Feedforward regulatory logic underlies robustness of the specification-to-differentiation transition and fidelity of terminal cell fate during *C. elegans* endoderm development

Chee Kiang Ewe[1,†], Erica M. Sommermann[1,†], Josh Kenchel[2,3], Sagen E. Flowers[1], Morris F. Maduro[4], Joel H. Rothman[1,2]*

[1] Department of MCD Biology and Neuroscience Research Institute, University of California Santa Barbara, Santa Barbara, CA 93106

[2] Program in Biomolecular Science and Engineering, University of California Santa Barbara, Santa Barbara, CA 93106

[3] Chemical and Biomolecular Engineering Department, University of California Los Angeles, Los Angeles, CA 90095

[4] Molecular, Cell and Systems Biology Department, University of California Riverside, Riverside, CA 92521

[†] These authors contributed equally to this work and are listed alphabetically.

* Correspondence: rothman@lifesci.ucsb.edu

Running title: The gene regulatory logic of endoderm development

Keywords: gene regulatory network, GATA factors, cell specification, morphological differentiation, developmental robustness, endoderm

ORCiD:

CKE - 0000-0003-1973-1308

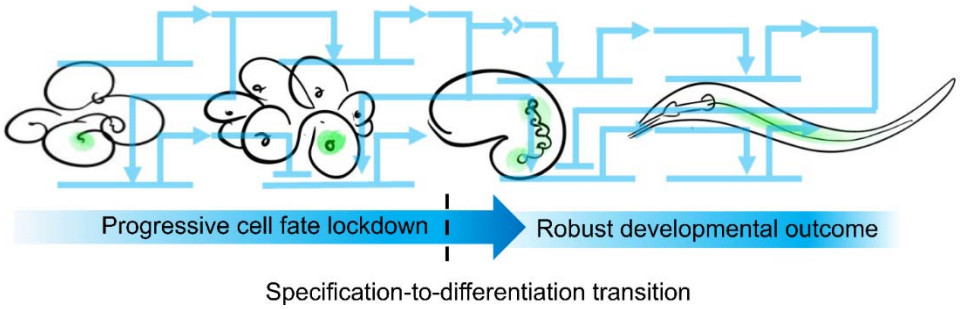
JK - 0000-0001-6378-6723

SEF - 0000-0002-7818-2188

MFM - 0000-0001-6257-7612

JHR - 0000-0002-6844-1377

Graphic abstract



Abstract

Development is driven by gene regulatory networks (GRNs) that progressively dictate specification and differentiation of cell fates. The architecture of GRNs directly determines the specificity and accuracy of developmental outcomes. We report here that the core regulatory circuitry for endoderm development in *C. elegans* is comprised of a recursive series of interlocked feedforward modules linking a cascade of six sequentially expressed GATA-type transcription factors. This structure results in a reiterated sequential redundancy, in which removal of a single factor or alternate factors in the cascade results in no, or a mild, effect on endoderm development and gut differentiation, while elimination of any two factors that are sequentially deployed in the cascade invariably results in a strong phenotype. The strength of the observed phenotypes is successfully predicted by a computational model based on the timing and levels of transcriptional states. The feedforward regulatory logic in the GRN appears to ensure timely onset of terminal differentiation genes and allows rapid and robust lockdown of cell fate during early embryogenesis**.** We further found that specification-to-differentiation transition is linked through a common regulator, the END-1 GATA factor that straddles the two processes. Finally, we revealed roles for key GATA factors in establishing spatial regulatory state domains by acting as transcriptional repressors that appear to define the boundaries of the digestive tract. Our findings support a comprehensive model of the core gene network that describes how robust endoderm development is achieved during *C. elegans* embryogenesis.

49

Introduction

Development is driven by progressive activation of transcriptional programs, or gene regulatory networks (GRNs) that direct cell specification and subsequent differentiation (Boveri, 1899; Davidson and Levine, 2008). The sequential restriction of cell identity and developmental potential through dynamic changes in transcriptional states was anticipated by Waddington's epigenetic landscape, a graphic metaphor describing canalization and robustness during development (Waddington, 1957).

