites plus a conserved AGATAG site in CR I.

As noted above, CR III fused to the basal heat shock promoter (Fig. 3A, Construct #8) produces strong expression from the 4E stage to adulthood. However, this activity is completely abolished by mutation of the four conserved TGATAA sites in CR III (Fig. 3B, Construct #9). Likewise, no activity is observed when a quadruply mutated CR III is combined with a quadruply mutated CR I (Fig. 3B, Construct #10). When an unmutated CR I or CR III is fused to a quadruply mutated CR III or CR I, respectively (Fig. 3B, Constructs #11,12), the resulting activity is close to that provided by the unmutated regions assayed in isolation. These results suggest that for CR III (and for CR I with the caveat that one of the mutated sites was AGATAG), TGATAA sites are necessary to provide enhancer activity. Further, the results suggest that there are no non-TGATAA sites that are sufficient for driving *elt-2* expression.

Each of the four TGATAA sites in CR III were mutated one at a time and then assayed as a fusion to an unmutated CR I (Fig. 3B, Constructs #13). Expression patterns produced by each of the four constructs are essentially indistinguishable and are similar to that produced by the unmutated CR III-CR I fusion (Construct #7). We conclude that there is no single TGATAA site within CR III that is necessary or obviously distinguished from the others. When two or even three of the four CR III TGATAA sites are mutated (Fig. 3B, Constructs #14, 15), expression remains strong.

END-1, ELT-7 and ELT-2 can bind *in vitro* to conserved TGATAA sites in the *elt-2* promoter

The above results show that CR III appears to be involved in both the initiation and maintenance of *elt-2* transcription. Electrophoretic mobility shift experiments show that END-1 (involved in *elt-2* initiation), ELT-2 (involved in *elt-2* maintenance) and ELT-7 (probably involved in both *elt-2* initiation and maintenance) can all bind *in vitro* to each of the four conserved TGATAA sites in CR III (Fig. 4A). Binding is specific in that it is competed efficiently by unlabelled double-stranded oligonucleotide but is competed much less efficiently by unlabelled oligonucleotides in which the TGATAA sites have been mutated to GTCGAC. Allowing for different specific activities of labelling of the four probes, for each of the three proteins we estimate that binding affinities to the individual TGATAA sites in CR III differ by at most 5- to 10-fold. These biochemical measurements support the general view derived from the previous transgenic experiments, namely that each TGATAA site contributes to overall promoter activity, even though individual sites may differ several fold in their influence.

ELT-2 binds directly *in vivo* to its own promoter and to promoters of intestinal differentiation genes

To test for direct ELT-2 occupancy at its own promoter *in vivo*, we performed chromatin immunoprecipitation and sequencing (ChIP-Seq) using an antibody specific to *C. elegans* ELT-2 (as used in Fig. 2A and Fig. S3) and extracts derived from wild-type L3 larvae. Within the *elt-2* promoter, three regions of ELT-2 occupancy can be

Development (2016) 143, 483-491 doi:10.1242/dev.130914

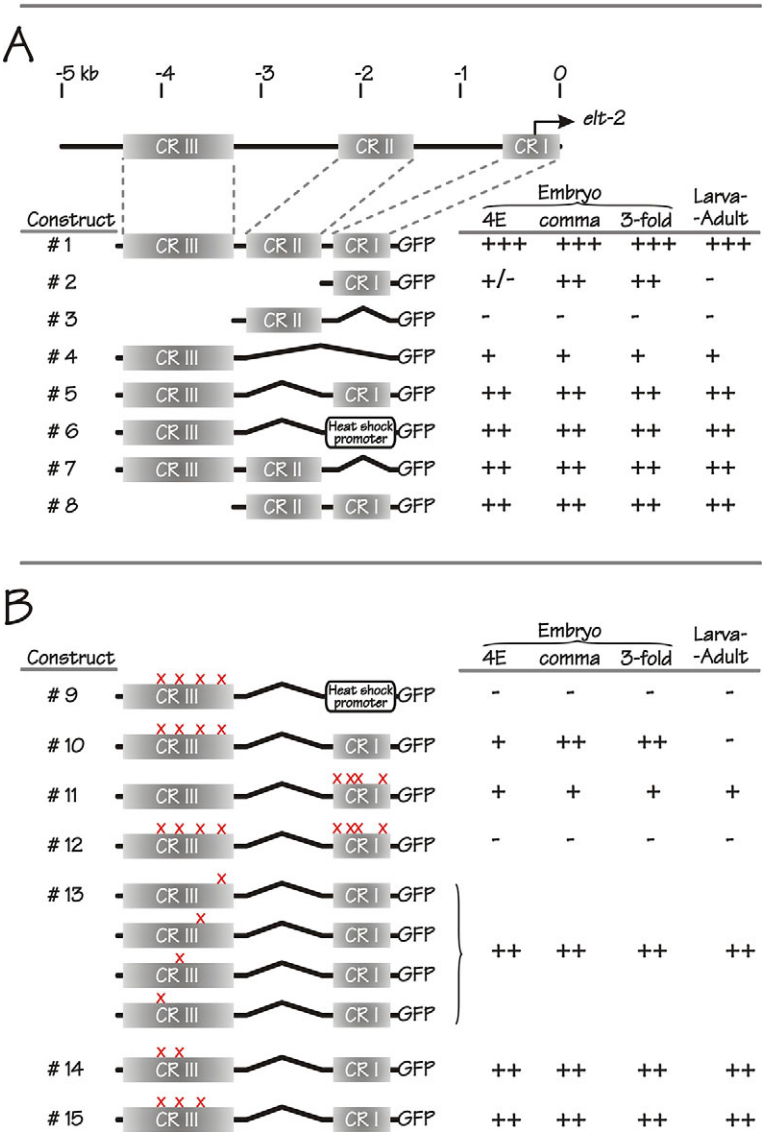


Fig. 3. Enhancer activities of the three conserved regions identified in the 5′ flanking region of *elt-2*. (A) The enhancer activity of CR I, CR II and CR III was tested individually and in combinations. Reporter expression patterns from multiple independent transgenic lines are summarised as: +++, similar pattern and intensity as the intact 5 kb promoter; ++ and +, decreasing (~by half) steps in intensity; –, no detectable reporter expression. (B) Importance of conserved TGATAA sites in CR III and CR I for *elt-2* enhancer activity. Mutated TGATAA sites are marked by 'X'.

identified that passed our threshold of significance (MACS2 peak scores $<10^{-30}$) and that aligned satisfactorily with CR I, CR II and CR III (Fig. 5, Fig. S5). Thus, the ChIP-Seq results support the view that ELT-2 interacts with all three of the conserved *cis*-regulatory regions in the *elt*-2 promoter. A peak of ELT-2 occupancy (below our significance threshold) can be detected upstream of the *elt-4* gene, suggesting that ELT-2 might also regulate *elt-4*. Fig. S6 shows that the expression of an integrated *elt-4* promoter::GFP transgenic reporter is largely abolished by *elt-2* RNAi.

An extensive ChIP-Seq analysis of ELT-2 binding to intestinal differentiation genes at different developmental stages will be provided elsewhere (E.O.N., J.D.L. and J.D.M., unpublished). However, the present ChIP-Seq data allow us to illustrate the direct binding of ELT-2 to two previously characterised intestinal differentiation genes, *ges-1* (gut esterase) and *cpr-6* (cysteine protease). As shown in Fig. 5B, a significant peak of ELT-2 occupancy can be detected 1.1 kb upstream of the *ges-1* initiation codon, aligning with the tandem pair of GATA sites (TGATAA and TGATAG) that have previously been shown to be functional in transgenic assays (Egan et al., 1995) and that were originally used to clone *elt-2* (Hawkins and McGhee, 1995). In L1 larvae, levels of *cpr-6* transcripts decrease ~100-fold in the absence of ELT-2 (McGhee et al., 2009) and a prominent peak of ELT-2 aligns with a TGATAA site immediately upstream of the *cpr-6* coding sequence (Fig. 5C).

ELT-2 levels recover from early perturbation caused by lack of END-3

Both quantitative *in situ* hybridisation and transgenic reporter assays show that loss of *end-3* significantly decreases the level of *elt-2* transcripts at the 4E to 8E cell stage (Boeck et al., 2011; Raj et al., 2010). Yet loss of *end-3* results in only 5-10% of embryos that lack an intestine (Maduro et al., 2007, 2005). Indeed, we determined that brood sizes in homozygous *end-3(-)* adults are essentially wild type [244±29 (n=1220) for *end-3(-)* adults, 253±33 (n=1265) for N2 (±s.d.); P=0.7, t-test]. Thus, either animals can survive and thrive with only a fraction of their normal ELT-2 levels, or the ELT-2 levels recover later in development.

ELT-2 protein levels in 8E cell stage *end-3(-)* embryos were measured by quantitative immunofluorescence and found to be 50±20% of those in 8E cell stage wild-type embryos (weighted mean±s.d. from three independent immunofluorescence comparisons, n=513 embryos), consistent with previous estimates (Raj et al., 2010; Boeck et al., 2011). Because immunofluorescence

Development (2016) 143, 483-491 doi:10.1242/dev.130914

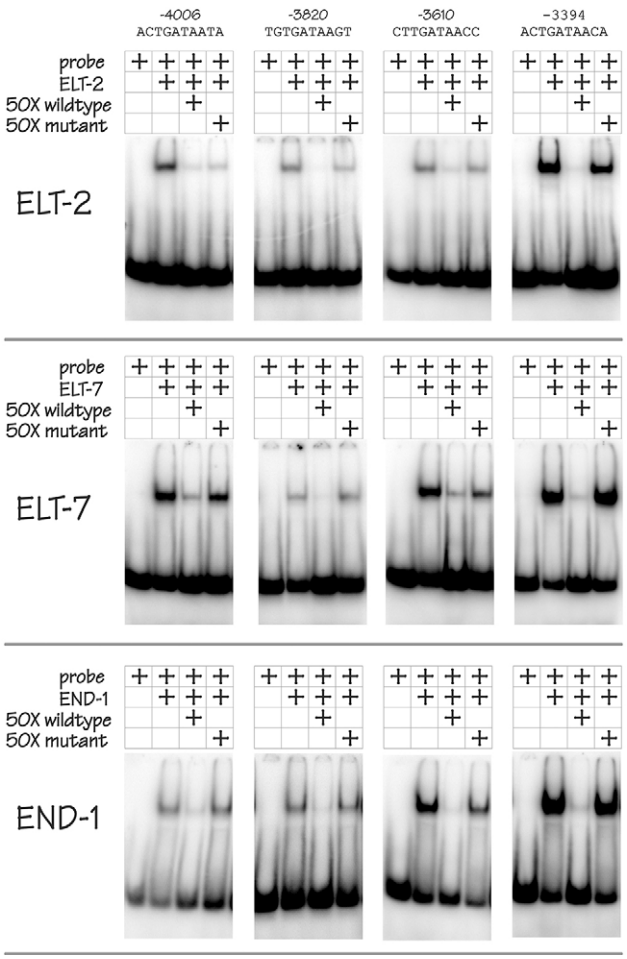


Fig. 4. DNA-protein interactions in the *elt-2* control region. Electrophoretic mobility shift assays to show that END-1, ELT-2 and ELT-7 proteins can all bind directly to each of the four conserved TGATAA sites in CR III of the *elt-2* promoter. The same set of labelled probes was used for all three proteins, with the coordinates of the individual conserved TGATAA sites shown at the top.

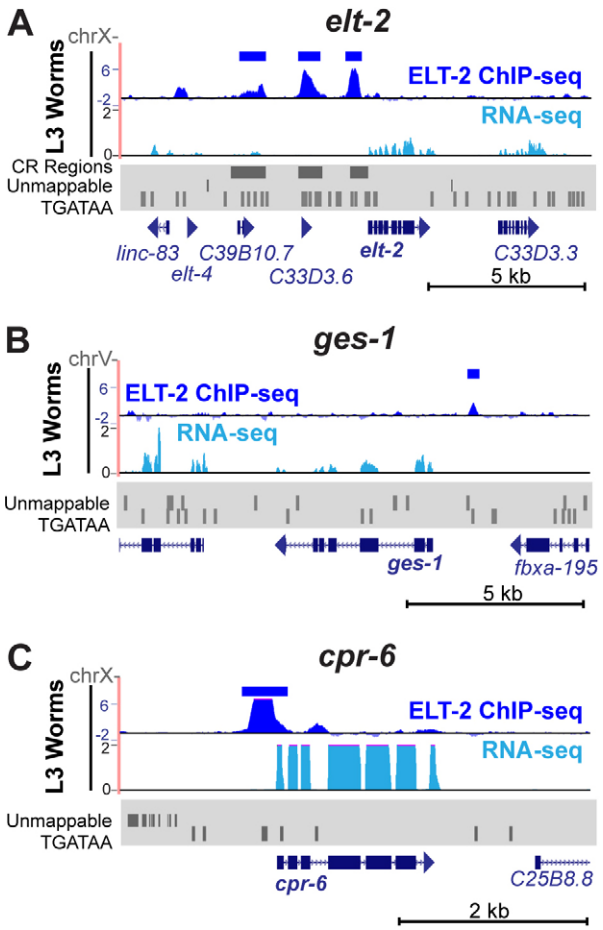


Fig. 5. ELT-2 ChIP-Seq on the *elt-2*, *ges-1* and *cpr-6* loci. (A) ELT-2 ChIP-Seq tracks (dark blue) from L3 larval worms are shown on and around the *elt-2* gene, with significant MACS2 peaks highlighted above (dark blue bars). ChIP-Seq reads were normalised with respect to read depth and IgG-only controls. The average of three replicates is shown. Individual replicates are shown in Fig. S5. The corresponding RNA-Seq (light blue) results obtained from the same chromatin preparation (whole L3 worms) are shown below. Regions of poor mappability owing to genomic repeats are depicted in the Repeatmasker (www.repeatmasker.org) trace (grey). The location of CR I, CR II and CR III and the occurrences of TGATAA motifs are also shown. (B,C) ELT-2 ChIP-Seq and RNA-Seq tracks at the (B) *ges-1* and (C) *cpr-6* loci.

of individual embryos during later stages of embryogenesis was too variable, we used quantitative western blotting to measure ELT-2 levels in newly hatched *end-3(-)* and wild-type L1 larvae (Fig. S7). We first established conditions in which the measured band intensity of a paramyosin UNC-15 control increased by 2.0±0.3-fold when twice the number of animals were loaded on the gel. From band intensities measured with three different loadings of both N2 and *end-3(ok1448)* L1 larvae (extracts of 500, 1000 and 2000 animals per lane) on each of two independent gels, we estimate that the ELT-2 levels in *end-3(-)* L1 larvae are 108±23% (±s.d.) of the UNC-15-normalised ELT-2 levels measured in N2 L1 larvae (P=0.4, t-test; thus not significantly different from 100%). Therefore, ELT-2 levels in *end-3(-)* animals have reattained wild-type levels by the time of hatching.

ELT-2 expressed earlier in development can replace all other endodermal GATA factors

In the course of normal *C. elegans* development, either END-1 or END-3 is necessary to specify the *C. elegans* endoderm (Maduro et al., 2015; Owraghi et al., 2009), whereas initial ELT-2 function is restricted to differentiation. We examined whether ELT-2 could replace END-1 and END-3 in specifying the endoderm, simply by being expressed earlier. If so, this would be a striking result, because these endodermal GATA factors have overall amino acid identities and similarities of only 12-24%.

We attempted to rescue an *end-1 end-3* double mutant by injecting a construct in which *elt-2* cDNA is under the control of either the *end-1* or *end-3* promoter. Strain MS1248 [*end-1(ok558) end-3(ok1448); irEx568 [end-1(+); end-3(+); sur-5::RFP]*; Owraghi et al., 2009] was injected so as to install a second extrachromosomal transgenic array containing, for example, *end-1prom::elt-2* cDNA, together with *elt-2prom::*GFP and *rol-6 (su1006)* marker constructs. Successful rescue was indicated by progeny worms that had lost the original rescuing array but had retained the replacement array, thereby producing green non-red rollers. Rescue was judged to be unsuccessful if such segregants could not be detected after several generations. Initial observations showed that both *end-3prom*::*elt-2* cDNA and *end-1prom::elt-2* cDNA constructs were able to rescue the *end-1 end-3* double mutant. Subsequent experiments were performed with the *end-1prom*::*elt-2* cDNA construct because it gave more efficient rescue. To establish a stable rescuing construct, one particular rescuing

Development (2016) 143, 483-491 doi:10.1242/dev.130914

array was integrated into a wild-type background, followed by outcrossing to remove extraneous mutations. This strain (JM230 *caIs85[end-1prom::elt-2* cDNA; *elt-2prom*::GFP; pRF4*]* I) serves as a control for phenotypes that are not connected to the rescuing ability of the transgenic array. The integrated array *caIs85* was then introduced into a quadruple homozygous null mutant in *end-1*, *end-3*, *elt-7* and *elt-4*.

The final rescued strain (JM229 *caIs85; elt-7 end-1 end-3; elt-4*) is surprisingly viable, healthy and fertile. JM229 shows only 3±4% embryonic lethality and 12±9% larval lethality (nine broods, 1467 total progeny), corresponding to an overall rescuing rate of ~85%. Control strain JM230 *caIs85* shows 2% and 0% embryonic and larval lethality, respectively, suggesting that the 15% overall lethality observed with JM229 is due to incomplete rescue and not to any unrelated property associated with the rescuing array. The rescued strain JM229 shows roughly the same level of lethality/arrest as seen with an *end-3* single mutant: we measured embryonic and larval lethality/arrest in strain RB1331 *end-3(ok1448)* as 3±3% and 12±7%, respectively (four broods, 912 total progeny); Maduro et al. (2007) previously reported that 5% of *end-3(ok1448)* embryos lack an intestine. Early embryonic phenotypes of JM229 (20°C) are mild and resemble the incomplete rescue of the *end-3* null phenotype described previously (Boeck et al., 2011). Specifically, the 2E cell stage is 2-4 min shorter than in control embryos, the 2E-to-4E cell division occurs closer to the ventral surface of the embryo [10±2 µm for JM229 (n=19); 15±1 µm for the N2 control (n=8); 15±1 µm for control strain JM230 (n=17)], and the division axis tends to be oriented in a more dorsoventral direction than normal. The most severe overall embryonic phenotype of the rescued strain is that the duration of embryogenesis (the time from the 2- to 4-cell stage to hatching) is extended by ~1 h compared with N2 and with the control strain JM230 (Fig. 6A).

An image of a newly hatched L1 larva of the rescued strain JM229 is shown in Fig. 6B. The number of intestinal nuclei is normal (20.1±0.8, n=31) as counted using the *elt-2prom::*GFP reporter that is part of the rescuing array (Fig. 6C). JM229 L1 larvae are 13-15% shorter than control L1 larvae [242±20 µm for JM229 (n=31); 284±25 µm for JM230 (n=32); 278±18 for N2 (n=30)]. The heads of rescued larvae sometimes appear more rounded than normal (compare the JM229 L1 shown in Fig. 6B with the wild-type L1 shown in Fig. 6D). The overall life cycle (time from 2- to 4-cell stage to first egg lay) is extended (~84 h for JM229 compared with ~60 h for the N2 control) but a similar delay is measured with the control strain JM230 and is thus more likely to reflect a property of the rescuing array than incomplete rescue. The most severe post-hatching phenotype is that the brood size is reduced [163±65 (n=9) for JM229; 262±52 (n=5) for N2; 221±28 (n=5) for JM230]. However, with respect to morphology and overall viability, the JM229 phenotypes are remarkably minor.

Using the same assay, we were unable to rescue the quadruple *elt-7 end-1 end-3; elt-4* mutant with a single copy of the *end-1prom*::*elt-2* cDNA construct inserted into the *ttTi5605* MosCI site on chromosome II (Frøkjaer-Jensen et al., 2012). Thus, *elt-2* overexpression in the early embryo might be a key feature of its ability to rescue the quadruple mutant. ELT-2 protein can always be detected immunologically in the 2E cell stage of JM229 embryos, which is one cell cycle earlier than it appears in wild-type embryos. However, the majority of JM229 1E cell embryos are ELT-2 negative, suggesting that there could be functional levels of ELT-2 below our detection limit or that endoderm specification can occur at the 2E cell stage. As assayed by immunofluorescence, a 2E cell stage JM229 embryo contains roughly the same amount of ELT-2 protein as found in an 8E to 16E cell stage wild-type embryo (data not shown). In other words, ELT-2 is indeed overexpressed in the earliest embryos of the rescued strain but not exceptionally so.

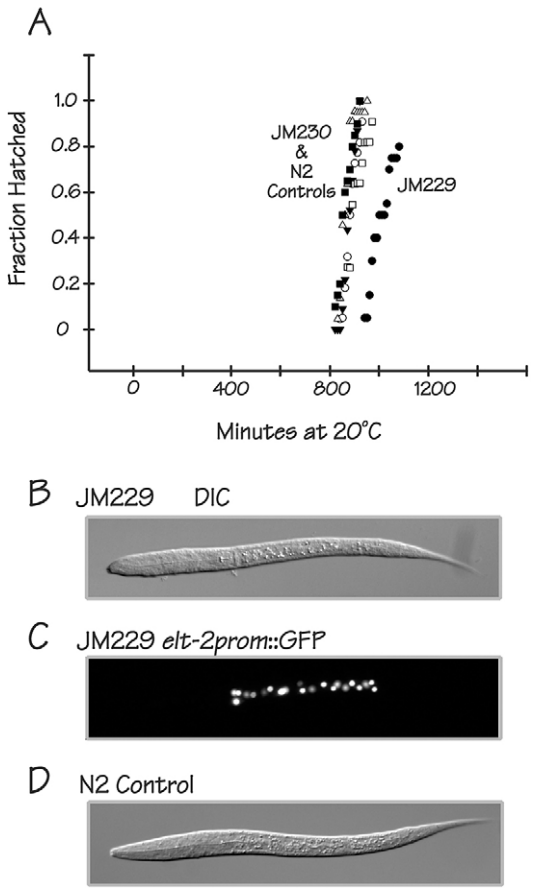


Fig. 6. ELT-2 can replace END-1, END-3, ELT-7 and ELT-4. (A) Embryos from the rescued quadruple mutant strain JM229 (black circles) hatch later than embryos from the control strain JM230 (white circles). Other symbols represent hatching curves measured for four different local versions of N2 wild-type worms (including a recent thaw). Time on the *x*-axis is minutes at 20°C from the 1- to 4-cell stage of embryogenesis. (B) DIC image of an L1 larva from the rescued quadruple mutant strain JM229. Average length of JM229 L1 larvae is 242±20 µm. (C) Fluorescent image of the same larva as in B, showing expression of an *elt-2prom*::GFP reporter incorporated into the integrated rescuing array *caIs85*. (D) DIC image of an L1 larva from N2 wild-type control. Average length of N2 L1 larvae is 278±18 µm.

elt-7 cDNA expressed under the control of the *end-1* promoter in a multicopy transgenic array was also able to rescue the quadruple *elt-7 end-1 end-3; elt-4* mutant. As expected, this rescue required *elt-2* (McGhee et al., 2007; Sommermann et al., 2010). However, not all GATA factors can specify the *C. elegans* endoderm. We used the *end-1* promoter to drive expression of cDNAs of either the *C. elegans* hypodermis-specific GATA factor ELT-3 (Gilleard et al., 1999) or the mouse endodermal GATA transcription factor GATA4 (Aronson et al., 2014), but neither construct was able to rescue the *end-1 end-3* double mutant.

DISCUSSION

The *C. elegans elt-2* gene is controlled by three conserved 'enhancers' distributed over ~5 kb of 5′ flanking region. Each of these three regions contributes to the transcriptional activation of *elt-2* but the exact contribution depends on which of the other

Development (2016) 143, 483-491 doi:10.1242/dev.130914

conserved regions are included in the construct. In other words, the three enhancers appear to interact with each other and to contribute synergistically to overall *elt-2* activity. At the present level of our analysis, there is no evidence of any particular subpattern of expression (e.g. adult stage only or anterior intestine only) being conveyed by any particular enhancer; rather, they all seem to contribute to overall *elt-2* transcriptional activity.