The generation of diverse animal forms largely relies on a common genetic toolkit. The GATA transcription factors play a conserved role in the development of diverse cell types, including those of the endoderm, the first of the three germ layers to have evolved during the late Precambrian era (Hashimshony et al., 2015; Rodaway and Patient, 2001). In the diploblastic phyla Poriferans and Cnidarians, GATA factors have been found to be specifically expressed in the endoderm, suggesting that these transcriptional regulators may have driven the invention of the endoderm germ layer and gastrulation at the dawn of metazoan evolution (Martindale et al., 2004; Nakanishi et al., 2014). This association of GATA factors with endoderm development persists throughout metazoan phylogeny, generally through successive deployment of multiple GATA factors. This sequential use of GATA factors is particularly striking in *C. elegans*, in which a cascade of six GATA factor-like transcription factors regulate specification and differentiation of the endoderm (Maduro, 2017; Maduro and Rothman, 2002; McGhee, 2007). In *Drosophila*, the GATA factor Serpent specifies endodermal fate and activates the expression of a second GATA factor, dGATAe, which is essential for the terminal differentiation of the intestine. These factors can act across wide phylogenetic spans, as first demonstrated with the *C. elegans* END-1 GATA factor, which ectopically activates endoderm development when expressed in the prospective ectoderm of *Xenopus* (Shoichet et al., 2000). Similarly, overexpression of *serpent* or *dGATAe* causes ectopic endoderm differentiation in non-

endodermal lineages in *Drosophila*, as well as in *Xenopus*, further supporting the functional conservation of the GATA factors (Murakami et al., 2005; Okumura et al., 2005). In sea urchin, Blimp1/Krox1 activates *otx1*, the product of which activates *gatae* expression (Davidson et al., 2002). Gatae in turn provides a positive input to *otx1*, in addition to activating the transcriptional program for endoderm development, thereby forming a stable circuit in the GRN (Davidson et al., 2002; Yuh et al., 2004). Accordingly, knocking down *gatae* severely blocks endoderm development and gastrulation (Davidson et al., 2002).

The endoderm in *C. elegans*, which arises from a single blastomere born at the 8-cell stage, the E cell (Boveri, 1899; Sulston et al., 1983), provides a highly tractable system for investigating the mechanisms of cell specification, differentiation, and organogenesis. This progenitor cell gives rise to a clone 20 cells comprising the intestine, which are arranged in nine rings (int1-9) spanning the length of the animal (Supplementary Figure 1A). Endoderm development is driven by pairs of duplicated genes encoding GATA-like transcription factors: the divergent MED-1/2 factors and the canonical END-1/3 and ELT-2/7 factors. The maternally provided SKN-1/Nrf transcription factor activates MED-1 and -2, which activate the specification of mesendodermal fate in the EMS blastomere (Bowerman et al., 1992). In the anterior daughter of EMS, the MS cell, the Wnt effector POP-1/Tcf represses the expression of *end-1/3*, and MED-1 and -2 activate *tbx-35*, whose product specifies mesodermal fate. In the E cell, the posterior daughter of EMS, a triply redundant signaling system (Wnt, Src, and MAPK) leads to the phosphorylation of POP-1 by nemo-like kinase LIT-1 (Bei et al., 2002; Maduro et al., 2002; Meneghini et al., 1999; Shin et al., 1999; Thorpe et al., 1997). Together with MED-1/2, Wnt-signaled POP-1 activates genes encoding the transiently expressed endoderm specification factors END-1 and -3, which in turn activate the expression of *elt-7* and *-2*, orthologues of vertebrate *GATA4/5/6*. Expression of the ELT factors is sustained throughout the remainder of the animal's life via a positive autoregulatory loop that "locks down" the differentiated state of

the intestine (Supplemental Figure 1B) (Maduro, 2017; Maduro and Rothman, 2002; McGhee, 2007). While *elt-7* loss-of-function mutants do not show a discernible phenotype, animals lacking ELT-2 arrest at early larvae stage (L1) owing to severely obstructed gut (Sommermann et al., 2010). Nonetheless, *elt-2(-)* mutant animals contain a well-defined intestinal lumen and the intestinal cells appear to be fully differentiated (Fukushige et al., 1998; Sommermann et al., 2010). In the absence of both ELT-2 and -7, however, the intestinal lumen is completely abolished and differentiation appears to proceed in only a subset of the endoderm-derived cells in a sporadic manner (Sommermann et al., 2010). This suggests that ELT-2 and -7 function synergistically in mediating morphological differentiation of the intestine, and additional input(s) mediate a bistable switch in the endodermal differentiation program in the absence of ELT-2 and -7 (Dineen et al., 2018; Sommermann et al., 2010).