Focusing on two of the *elt-2* enhancers (CR I and CR III), we showed that mutation of all conserved TGATAA sites (four in CR III, three in CR I plus an AGATAG site) abolished both enhancer and basal promoter activity when assayed in transgenic reporters. We conclude that these conserved sites are necessary for *elt-2* expression and that no other site within the enhancers is sufficient for reporter expression (barring sites that overlap with the mutated TGATAA sites). Thus, we have no evidence that the core developmental control of *elt-2* expression is mediated by anything other than conserved TGATAA sites and, by implication, by the known set of endodermal GATA factors: END-1/3 and ELT-2/7. However, we fully expect that there will be other types of intestinal transcription factors and other *cis*-acting sites that, at least in post-embryonic stages, participate in *elt-2* control, maintaining physiological homeostasis and responding to nutritional or environmental signals.

The transcriptional activity of the most active enhancer, CR III, persisted even with only one remaining wild-type TGATAA site, at least qualitatively. Furthermore, we could find no evidence that any individual TGATAA site was functionally distinct from any other; for example, being solely responsible for the initiation of *elt-2* transcription or responsive to only one of the several GATA factors present in the early endoderm. This model is supported by the *in vitro* demonstration that END-1, ELT-7 and ELT-2 could all bind directly to each of the four conserved TGATAA sites within CR III. Even though different sites showed modestly different affinities for the different factors, we were not able to identify a 'GATA' site to which one factor could bind but another factor could not (note that END-3 protein was not available).

The above results lead to a model for *elt-2* control that is redundant, robust and flexible. Indeed, when ELT-2 levels are halved by loss of *end-3* in the early embryo, they are able to recover to wild-type levels by the time that the animals hatch, several hours later. Thus, the endoderm network is capable of dynamic re-equilibration or self-correction during embryogenesis. A different perturbation of the early endoderm regulatory network has recently been shown to lead to increased numbers of intestinal nuclei in the adult (Maduro et al., 2015). It will be important to determine if such adult phenotypes are due to a persistent perturbation of the transcriptional network and its downstream biochemical pathways or are rather due to some irreversible early cellular defect, such as aberrant cell division in the early embryo.

An unexpected result of the present study is the demonstration that ELT-2, when expressed under a transgenic *end-1* promoter as well as under its own endogenous promoter, is able to replace, essentially completely, the other four GATA factors involved in development of the *C. elegans* endoderm, namely END-1, END-3, ELT-7 and (for completeness) ELT-4. Rescue is highly efficient using a multiple copy transgene of *end-1prom*::*elt-2* cDNA, but is not perfect. The strain has low levels of embryonic/larval lethality, slightly perturbed gastrulation, marginally slower embryonic development and a lower brood size. However, overall, we regard these phenotypes as remarkably modest considering the extensive rearrangement of the core transcriptional network. We were unable to achieve rescue of the quadruple *elt-7 end-1 end-3; elt-4* mutant using a single integrated copy of the *end-1prom::elt-2* cDNA transgene. This failure could possibly reflect a position effect, although single-copy insertions of *end-1* into this same genomic locus appear to function well (Maduro et al., 2015). Alternatively, perhaps the larger ELT-2 molecule takes longer to be produced than the smaller END-1 and/or END-3 molecules (see below), or perhaps ELT-2 binds less optimally to the early endoderm-specifying genes that are the normal targets of END-1/END-3. Both of these inefficiencies might require that ELT-2 is expressed at higher levels in order to compensate.

Regulatory pathways are thought to evolve in a retrograde manner, with genes expressed late brought under control of factors expressed earlier, which in turn are brought under the control of factors expressed even earlier (Wilkins, 1995). Thus, genes encoding intestinal digestive enzymes or intestinal structural proteins might originally have been under the sole control of ELT-2. ELT-2 might subsequently have come under control of the earlier expressed *end-1* and *end-3*, with the redundant activation of *elt-2* by *elt-7* introduced as an intermediate step or as a later intercalation.

In light of the demonstration that ELT-2 can perform all necessary functions of END-1, END-3 and ELT-7, why did the *C. elegans* endoderm pathway evolve as it did and not remain with only ELT-2 as the transcriptional activator of the genes performing specification, differentiation, growth and intestinal maintenance? Three reasons come to mind. The first reason is possible selection for greater fidelity of endoderm development controlled by a redundant pathway (Cooke et al., 1997). A second possible reason is to separate *elt-2* control from the influence of *skn-1* and *pop-1*, the two maternal-effect genes that activate *end-1* and *end-3* in the E blastomere. Both *skn-1* and *pop-1* also function zygotically within the differentiating and mature intestine and the transient expression of *end-1* and *end-3* would free ELT-2 from being controlled by *skn-1* and/or *pop-1* throughout the animal's lifespan. A third reason could be that the use of the smaller *end-1* and *end-3* genes allows more rapid transcription and translation and hence more rapid specification of the E cell than if specification depended on *elt-2*. The differences are greatest when comparing *end-3* and *elt-2*. The transcript lengths are 1276 and 2344 nucleotides, respectively; protein sizes are 242 and 433 amino acids, respectively. Cell cycle times in the early *C. elegans* embryo are only 15-20 min, perhaps providing sufficient time for *end-3* but not *elt-2* to be transcribed and translated. In the early *Drosophila* embryo, cell cycle times are even shorter and it has been shown that the transcription of long genes is aborted by the intervention of mitosis (Shermoen and O'Farrell, 1991). However, *elt-2* (~2.3 kb) is much shorter than the *Drosophila* gene (*Ubx*, 57 kb), the transcription of which is aborted by mitosis, and to our knowledge it has not yet been shown that similar mitosis-aborted transcription occurs in early *C. elegans* development. Moreover, our rescuing transgene is based on *elt-2* cDNA, which is roughly the same size as the *end-3* transcript, suggesting that perhaps it is translation that is limiting, not transcription. In any event and independent of any particular explanation, the quadruple mutant *elt-7(-) end-1(-) end-3(-); elt-4(-)* embryos rescued by the *end-1prom::elt-2* cDNA transgene hatch ~1 h later than wild-type embryos, a potentially huge fitness disadvantage.

The *C. elegans* endoderm is one of only several developmental cell lineages in which a plausible direct molecular chain of command can be proposed to connect factors in the maternal cytoplasm with factors controlling tissue-specific gene transcription in the mature adult. In the present paper, we have defined core

Development (2016) 143, 483-491 doi:10.1242/dev.130914

features of both the *cis*-acting sequences and the *trans*-acting factors controlling transcription of the gene encoding ELT-2, the predominant transcription factor associated with endoderm differentiation and function. We have provided a clear example of how the regulatory network is able to overcome a severe early perturbation, namely the low concentration of ELT-2 in the early embryo caused by loss of the *end-3* gene. Finally, we have explored the regulatory potential of the ELT-2 protein and have shown that it is capable of replacing all other endodermal GATA factors, in particular replacing the END-1 and END-3 factors that normally specify the endoderm. Our results contribute to a long-term goal of describing development of the *C. elegans* endoderm quantitatively, in terms of binding affinities for particular regulatory sites and in terms of transcription factor activities, redundancies and stabilities.

MATERIALS AND METHODS

Nematode strains

C. elegans strains were grown on OP50-seeded NGM plates (Brenner, 1974). Transgenic animals were produced by standard gonadal injection (Mello et al., 1991), with the DNA construct to be tested present at 25-50 µg/ml, usually together with pRF4 at 50 µg/ml as a phenotypic marker. Plasmids were constructed by standard methods and are described in more detail in the supplementary Materials and Methods. Selected transforming arrays were integrated into the genome using γ-irradiation as previously described (Egan et al., 1995), followed by outcrossing at least four times. The full genotype of the rescued quadruple mutant strain JM229 is: *caIs85[end-1prom::elt-2*cDNA (pJM513), *rol-6(su1006)* (pRF4), *elt-2prom*::GFP (pJM370) *I]*; *elt-7(tm840) end-1(ok558) end-3(ok1448) V; elt-4(ca16) X*. The genotype of the control strain JM230 is: *caIs85 I*. All mutant alleles of the GATA factor genes are deletions (presumed nulls) and were followed in genetic crosses by PCR; primer sequences and expected product sizes are detailed in Table S1. The homozygosity of the *end-1* deletion in the final strain JM229 was also verified by Southern blotting. The *end-1prom::elt-2* cDNA sequence was cloned into plasmid pCFJ350 (Addgene) and a single copy inserted into chromosome II using strain EG6699 and the MosCI technique developed by Frøkjaer-Jensen et al. (2012).

Antibodies and immunodetection

The anti-ELT-2 monoclonal antibody 455-2A4 (isotype IgG1) used in immunofluorescence, ChIP and western blotting was produced by the Southern Alberta Cancer Research Institute Antibody Services, using as antigen a purified polyhistidine-tagged full-length ELT-2 protein produced in *E. coli*. Immunostaining of embryos dissected from adult hermaphrodites was performed as described (Van Furden et al., 2004). Western blots were performed by standard methods, probed with anti-ELT-2 monoclonal antibody 455-2A4 together with monoclonal antibody MH16 (Developmental Studies Hybridoma Bank, University of Iowa) to detect paramyosin as a loading control and developed with HRP-conjugated secondary antibodies and the Amersham ECL Prime reagent. Band intensities were measured with a LAS4000 Imaging Station (GE Healthcare) and quantitated using Fiji. Further details of the antibodies used and immunodetection methods are provided in the supplementary Materials and Methods.

DNA-protein interactions

Electrophoretic mobility shift assays (band shifts) were performed as previously described (Kalb et al., 1998). The four TGATAA sites (with 10 bp flanking sequence on either side) were all synthesized as self-complementary hairpins (four C residues in the loop) to guarantee equal stoichiometry and a double-stranded conformation. Only the wild-type hairpin was used as a labelled probe, in which case a missing 3′ terminal C residue was filled in using Klenow polymerase and α^{32}P-dCTP. Binding specificities were tested using competition with non-labelled hairpins at a 50-fold excess. Full-length His-tagged END-1, ELT-7 and ELT-2 proteins were synthesized in baculovirus-infected insect cells, either by ourselves or by ARVYS Proteins Inc., and purified by metal-affinity chromatography (estimated purity >90%).

Chromatin immunoprecipitation followed by deep sequencing (ChIP-Seq)

ChIP-Seq was performed and the data were analysed as previously described (Berkseth et al., 2013). ChIP-Seq and RNA-Seq methods, analysis parameters and instructions for data access (GEO accession numbers) are described in detail in the supplementary Materials and Methods, with ELT-2 ChIP-Seq peaks and summits listed in Tables S2 and S3.

Acknowledgements

We thank the following people who have variously contributed to this project over the years: Jamie Feng, Anne Formaz-Preston, Tetsunari Fukushige, Mark Hawkins, Sai Ravikumar, Fran Snider and Lana Wong.

Competing interests

The authors declare no competing or financial interests.

Author contributions

T.W. and J.Y.B. performed and analyzed the immunohistochemical and transgenic reporter experiments. B.G. performed the EMSA assays for *in vitro* DNA-protein interactions. E.O.N., A.G.R. and J.D.L. performed and analyzed the ELT-2 ChIP-Seq and RNA-Seq experiments. T.W. performed the mutant rescue experiment. T.W., E.O.N., J.D.L. and J.D.M. wrote and edited the manuscript.

Funding

Work in Calgary was supported by an operating grant from the Canadian Institutes of Health Research to J.D.M.; the ChIP-Seq/RNA-Seq experiments were funded by the National Institutes of Health [grant 5R01GM104050 to J.D.L.]. E.O.N. was supported by a Damon Runyon Cancer Research Foundation Postdoctoral Fellowship Award [2083-11]. J.D.M. gratefully acknowledges salary support from the Alberta Heritage Foundation for Medical Research (now AIHS) and from the Canada Research Chairs Program. Some strains were provided by the CGC, which is funded by the NIH Office of Research Infrastructure Programs [P40 OD010440]. Deposited in PMC for release after 12 months.

Supplementary information

Supplementary information available online at
http://dev.biologists.org/lookup/suppl/doi:10.1242/dev.130914/-/DC1

References

Aronson, B. E., Stapleton, K. A. and Krasinski, S. D. (2014). Role of GATA factors in development, differentiation, and homeostasis of the small intestinal epithelium. *Am. J. Physiol. Gastrointest. Liver Physiol.* **306**, G474-G490.

Berkseth, M., Ikegami, K., Arur, S., Lieb, J. D. and Zarkower, D. (2013). TRA-1 ChIP-seq reveals regulators of sexual differentiation and multilevel feedback in nematode sex determination. *Proc. Natl. Acad. Sci. USA* **110**, 16033-16038.

Boeck, M. E., Boyle, T., Bao, Z., Murray, J., Mericle, B. and Waterston, R. (2011). Specific roles for the GATA transcription factors end-1 and end-3 during C. elegans E-lineage development. *Dev. Biol.* **358**, 345-355.

Brenner, S. (1974). The genetics of Caenorhabditis elegans. *Genetics* **77**, 71-94.

Cooke, J., Nowak, M. A., Boerlijst, M. and Maynard-Smith, J. (1997). Evolutionary origins and maintenance of redundant gene expression during metazoan development. *Trends Genet.* **13**, 360-364.

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

Frøkjaer-Jensen, C., Davis, M. W., Ailion, M. and Jorgensen, E. M. (2012). Improved Mos1-mediated transgenesis in C. elegans. *Nat. Methods* **9**, 117-118.

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

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

Fukushige, T., Goszczynski, B., Tian, H. and McGhee, J. D. (2003). The evolutionary duplication and probable demise of an endodermal GATA factor in Caenorhabditis elegans. *Genetics* **165**, 575-588.

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

Hawkins, M. G. and McGhee, J. D. (1995). elt-2, a second GATA factor from the nematode Caenorhabditis elegans. *J. Biol. Chem.* **270**, 14666-14671.
Iizuka, R., Yamagishi-Shirasaki, M. and Funatsu, T. (2011). Kinetic study of de novo chromophore maturation of fluorescent proteins. *Anal. Biochem.* **414**, 173-178.
Kalb, J. M., Lau, K. K., Goszczynski, B., Fukushige, T., Moons, D., Okkema, P. G. and McGhee, J. D. (1998). pha-4 is Ce-fkh-1, a fork head/HNF-3alpha, beta, gamma homolog that functions in organogenesis of the C. elegans pharynx. *Development* **125**, 2171-2180.
Kruesi, W. S., Core, L. J., Waters, C. T., Lis, J. T. and Meyer, B. J. (2013). Condensin controls recruitment of RNA polymerase II to achieve nematode X-chromosome dosage compensation. *eLife* **2**, e00808.
Maduro, M. F., Hill, R. J., Heid, P. J., Newman-Smith, E. D., Zhu, J., Priess, J. R. and Rothman, J. H. (2005). Genetic redundancy in endoderm specification within the genus Caenorhabditis. *Dev. Biol.* **284**, 509-522.
Maduro, M. F., Broitman-Maduro, G., Mengarelli, I. and Rothman, J. H. (2007). Maternal deployment of the embryonic SKN-1→MED-1,2 cell specification pathway in C. elegans. *Dev. Biol.* **301**, 590-601.
Maduro, M. F., Broitman-Maduro, G., Choi, H., Carranza, F., Wu, A. C. and Rifkin, S. A. (2015). MED GATA factors promote robust development of the C. elegans endoderm. *Dev. Biol.* **404**, 66-79.
McGhee, J. D. (2013). The Caenorhabditis elegans intestine. *Wiley Interdiscip. Rev. Dev. Biol.* **2**, 347-367.
McGhee, J. D., Sleumer, M. C., Bilenky, M., Wong, K., McKay, S. J., Goszczynski, B., Tian, H., Krich, N. D., Khattra, J., Holt, R. A. et al. (2007). The ELT-2 GATA-factor and the global regulation of transcription in the C. elegans intestine. *Dev. Biol.* **302**, 627-645.
McGhee, J. D., Fukushige, T., Krause, M. W., Minnema, S. E., Goszczynski, B., Gaudet, J., Kohara, Y., Bossinger, O., Zhao, Y., Khattra, J. et al. (2009). ELT-2 is the predominant transcription factor controlling differentiation and function of the C. elegans intestine, from embryo to adult. *Dev. Biol.* **327**, 551-565.
Mello, C. C., Kramer, J. M., Stinchcomb, D. and Ambros, V. (1991). Efficient gene transfer in C.elegans: extrachromosomal maintenance and integration of transforming sequences. *EMBO J.* **10**, 3959-3970.
Nair, G., Walton, T., Murray, J. I. and Raj, A. (2013). Gene transcription is coordinated with, but not dependent on, cell divisions during C. elegans embryonic fate specification. *Development* **140**, 3385-3394.
Owraghi, M., Broitman-Maduro, G., Luu, T., Roberson, H. and Maduro, M. F. (2009). Roles of the Wnt effector POP-1/TCF in the C. elegans endomesoderm specification gene network. *Dev. Biol.* **340**, 209-221.
Pauli, F., Liu, Y., Kim, Y. A., Chen, P. J. and Kim, S. K. (2006). Chromosomal clustering and GATA transcriptional regulation of intestine-expressed genes in C. elegans. *Development.* **133**, 287-295.
Raj, A., Rifkin, S. A., Andersen, E. and van Oudenaarden, A. (2010). Variability in gene expression underlies incomplete penetrance. *Nature* **463**, 913-918.
Shermoen, A. W. and O'Farrell, P. H. (1991). Progression of the cell cycle through mitosis leads to abortion of nascent transcripts. *Cell* **67**, 303-310.
Sommermann, E. M., Strohmaier, K. R., Maduro, M. F. and Rothman, J. H. (2010). Endoderm development in Caenorhabditis elegans: the synergistic action of ELT-2 and -7 mediates the specification→differentiation transition. *Dev. Biol.* **347**, 154-166.
Sulston, J. E., Schierenberg, E., White, J. G. and Thomson, J. N. (1983). The embryonic cell lineage of the nematode Caenorhabditis elegans. *Dev. Biol.* **100**, 64-119.
Van Furden, D., Johnson, K., Segbert, C. and Bossinger, O. (2004). The C. elegans ezrin-radixin-moesin protein ERM-1 is necessary for apical junction remodelling and tubulogenesis in the intestine. *Dev. Biol.* **272**, 262-276.
Wilkins, A. S. (1995). Moving up the hierarchy: a hypothesis on the evolution of a genetic sex determination pathway. *Bioessays* **17**, 71-77.

SUPPLEMENTARY MATERIALS AND METHODS

Reporter constructs

The series of promoter deletions shown in Supplementary Figure S4 was based on the GFP/*lacZ* reporter vector pPD96.04 (Fire et al., 1990), fused to the *elt-2* 5'-region at a BamHI site 32 amino acids into the *elt-2* coding region. Expression was detected in transgenic embryos by β-galactosidase activity. As a secondary point of interest, we occasionally found that it was difficult to produce transgenic strains with particular constructs (more often if the promoter had been extensively deleted); moreover, we occasionally detected transformed worms showing a Gob (gut-obstructed) phenotype similar to that of *elt-2(-)* arrested larvae, suggesting that some sort of sporadic co-suppression effect could be present. Subsequent reporter constructs (Fig. 2A and 3) incorporated the following changes. The 5'-flanking region of *elt-2* was modified to insert a BamHI site immediately downstream of the *elt-2* initiation codon. The basic GFP reporter vector (pJM355) was a derivative of the GFP-histone H2B construct pAP.10 (provided by Dr. Jeb Gaudet and described briefly in (Gaudet and Mango, 2002)), incorporating the following modifications: (i) removal of the *pes-10* basal promoter sequence; (ii) removal of duplicate multiple cloning sites; (iii) alteration of the reading frame to match that of pPD96.04; (iv) removal of the histone H2B 3'-UTR together with an unwanted BamHI site, and; (v) truncation of the standard *unc-54* 3'-UTR to remove sequences from *aex-1*. Any differences between the results shown in Figure 3 and in Supplementary Figure 4 can reasonably be ascribed to the GFP vector being more sensitive and less variable than the earlier GFP/*lacZ* reporter. The various TGATAA motifs were mutated to GTCGCC using overlapping PCR strategies with mutated primers; for some later constructs, the local region containing the mutation was synthesized commercially. All constructs were verified by sequencing. Plasmid sequences can be obtained from the authors.

Antibodies and immunodetection

Hybridoma supernatants for the anti-ELT-2 monoclonal antibody 455-2A4 (isotype IgG1) were concentrated ~10-fold by pressure filtration, dialyzed extensively against PBS and then concentrated a further 2-3 fold by dialysis against PBS-50 % glycerol. Sodium azide was added to 0.01% before storage at -80ºC. Immunostaining of embryos used the concentrated 455-2A4 stock at a dilution of 1/2000. For the Western blots performed to measure ELT-2 protein levels, 500-2000 L1 larvae (hatched in the absence of food) were boiled in sample buffer and the extracts electrophoresed on 8% polyacrylamide-SDS gels; gel contents were electrophoretically transferred to PVDF membranes and probed with anti-ELT-2 monoclonal antibody 455-2A4 (1/1000 dilution of concentrated stock) together with monoclonal antibody MH16 to detect paramyosin as a loading control. Both primary antibodies were detected using a

donkey anti-mouse HRP-conjugated secondary antibody diluted 1:5000, followed by Amersham ECL Prime Western reagent (GE Healthcare HRP2232).

Chromatin immunoprecipitation followed by deep sequencing (Chip-Seq)

N2 worms were grown in liquid culture (S-medium) at 25°C, fed with HB101 *Escherichia coli* bacteria, and harvested for ChIP-seq at the L3 stage of development (24 – 26 hours post feeding). Flash frozen larvae were homogenized in a mixer mill, fixed in 1 % formaldehyde for 10 minutes, quenched with 50 mM glycine for 5 min, sheared by sonication in a Bioruptor, and the resulting chromatin was extracted. Immunoprecipitation was performed using either 5 µg α-ELT-2 (455-2A4) antibody, 5 µg α-H3K4me^3 antibody (WAKO 305-34819), or mock conditions (IgG only). ChIP-recovered DNA and input DNA were ligated to sequencing adapters containing an eight-base multiplexing barcode sequence. DNA libraries were sequenced on either the Hi-Seq2000 or Hi-Seq2500 Illumina sequencers with 50 rounds of single-end sequencing, 8 cycles of which were used for multiplex barcode indices. Data from this study are available on NCBI GEO (GSE71720).

ChIP-seq Analysis

Quality filtering: Sequenced reads were trimmed of barcodes (FASTX-Toolkit, 11sep2008, http://hannonlab.cshl.edu/fastx_toolkit/index.html) and filtered for primer and adapter sequences using Tagdust (1.12), (Lassmann et al., 2009). **Alignment:** Resulting high quality reads were aligned to the *C. elegans* ce10 genome using Bowtie/1.1.0 (Langmead et al., 2009). Resulting bam files were normalized to read depth and converted to bigwig format using bedtools genomecov (2.22.1) (Quinlan et al., 2010), followed by bedGraphToBigWig (UCSC Genome Browser) to create individual bigwig tracks. The resulting ELT-2 and H3K4me3 ChIP-seq bigwig files were merged by subtracting IgG-only bigwig tracks and averaging the result across replicates using java-genomics-toolkit wigmath.Average (java/1.8.0_11), (http://palpant.us/java-genomics-toolkit/) to create average_track.bw files. **Peak Calling:** Significant peaks were identified by MACS2 (2.0.9) using a concatenated set of ChIP-seq replicates as treatment and a concatenated set of input samples as a control (Zhang et al., 2008). Peaks were filtered for a minimum threshold of $-\log_{10}(\text{q-value}) \geq 30$. HOT regions (Chen et al., 2014), blacklisted regions (Araya et al., 2014) and IgG-only peaks were subtracted from the set of identified MACS2 peaks.

RNA-seq

Total RNA was prepared using TRIZOL and Qiagen RNeasy spin kit. mRNA was isolated from total RNA preparations and RNA-seq was performed using a standard protocol (NEBNext Poly(A) mRNA Magnetic Isolation Module E7490, NEBNext Ultra RNA Library Prep Kit E7530 or E7420, and Agencourt Ampure XP beads). The resulting fragmented cDNA was sequenced on Hi-Seq2500 Illumina instruments as 50 bp single-end reads. Data from this study are available on NCBI GEO (GSE71720).