In this study, we sought to decipher the functional requirements and interactions between the GATA transcription factors and how they allow rapid and faithful deployment of the endoderm GRN. We found that the endoderm GRN consists of a series of interlocking feedforward loops, creating “sequential redundancy” in the cascade and culminating in the rapid lockdown of cell fate. We further report that END-1 acts at a transition point, participating in both specification and differentiation. Finally, we demonstrated the important roles of GATA factors in safeguarding intestinal cell fate and defining the boundaries of the digestive tract. Overall, our findings reveal the nature of the extensive genetic redundancy in the regulatory circuitry and how this GRN architecture dictates robust cell specification and differentiation during embryonic development.

Methods and Materials

C. elegans cultivation and genetics

Worm strains were cultured using standard procedure (Brenner, 1974) and all experiments were performed at room temperature (20-23°C). All genetic manipulations were performed according to standard techniques (Fay, 2013). *him-5(-)* or *him-8(-)* was introduced into some strains to generate males and facilitate crosses. See Supplementary Table 1 for a complete list of strains used in this study.

Immunofluorescence analysis

Antibody staining with methanol-acetone fixation was performed as previously described (Sommermann et al., 2010). Antibodies MH27 (AB_531819), MH33, and 455-2A4 (AB_2618114) were used to detect AJM-1, IFB-2, and ELT-2, respectively. Alexa Fluor® 488 goat anti-mouse secondary antibody was used at 1:1000 dilution.

RNAi

RNAi feeding clones were obtained from the Ahringer (Kamath et al., 2003) or the Vidal (Rual et al., 2004) library. The bacterial strain was inoculated overnight at 37°C in LB containing 50 µg/ml ampicillin. The bacterial culture was then diluted 1:10 and incubated for an additional 4 h. Next, 1 mM of IPTG was added to the bacterial culture and 100 µL was seeded onto 35 mm NGM agar plates containing 1 mM IPTG and 25 µg/mL carbenicillin. For simultaneous knockdown of *elt-2* and *elt-7*, the two bacterial strains, each expressing dsRNA for one gene, were concentrated and resuspended in 1 mL of LB in 1:1 ratio before seeding the NGM plates. Seeded plates were allowed to dry for 48 h before use. Next, 10-20 L4 animals were placed on the RNAi plates. 24 h later, the animals were transferred to fresh RNAi plates to lay eggs. The progeny was then collected for analyses.

Imaging and fluorescence quantification

The animals were immobilized using 10 mM Levamisole and mounted on 4% agarose pads. Images were acquired, typically at 60X, using Nikon Eclipse Ti-E inverted microscope fitted with ORCA-Flash2.8 camera. For expression studies, maximum intensity Z-projection was generated on the Nikon NIS-Elements AR v4.13.05. Images were then analyzed using ImageJ or Imaris v9.7.2.

RT-qPCR

RNA was extracted from synchronized L1 animals using Monarch® Total RNA Miniprep Kit (#TS010S). cDNA synthesis was performed using SuperScript™ III First-Strand Synthesis SuperMix (Thermo Fisher Scientific, #18080400). Quantitative PCR was performed using BioRad CFX96 Real-Time System. Each 15 µL reaction contained cDNA, primers, and PowerUp™ SYBR™ Green Master Mix (Thermo Fisher Scientific, #A25743). The data were analyzed using the standard $2^{-\Delta\Delta CT}$ method (Livak and Schmittgen, 2001).

The primers sequences are:

act-1 - 5'-TCCATTGTCGGAAGACCACG-3' and 5'-GGTGACGATACCGTGCTCAA-3'

act-5 - 5'-GTCACTCACACCGTTCCAATC-3' and 5'-GTGAGGATCTTCATCATGTAGTCG-3'