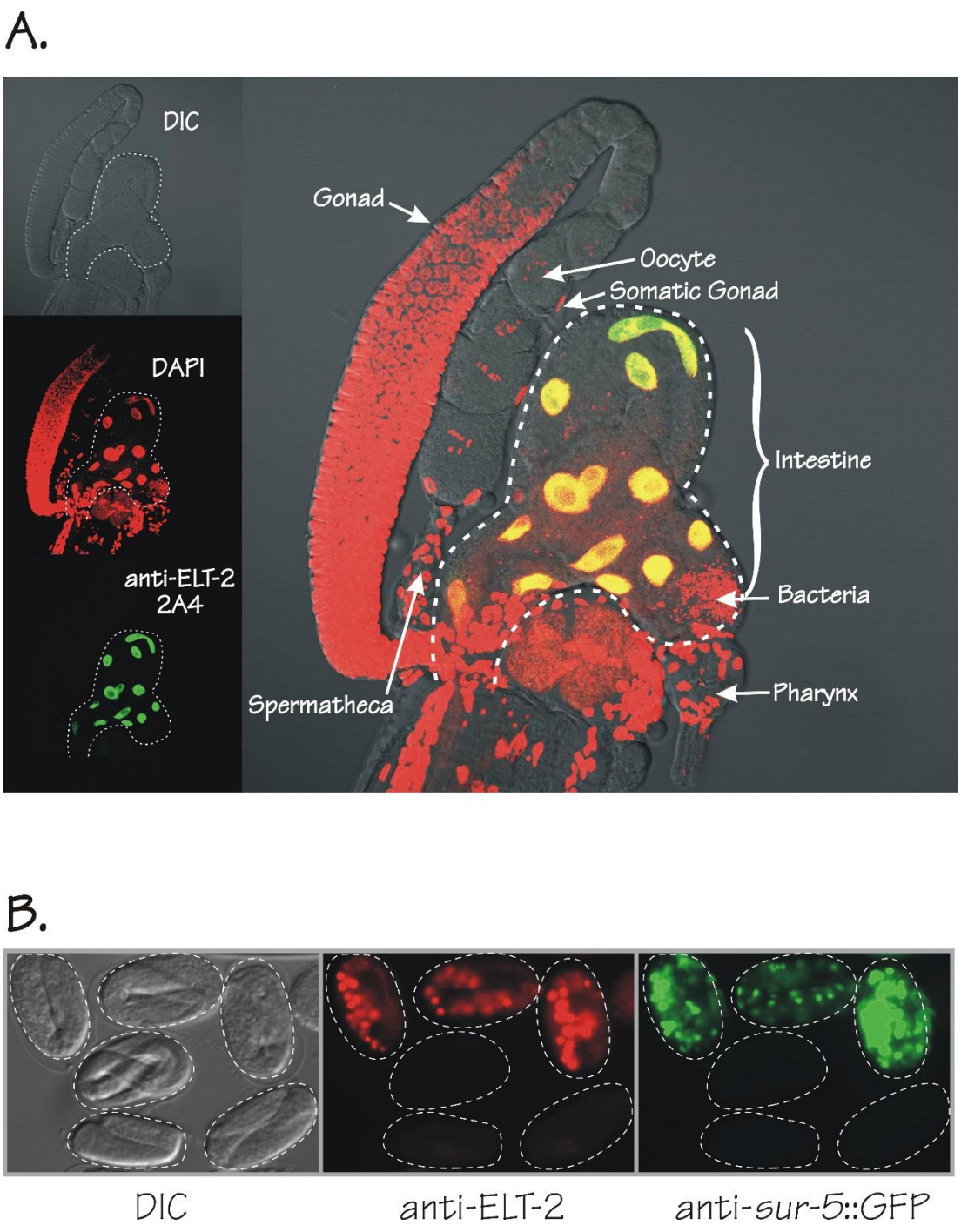
RNA-seq Analysis

Quality filtering: Sequenced reads were filtered for primer and adapter sequences using Tagdust (1.12) (Lassmann et al., 2009). **Alignment:** High quality reads were aligned to the *C. elegans* ce10 genome using Tophat2 (2.0.14), (Kim et al., 2013). Resulting bam files were normalized to read depth and converted to bigwig format using bedtools genomecov (2.22.1) (Quinlan et al., 2010) followed by bedGraphToBigWig (UCSC Genome Browser) conversion to bigWig format. Replicate bigwig files were averaged using java-genomics-toolkit wigmath.Average (java/1.8.0_11), (http://palpant.us/java-genomics-toolkit/).

Publicly available dataset

Dataset submitted to NCBI/GEO:
01_EO025_rep1_ELT2.fastq.gz
01_EO027_rep1_IgG.fastq.gz
01_EO028_rep1_input.fastq.gz
01_EO036_rep2_H3K4me3.fastq.gz
01_EO037_rep2_ELT2.fastq.gz
01_EO041_rep2_input.fastq.gz
01_AR135_rep3_H3K4me3.fastq.gz
01_AR136_rep3_ELT2.fastq.gz
01_AR137_rep3_IgG.fastq.gz
01_AR138_rep3_input.fastq.gz
01_AR501_L3_mRNA.fastq.gz
01_AR505_L3_mRNA.fastq.gz
08_EO025_rep1_ELT2.bw
08_EO027_rep1_IgG.bw
08_EO028_rep1_input.bw
08_EO036_rep2_H3K4me3.bw
08_EO037_rep2_ELT2.bw
08_EO041_rep2_input.bw
08_AR135_rep3_H3K4me3.bw
08_AR136_rep3_ELT2.bw
08_AR137_rep3_IgG.bw
08_AR138_rep3_input.bw
12_EO025_EO037_AR136_average_ELT2_track.bw

12_EO036_AR135_average_H3K4me3_track.bw
14_ELT2_IggBlackHOTMinus_gt30_peaks.bed
14_ELT2_IggBlackHOTMinus_gt30_summits.bed
14_H3K4me3_IggBlackHotMINUS_gt30_peaks.bed
14_H3K4me3_IggBlackHOTMinus_gt30_summits.bed
05_AR501_L3_mRNA.bw
05_AR505_L3_mRNA.bw



Supplementary Figure S1. A. ELT-2 protein is present in the adult hermaphrodite intestine. Immunofluorescence staining using anti-ELT-2 monoclonal antibody 455-2A4 on a squashed adult hermaphrodite. DAPI staining is pseudo-coloured red; ELT-2 immunofluorescence is green. B. Test for specificity of the anti-ELT-2 monoclonal antibody 455-2A4. Embryos produced by strain JM147 *elt-2*(*ca15*); *caEx3*[pJM276 = *elt-2* genomic region; pTG96-2 = *sur-5*promoter::GFP; pRF4 = *rol-6(su1006)*]

were stained for ELT-2 with 455-2A4 (secondary antibody tagged red) and with anti-GFP (secondary antibody tagged green). The images show six morphogenesis stage embryos (past the stage when expression of both *elt-2* and *sur-5* initiate). Three embryos stain green and hence contain the *sur-5*promoter::GFP rescuing array; these same three embryos also stain red and hence express ELT-2. Three embryos do not stain green, are not rescued and hence are *elt-2*(-); these embryos do not stain red for ELT-2. On the same slide, and assaying only embryos at the 8E cell stage and later, we counted zero Red-NOT-Green embryos compared to 120 Red-AND-Green embryos.

A. Conserved Region I (CR I). Coordinates can be translated into coordinates upstream of the ATG as follows: for *C. briggsae elt-2* CR I, "1" corresponds to 670"bps upstream of ATG. For *C. elegans elt-2* CR I, "1" corresponds to "575" bps upstream of ATG. Three conserved TGATAA sites and 1 (reverse) AGATAG sites are indicated.

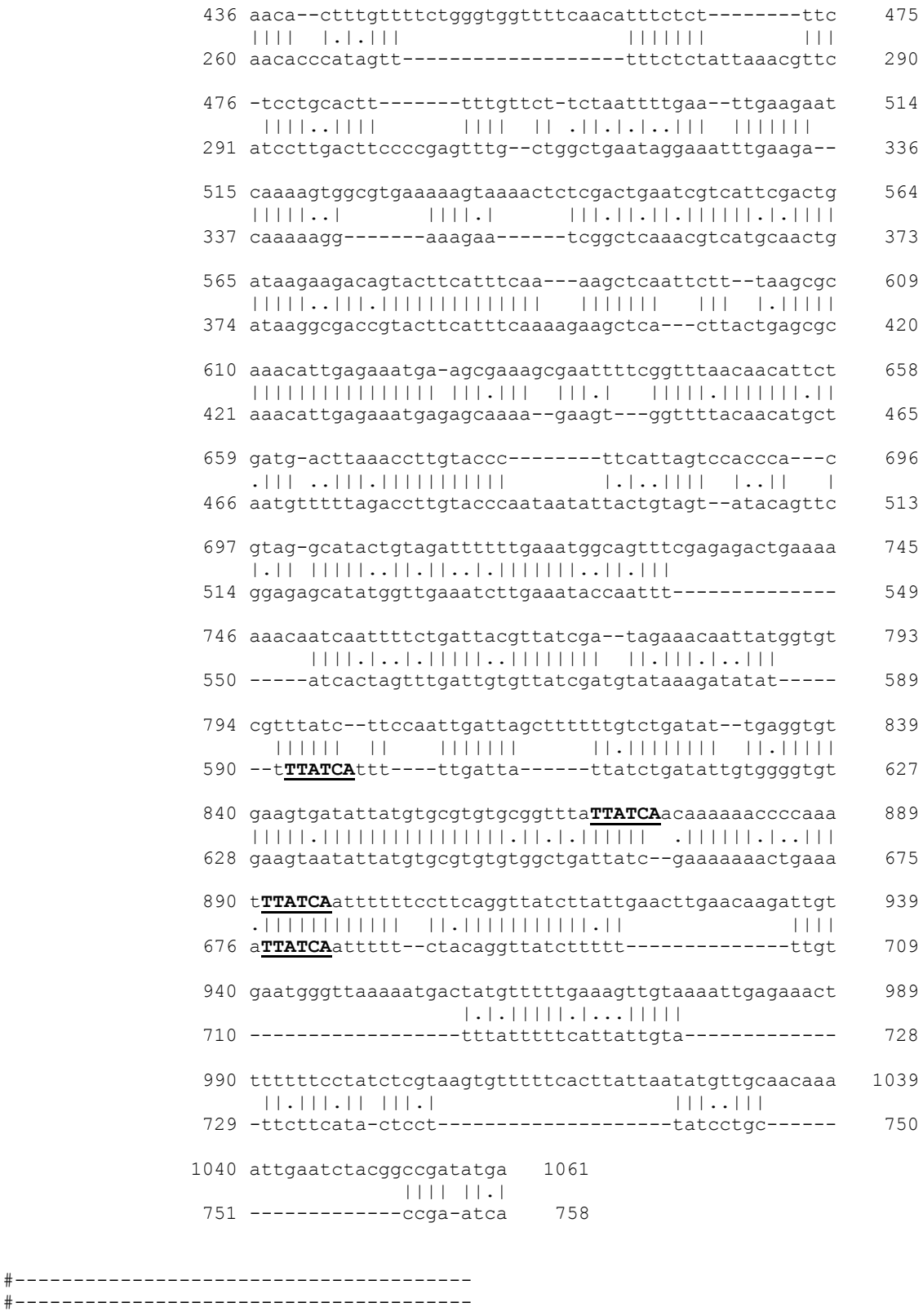
```
#=======================================
#
# Aligned_sequences: 2
# 1: C. briggsae elt-2 conserved region I
# 2: C. elegans elt-2 conserved region I
# Matrix: EDNAFULL
# Gap_penalty: 10.0
# Extend_penalty: 0.5
#
# Length: 803
# Identity:     364/803 (45.3%)
# Similarity:   364/803 (45.3%)
# Gaps:         358/803 (44.6%)
# Score: 828.5
#
#=======================================

       C. briggsae 1 taagtaggcggtctgaaaacctgtgctgaaaactctgagctgtctgcgtg     50
                     .||||.||.|||                              |||||||||.
       C. elegans  1 gaagtgggtggt----------------------------tgtctgcgta     22

                  51 tatggtttt-------gaggaatcgtcgc----acTGATAAgggctgtgt     89
                     ||||            ||||  |||..||    |||||||||
                  23 tatg-----agcgacagagg--tcggggctgaaacTGATAAg--------     57

                  90 gtgca----tcg---ctaacggcataatcgccagcacagcagtcaaaaaa    132
                         |    |||   ||||||.|||||||||.||| ||||..|||
                  58 ----aatagtcgacactaacgccataatcgctagc-cagccatca-----     97

                 133 gagagggagtgggagacagagagcagtacaacctctcact-tgcacgcca    181
                              .|.|.||.|||             |||..| ||||
                  98 ----------tgcacaccgag-------------ctcggtgtgca-----    119

                 182 ataccaccgtcttcttctacttatttcattcca-----------gccctc    220
                         ||||.|||  ||||     |||||..|||           |||.||
                 120 ----caccatct--ttct-----tttcaaaccaatacgctttgtgccttc    158

                 221 att----atttcttgtcagttatacacatcaaacacttttttTGATAAaa    266
                     |||    ||||                          |.|||||||||||
                 159 attgacaattt------------------------tctttTGATAAaa    183

                 267 tcaacCTATCTatacttccc----------agtcttatcgttgcaaggcc    306
                     |||.|||||||||||||||          |||||||||||||.|..||.
                 184 tcagcCTATCTatacttcccaatcatttttagtcttatcgttgaacagct    233
```

```
307 ctcaagttactactgttccaactgatatcttctagattgcgatatcagga    356
    .||.||.|.|.|||||||...|||||||||||||..||.|.||..||..|
234 atcgaggtgccactgttttcactgatatcttctaagttactatggcatta    283

357 accggtgaaagtgatagggaataattcagagcgttgatttgcaaggagtg    406
    ||                  ||..||    |.|||..|.|
284 ac------------------atcttt----gtgttcttat----------    301

407 ttgatggggaatcgaatgttt--tgaa--tttgaacttaa------caat    446
    ||.||||||..||      |||  |.||  ||||.|.||||      .|||
302 tttatgggttat------tttaattaatttttgcagttaatttttggaat    345

447 gagcaatataaaactgttgattgaaaaatcccctgttctttctg------    490
    ||||||   .|||.||||.||||.||.|||     ||| .||||
346 gagcaa---gaaaatgttaattgtaatatc-----ttc-gtctgaaaatt    386

491 --cttc-----ctgttca---taaacttccggaggcacacgtcattt---    527
      ||||     .|.||.|   |||        |||||   || ||||
387 gtcttcaaatagttttaattttaa--------aggca---gt-atttaag    424

528 ------ccattcccaggtcgaggggtagcctctcattttcttga--ttca    569
          |..|||     ||||.|          |||||  ||||  ||..
425 aaaatacacttc-----tcgaag----------cattt--ttgaattttt    457

570 aaatttttt------tttgatg-------------------aacTGATAAc    594
    .|||||||      .||||||                    .|.||||||
458 gaatttttaaactgcttgatgttttaggtgccactgttttcatTGATAA-    506

595 gaaaaccaatcgtcccctcatttcatgaat-----ct---ttttaactt-    635
                ||       |||.|||.||     ||   ||||  |||
507 -----------gt-------tttgatgtataaatgcttgatttt--cttg    536

636 ---tcccaatactcccctaattaccg-tactcttgcaga-----------    673
       |.|.||||       ||.||  | .|||     ||||..
537 gcattctaata-------aaata--gaaact-----agaaaatagattat    572

673 ---    673

573 aga    575
```

```
#---------------------------------------
#---------------------------------------
```

B. Conserved Region II (CR II). Coordinates can be translated into coordinates upstream of the ATG as follows: for *C. briggsae elt-2* CR II, "1" corresponds to "2746"bps upstream of ATG. For *C. elegans elt-2* CR II, "1" corresponds to "2226" bps upstream of ATG. Three conserved TGATAA sites are shown, as is a TGATAA site in *C. elegans* and a nearby CGATAA site in *C. briggsae*.

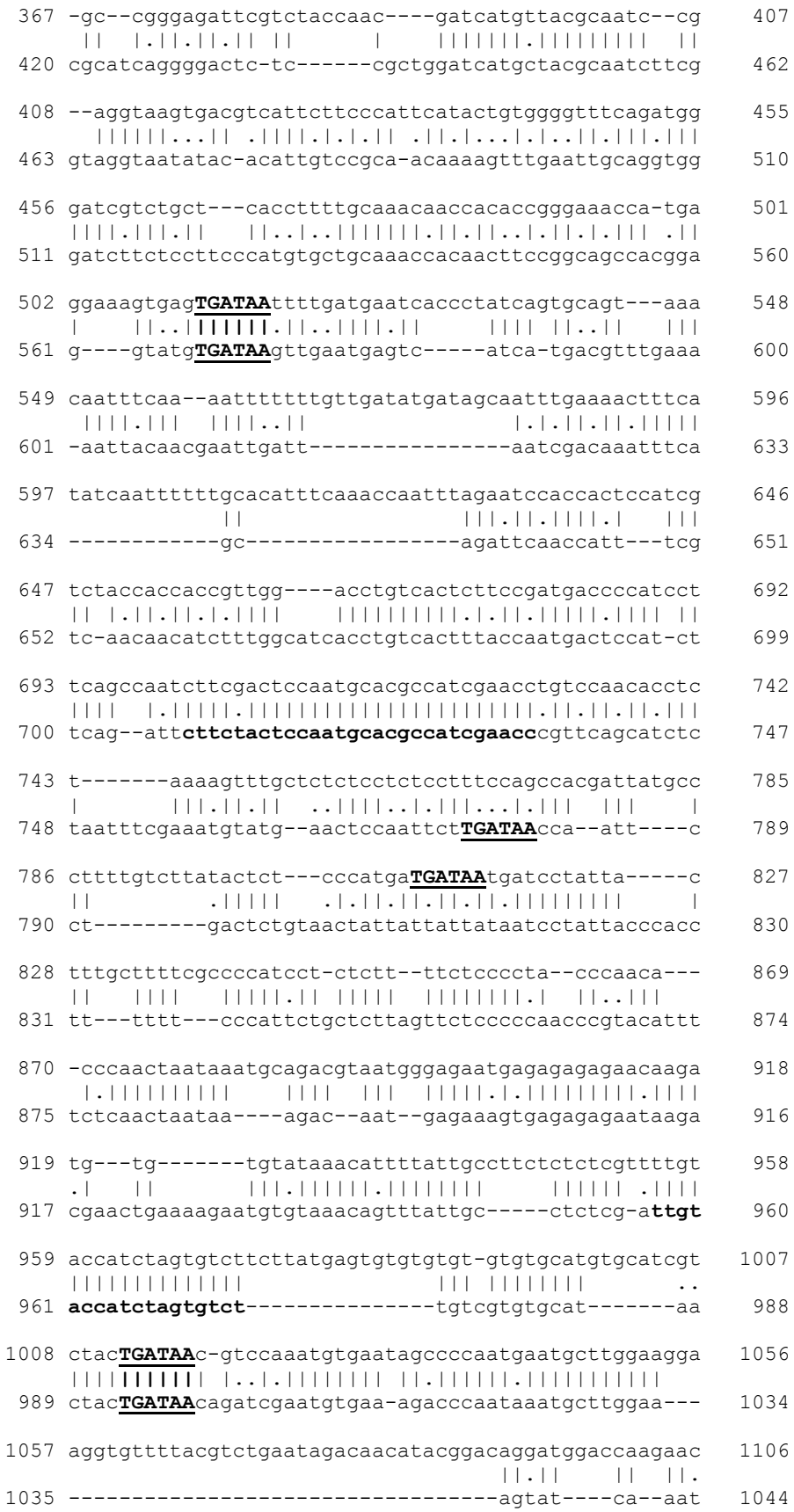
```
#=======================================
#
# Aligned_sequences: 2
# 1: C. briggsae elt-2 conserved region II
# 2: C. elegans elt-2 conserved region II
# Matrix: EDNAFULL
# Gap_penalty: 10.0
# Extend_penalty: 0.5
#
# Length: 1122
# Identity:     588/1122 (52.4%)
# Similarity:   588/1122 (52.4%)
# Gaps:         425/1122 (37.9%)
# Score: 1742.0
#
#=======================================

       C. briggsae 1 ---atgtccttttctcactaaaacgga-----acagca-tgtgaatgcag     41
                       .||||.|||           ||||     |.||.| |||||  ||||
       C. elegans  1 attgtgtcattt-----------cggatatagagagtagtgtga--gcag     37

                  42 tttaaacttttttatgactaggttatctgttcgcgattctttgaaaaatca     91
                                       ||.||||        ||||      ||
                  38 ------------------------atgtgtt--------tttg------ca     50

                  92 gaaattttgaaataagtccgtccgggcgagaaattggacagaaaagttgg    141
                     |||.|                          ||||||
                  51 gaact--------------------------aaattg--------------     61

                 142 tacaaaacttaagaatcaataagacactctggaattttcgaaactgcaag    191
                                        |.||||||     |||.|.|||
                  62 ---------------------atgacact-----attatggaa--------     78

                 192 gagcatgacaagtcttcaaaatagagaagcaaaaacaactaaaaacactt    241
                                                       ||.||
                  79 ------------------------------------------tataa------     83

                 242 ttcaacaaatgatcaccaccacattggtgaaaatgtctggtatcaaaacc    291
                      |.|.||||||.|       |.||  |||..||.||.||.||||||||||
                  84 -tgaccaaatgtt-------aaat--gtgttaaggtttgatatcaaaacc    123

                 292 ggtattctctttttatacagaatTGATAAtgttatcttcaattgatttct    341
                     .|||||.||||||||||||||||||||||||||||||||||||||||||||
                 124 tgtattttctttttatacagaatTGATAAtgttatcttcaattgatttct    173

                 342 cttccaaccatttctgagctacggcgatatca--acgcaatcttctccaC    389
                              |.||||||||||||||||..|  |||||   |||||.||
                 174 ---------acttctgagctacggcgatacgaggacgca---ttctcaac    211

                 390 GATAAtgttgccataattt---ctgatta-tcagttactgatagtttcag    435
                     ||||||||||||||  |||   |||||.| |...||||||||.|||||||
                 212 gataatgttgccat--tttgtccTGATAAttttttactgattgtttcag    259
```


436 aaca--ctttgttttctgggtggttttcaacatttctct--------ttc 475
260 aacacccatagtt-------------------tttctctattaaacgttc 290
476 -tcctgcactt-------tttgttct-tctaattttgaa--ttgaagaat 514
291 atccttgacttccccgagtttg--ctggctgaataggaaatttgaaga-- 336
515 caaaagtggcgtgaaaaagtaaaactctcgactgaatcgtcattcgactg 564
337 caaaaagg-------aaagaa------tcggctcaaacgtcatgcaactg 373
565 ataagaagacagtacttcatttcaa---aagctcaattctt--taagcgc 609
374 ataaggcgaccgtacttcatttcaaaagaagctca---cttactgagcgc 420
610 aaacattgagaaatga-agcgaaagcgaattttcggtttaacaacattct 658
421 aaacattgagaaatgagagcaaaa--gaagt---ggttttacaacatgct 465
659 gatg-acttaaaccttgtaccc--------ttcattagtccaccca---c 696
466 aatgtttttagaccttgtacccaataatattactgtagt--atacagttc 513
697 gtag-gcatactgtagatttttgaaatggcagtttcgagagactgaaaa 745
514 ggagagcatatggttgaaatcttgaaataccaattt-------------- 549
746 aaacaatcaattttctgattacgttatcga--tagaaacaattatggtgt 793
550 -----atcactagtttgattgtgttatcgatgtataaagatatat----- 589
794 cgtttatc--ttccaattgattagctttttgtctgatat--tgaggtgt 839
590 --tTTATCAttt----ttgatta------ttatctgatattgtggggtgt 627
840 gaagtgatattatgtgcgtgtgcggtttaTTATCAacaaaaaaaccccaaa 889
628 gaagtaatattatgtgcgtgtgtggctgattatc--gaaaaaaactgaaa 675
890 tTTATCAatttttttccttcaggttatcttattgaacttgaacaagattgt 939
676 aTTATCAattttt--ctacaggttatctttttt--------------ttgt 709
940 gaatgggttaaaaatgactatgtttttgaaagttgtaaaattgagaaact 989
710 ------------------tttatttttcattattgta------------- 728
990 ttttttcctatctcgtaagtgtttttcacttattaatatgttgcaacaaa 1039
729 -ttcttcata-ctcct--------------------tatcctgc------ 750
1040 attgaatctacggccgatatga 1061
751 -------------ccga-atca 758

#----------
#----------

C. Conserved Region III (CR III). Coordinates can be translated into coordinates upstream of the ATG as follows: for *C. briggsae elt-2* CR III, "1" corresponds to "4099"bps upstream of ATG. For *C. elegans elt-2* CR III, "1" corresponds to "4387" bps upstream of ATG. Three conserved TGATAA sites are shown, as is a TGATAA site in *C. elegans* and a CGATAG site in *C. briggsae*.

```
#=======================================
#
# Aligned_sequences: 2
# 1: C. briggsae elt-2 conserved region III
# 2: C. elegans elt-2 conserved region III
# Matrix: EDNAFULL
# Gap_penalty: 10.0
# Extend_penalty: 0.5
#
# Length: 1366
# Identity:     768/1366 (56.2%)
# Similarity:   768/1366 (56.2%)
# Gaps:         409/1366 (29.9%)
# Score: 2072.0
#
#=======================================

      C. briggsae  1 caagagacaatcctaaacccgattatttcatcagaatctgtgtatatttg     50
                     ||| ||..||||.|.|.|       |||||  .||||.||||||||   |
      C. elegans   1 caa-agcgaatcttcagc-------tttca--cgaatttgtgtata---g     37

                   51 --ggacctaagatggatattc-tccccgttaagaa-tgccaaaagaagat     96
                        ||||||.|||.||||||.| |  |.|||.|||| |||| |||||||||
                   38 gaggacctgagacggatatacat--cggtttagaagtgcc-aaagaagat     84

                   97 ttgatgtgcgtcacaaacaacgtaaaatg-gggtccgagaagagcccatt    145
                      ||.|||||||||||||.||.||||||||| ||.|.|.|||.| |.||||..
                   85 ttaatgtgcgtcacaatcatcgtaaaatgcggcttcaagacg-ggccagg    133

                  146 tgctcatttgttc-------------------tcaa--------------    162
                      .|||||||| |||                   ||||
                  134 agctcattt-ttcaattttatactgccacctgtcaacaattctcttgtgc    182

                  163 ----------------aact-----atggtcgacttgaaagctattgcct    191
                                      ||||     |||||.|||.|.|||.||.|..|.|
                  183 attcctaacaactgagaacttaaaaatggtagacattaaaactttaactt    232

                  192 ctcccaaggtagctgttccaattgctggaattgtgattctcattgcggtc    241
                      |.||.||.||.||..||||.|||...|||||.|||.||.|..|.|..|||
                  233 cccctaaagttgcaattcctattataggaatcgtggttttagtggtcgtc    282

                  242 tccacgttggtagcgatattctatatcaggtaagc----cacca------    281
                      ||.||||||||.||.||.||||||.|.||||||.|    ||..|
                  283 tcaacgttggtcgcaatgttctatgttaggtaatcattacaatatgatag    332

                  282 -aattgtgttcttcaaaaaaat---tta--agaactccagaaacccaaca    325
                       |||||       |..|||.||   |||  |..|.|.||||||||||||.
                  333 gaattg-------ccgaaagatgaattactactattacagaaacccaacg    375

                  326 gtaacCGATAGcctctttggcactggtggcaattctaccga---------    366
                      ||.||.||||...|.|||||||.|    ||||  .|||.||
                  376 gtcacTGATAAtatttttggcaat----gcaa--atacggatgagaaaga    419
```


367 -gc--cgggagattcgtctaccaac----gatcatgttacgcaatc--cg 407
420 cgcatcaggggactc-tc------cgctggatcatgctacgcaatcttcg 462
408 --aggtaagtgacgtcattcttcccattcatactgtggggtttcagatgg 455
463 gtaggtaatatac-acattgtccgca-acaaaagtttgaattgcaggtgg 510
456 gatcgtctgct---caccttttgcaaacaaccacaccgggaaacca-tga 501
511 gatcttctccttcccatgtgctgcaaaccacaacttccggcagccacgga 560
502 ggaaagtgagTGATAAttttgatgaatcaccctatcagtgcagt---aaa 548
561 g----gtatgTGATAAgttgaatgagtc-----atca-tgacgtttgaaa 600
549 caatttcaa--aatttttttgttgatatgatagcaatttgaaaactttca 596
601 -aattacaacgaattgatt----------------aatcgacaaatttca 633
597 tatcaatttttttgcacatttcaaaccaatttagaatccaccactccatcg 646
634 ------------gc-----------------agattcaaccatt---tcg 651
647 tctaccaccaccgttgg----acctgtcactcttccgatgaccccatcct 692
652 tc-aacaacatctttggcatcacctgtcactttaccaatgactccat-ct 699
693 tcagccaatcttcgactccaatgcacgccatcgaacctgtccaacacctc 742
700 tcag--attcttctactccaatgcacgccatcgaaccgttcagcatctc 747
743 t-------aaaagtttgctctctcctctcctttccagccacgattatgcc 785
748 taatttcgaaatgtatg--aactccaattctTGATAAcca--att----c 789
786 cttttgtcttatactct---cccatgaTGATAAtgatcctatta-----c 827
790 ct---------gactctgtaactattattattataatcctattacccacc 830
828 tttgcttttcgccccatcct-ctctt--ttctccccta--cccaaca--- 869
831 tt---tttt---cccattctgctcttagttctcccccaacccgtacattt 874
870 -cccaactaataaatgcagacgtaatgggagaatgagagagagaacaaga 918
875 tctcaactaataa----agac--aat--gagaaagtgagagagaataaga 916
919 tg---tg-------tgtataaacattttattgccttctctctcgttttgt 958
917 cgaactgaaaagaatgtgtaaacagtttattgc-----ctctcg-attgt 960
959 accatctagtgtcttcttatgagtgtgtgtgt-gtgtgcatgtgcatcgt 1007
961 accatctagtgtct---------------tgtcgtgtgcat-------aa 988
1008 ctacTGATAAc-gtccaaatgtgaatagccccaatgaatgcttggaagga 1056
989 ctacTGATAAcagatcgaatgtgaa-agacccaataaatgcttggaa--- 1034
1057 aggtgttttacgtctgaatagacaacatacggacaggatggaccaagaac 1106
1035 ----------------------------------agtat----ca--aat 1044

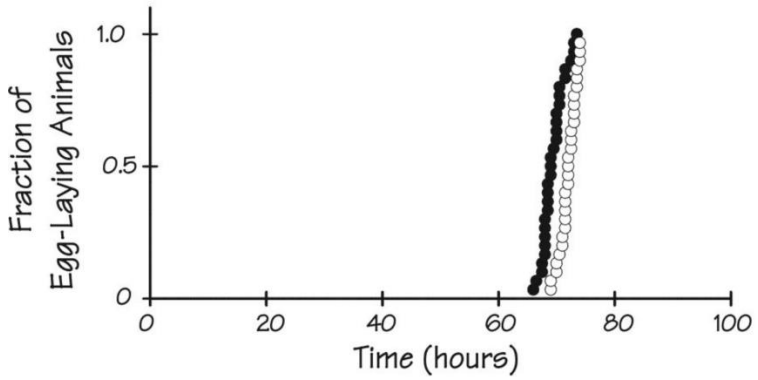
```
1107 atgggccgtagtcaaacgattgtaacctgaaattctgaaaatatttatca   1156
     |||                 |.||    |||||      |.|||||| .|
1045 atg-----------------ttta----gaaat------agtattta-ga   1066

1157 g-ctt------gacaaaggcactttcttaaatgcctaaaacagta--atg   1197
     | |||      |||.||                   |||||||.|  ||.
1067 gacttatgtgggacgaa-------------------aaaacaggagcatt   1097

1198 gacaat-cagt--caa   1210
     .|.||| ||||  |||
1098 aagaatacagttgcaa   1113


#---------------------------------------
#---------------------------------------
```

Supplementary Figure S2. Global pairwise alignments (Needleman-Wunsch algorithm as implemented in EMBOSS) between the three conserved regions of the *C. elegans elt-2* promoter (lower sequence) and the *elt-2* promoter of *C. briggsae* (upper sequence). Coordinates are base pairs upstream of the ATG (with the base pair immediately upstream of Atg counted as "0"). TGATAA sites in both sequences are bold underlined upper case.

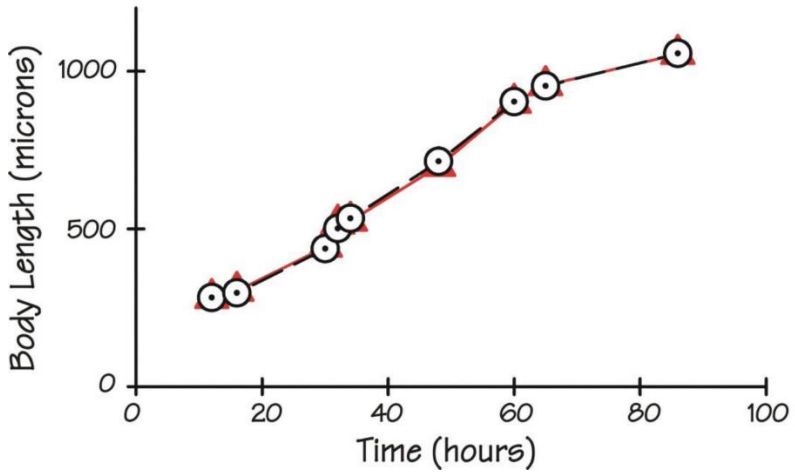
A. Strain VC231 (*gk153*) shows comparable defecation timing, brood sizes and embryonic lethality to wildtype controls.

Strain	Defecation Cycle (secs)	Brood Size	Embryonic Lethality
N2	42.2 ±1.8	226.6±23.7	0.8 %
VC231	43.0±2.7	209.2±15.9	0.2 %

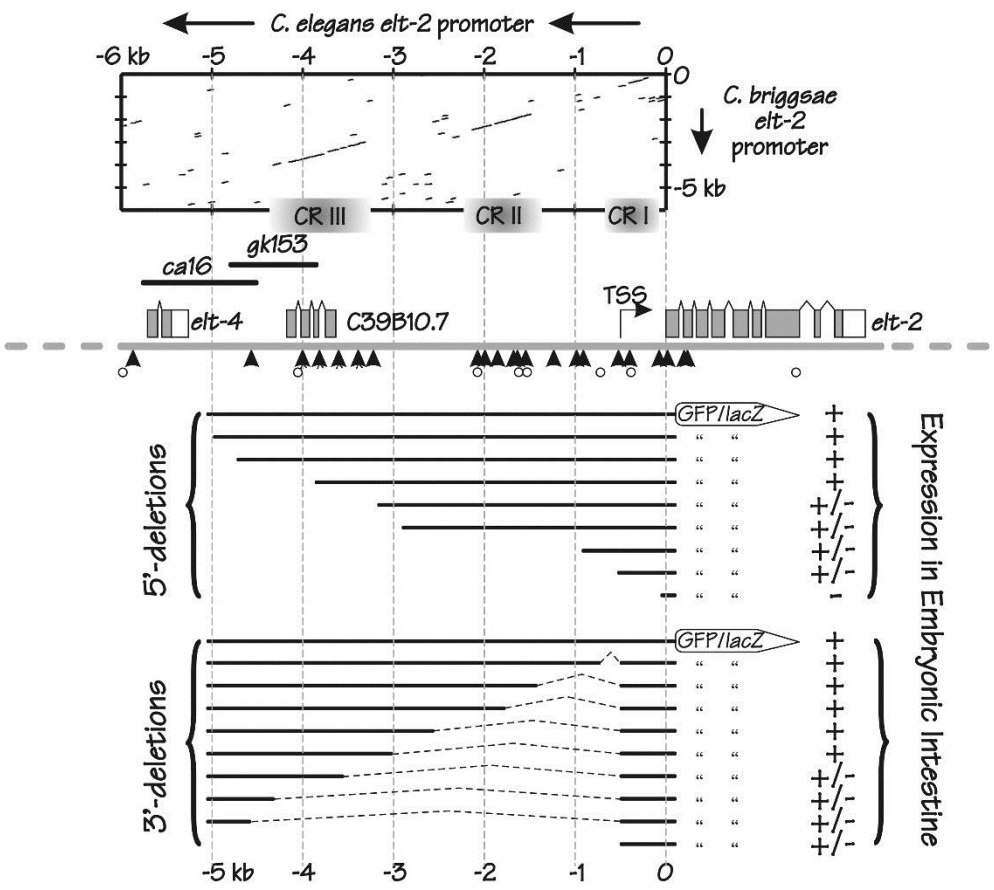
B. Egg-to-egg time (hours) for strain VC231 (*gk153*) (open symbols) and N2 controls (closed symbols). Thirty animals were assayed.



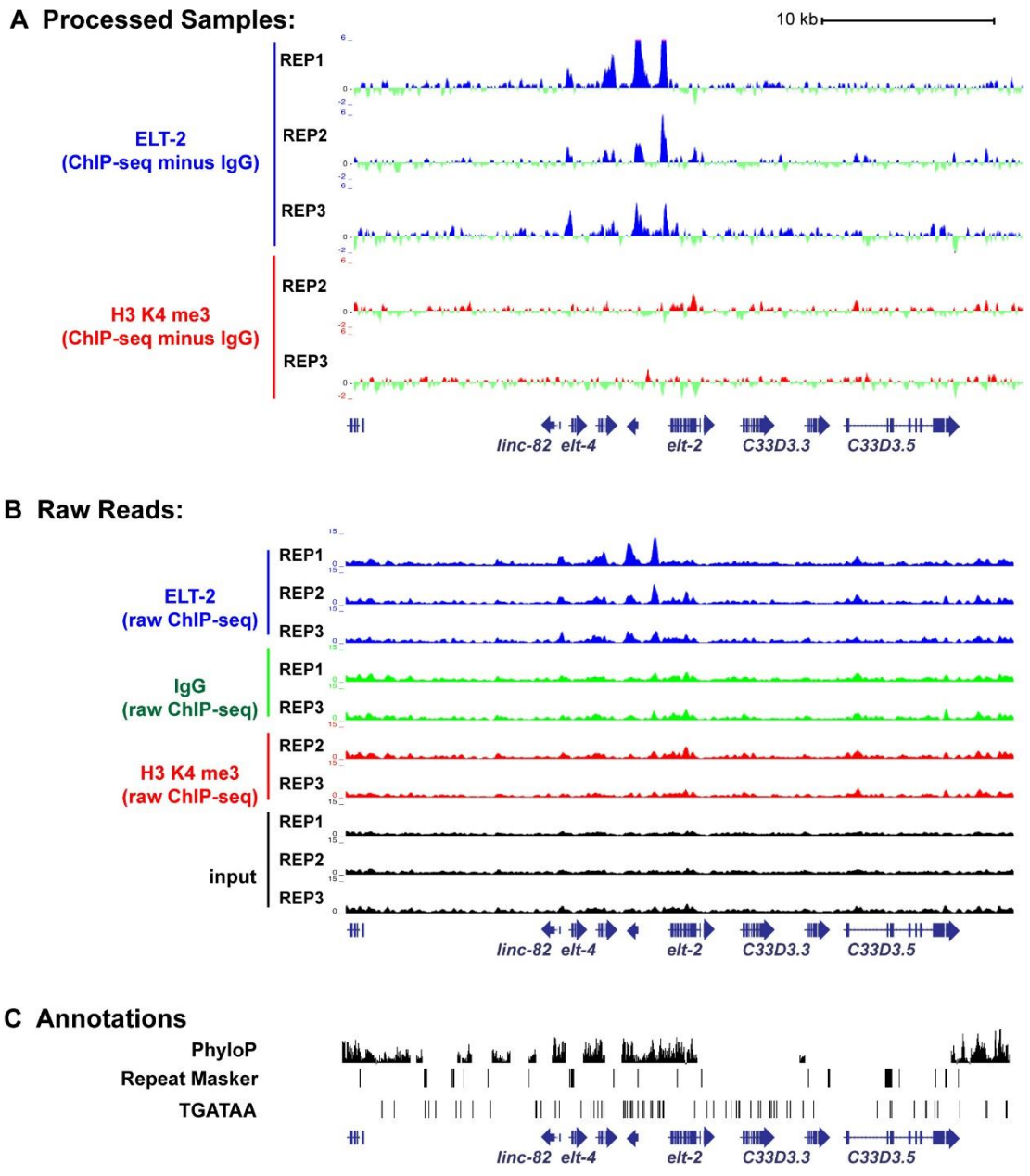
C. Growth curves (body length in microns) for strain VC231(*gk153*) (open symbols) and for N2 control (red symbols and line). Nine animals were measured for each strain at each time point.



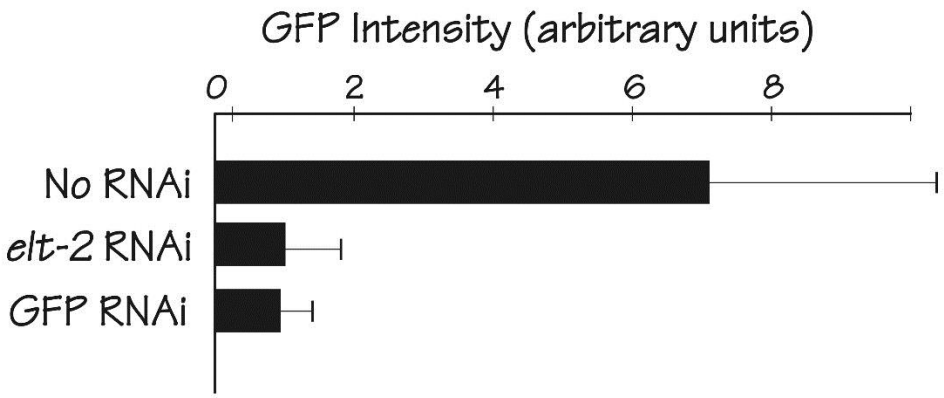
Supplementary Figure S3. Comparison of strain VC231(*gk153*) to N2. Temperature = 20ºC A. Strain VC231 (*gk153*) shows comparable defecation timing, brood sizes and embryonic lethality to wildtype controls. B. Egg-to-egg time (hours) for strain VC231 (*gk153*) (open symbols) and N2 controls (closed symbols). Thirty animals were assayed. C. Growth curves (body length in microns) for strain VC231 (*gk153*) (open symbols) and for N2 control (red symbols and line). Nine animals were measured for each strain at each time point.



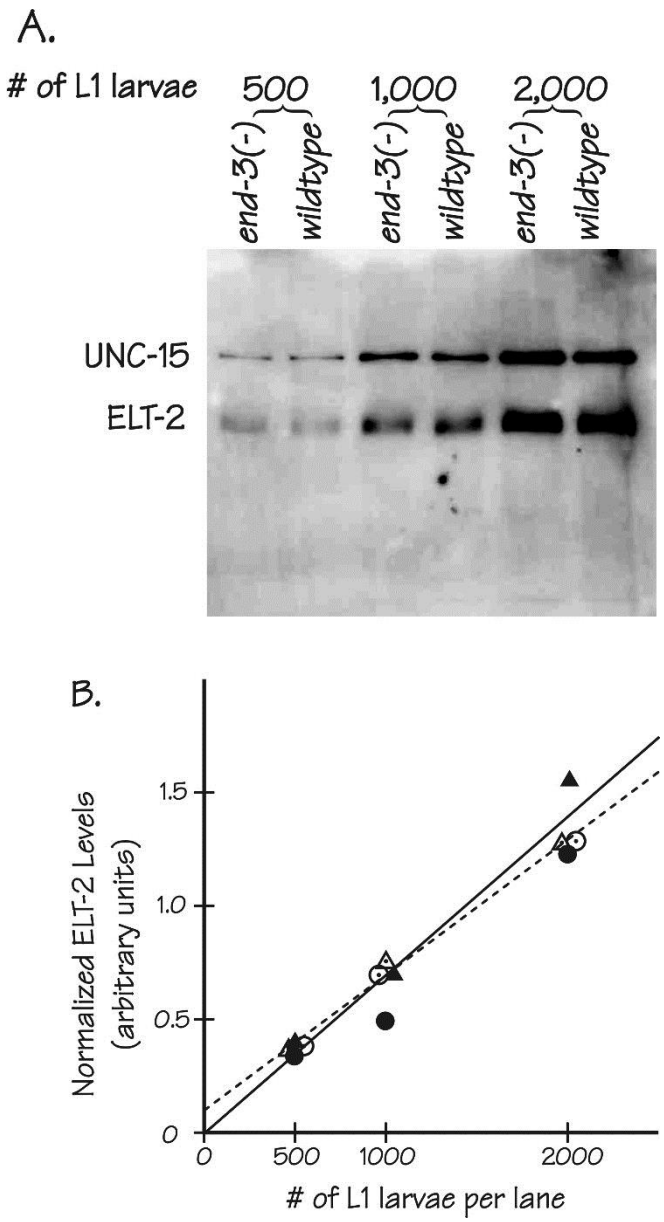
Supplementary Figure S4. Results of two series of transgenic deletion constructs (GFP/lacZ reporter); the 5'-deletion series begins 5 kb upstream of *elt-2* and proceeds toward the gene; the 3'-deletion series all retain the proximal 0.5 kb of the *elt-2* promoter with deletions proceeding upstream. Expression of the transgenic reporters was evaluated in embryos and is reported as "~ wild type" = +, or "sporadic, weak or ectopic" = +/-. In both series, expression patterns change when deletions proceed through the proximal region of CR III.



Supplementary Figure S5. Raw reads and processed reads of individual replicates of ELT-2 ChIP-seq. A. Processed replicates of ELT-2 and H3K4me3 ChIP-seq samples are shown. Processing involved subtracting IgG raw tracks from ChIP-seq raw tracks. The averages of these tracks are illustrated in Figure 5. B. Raw reads that generated the processed samples are shown. Tracks were normalized for read-depth only. (PDF FILE)



Supplementary Figure S6. ELT-2 activates the *elt-4* gene. Strain JM119 contains an integrated transgenic array incorporating plasmid pJM188 (~4 kb 5'-flanking and ~ 6 kb 3'-flanking regions of the *elt-4* gene, with GFP inserted at the *elt-4* C-terminus (Fukushige et al., 2003)). RNAi was performed by feeding against GFP (positive control) and against *elt-2*; no RNAi negative controls were performed in parallel. Total intestinal GFP intensities (arbitrary units) were measured using FIJI on 26-28 animals for each condition. Error bars correspond to standard deviations.



Supplementary Figure S7. A. Western blots to measure the ELT-2 levels in *end-3(-)* L1 larvae and in wildtype animals. The ELT-2 protein in extracts of 500, 1000 and 2000 L1 larvae from each strain were detected with anti-ELT-2 antibody 455-2A4 followed by an HRP-tagged secondary antibody. B. To correct for loading errors and to demonstrate linearity of response, ELT-2 band intensities were quantitated for each lane of replicate gels and normalized to the intensity of the paramyosin band (UNC-15; detected by monoclonal antibody MH16.) measured for 2,000 larvae in the same set of samples. Circles and triangles represent replicate gels. Open symbols = wildtype larvae; closed symbols = *end-3(-)* larvae. Dashed line and solid line represent linear regression fit to wildtype data and *end-3(-)* data, respectively.

Supplementary Table S1. Summary of PCR primers used to detect gene deletions during strain construction. Mutant alleles are listed in the Methods section.

Primer Name	Sequence	Short Description	Product Length (bps)	
			Wildtype	Mutant
oJM345	TGCAAGTGAGTTGAGGTTTTTG	*end-1* forward	1255	376
oJM346	CCCCATCCCAGTGTAGGAG	*end-1* reverse		
oTW7	CACTCTCGCACGTGAAAAAC	*end-3* forward	2100	1400
oTW8	CAATGCCTGTCTTTTGAGCA	*end-3* reverse		
oAD15a	AGACCGTTTACCTTCCCAAAA	*elt-4* forward	1500	230
oAD15b	ACACAAATTCGTGAAAGCTGAA	*elt-4* reverse		
oJM314	CCAACTTTTGGCAACTTCTTG	*elt-7* forward	976	360
oJM315	CCGATTTTTCGGAAATTGAA	*elt-7* reverse		

Supplementary Table S2. ELT-2 ChIP-seq peaks**.** A text file containing the genome coordinates of all 624 ELT-2 ChIP-seq peak regions identified in this study, in BED format. (TXT FILE).

Click here to Download Table S2

Supplementary Table S3. ELT-2 ChIP-seq summits. A text file containing the genome coordinates of all 624 ELT-2 ChIP-seq peak summits (1 bp) identified in this study, in BED format. (TXT FILE).

Click here to Download Table S3