Modeling endoderm gene regulatory circuits

The topology of the gene circuits with temporal information was written as a system of differential equations, with expression of each factor dependent on the concentration of its activators (See Supplementary Materials and Supplementary File 1). The gene cascade is initiated by SKN-1, which was modelled as a square wave in the EMS blastomere. Similarly, the positive inputs of (phosphorylated) POP-1 into *end-3* and *end-1* were modelled as a square wave in the E blastomere (23 mins after the four-cell stage). Model runs were calculated as

time-discretized Euler approximations (Hahn, 1991) with time steps of 0.01 s. An iterative least-squares algorithm following a modified Gauss-Newton method (Ruhe, 1979; Yip et al., 2010) was used to fit the model parameters to published transcriptomics data (Baugh et al., 2003; Tintori et al., 2016). The performance of the final model was then evaluated by comparing the predicted phenotypes of the single mutants with published results (Boeck et al., 2011; Dineen et al., 2018; Maduro et al., 2005a; Maduro et al., 2015). Finally, predictions of *elt-2* activation (a readout for the commitment to E fate) in the mutant combinations were generated by holding the concentration of knockout genes at zero and otherwise running the model as described. The source code is available on https://github.com/RothmanLabCode/endoderm_GRN_model.

Statistics and figure preparation

Statistics were performed using R software v3.4.1 (https://www.r-project.org/). The specific statistical tests were reported in the figure legends. Plots were generated using R package ggplot2 or Microsoft Excel. Figures were assembled in Inkscape v0.92.4 (https://inkscape.org/).

Results

Sequential redundancy suggests feedforward regulatory circuitry in the endoderm GRN

Owing to the extensive genetic redundancy, single mutants of most of the genes throughout the *C. elegans* endoderm GRN either show no overt phenotype (single *med* mutants, *end-1(-)*, *elt-7(-)*) or an extremely weakly penetrant phenotype (*end-3(-)*). Moreover, while the *elt-2(-)* mutant undergoes larval arrest immediately after hatching with a dysfunctional gut, the gut appears fully differentiated in these mutants. A strong gut differentiation defect requires the removal of both *elt-2* and *elt-7* (Figure 1A) (Fukushige et al., 1998; Sommermann et al., 2010). Similarly, strong phenotypes are observed only when both *meds* or both *ends* are removed in pairs. In this study, we sought to examine the basis for this extensive redundancy in the pathway and to illuminate how it might contribute to faithful specification and differentiation in endoderm GRN.

Temporally resolved transcriptomic analyses revealed that the GATA transcription factors are sequentially activated in the mesendoderm regulatory pathway (Baugh et al., 2003; Tintori et al., 2016). This observation suggests that, rather than pairs of factors (MEDs, ENDs, ELTs) acting together at specific tiers in the cascade (Supplementary Figure 1B), each factor is redundant with its immediate upstream or downstream factor in a continuously sequential cascade ("sequential redundancy"). For example, as is often seen in forward-driven biological switches, each may act through a feedforward motif in which a given factor activates both its immediate target gene and the target of that gene (Mangan and Alon, 2003).

To test the model that endoderm differentiation is directed by a series of sequentially redundant feedforward regulatory motifs, we constructed a series of double mutants and compared the penetrance of lethality and the extent of gut differentiation. We found that unlike the mild or undetectable phenotypes observed with single mutants, removing pairs of genes that

are expressed at sequential steps in the cascade invariably results in severely diminished viability (Figure 1A) and defects in endoderm specification/differentiation, as revealed by disrupted gut lumen morphology (Figure 1A, B). This effect is associated with sporadic expression and in some cases the complete absence of gut-specific rhabditin granules (Clokey and Jacobson, 1986) (Figure 1C). We found that the expression of the gut-specific IFB (Intermediate Filament, B)-2 protein is strongly diminished in the double mutants (Supplementary Figure 2A). Moreover, reduced AJM (Apical Junction Molecule)-1 expression suggests gut epithelialization defects in double mutants of sequential members in the endoderm GRN (Supplementary Figure 2B).

In *med-2(-); med-1(-)* double mutants, all embryos die as embryos as MS blastomere fails to be specified and its fate is transformed to that of the C-like mesectodermal progenitor in the early embryo (Figure 1A), as is also the case in embryos lacking the initiating, maternally provided SKN-1 transcription factor (Bowerman et al., 1992). The E cell similarly adopt a C cell fate in some, but not all, of the *med-2(-); med-1(-)* arrested embryos (Maduro et al., 2007). As previously reported (Maduro et al., 2005a), many *end-1(-) end-3(-)* mutants also die as embryos owing to transformation of the E cell into a C-like cell (Figure 1A). Thus, END-1 and -3 act as redundant selector genes that promote specification of endoderm over other cell fates.

Consistent with the sequential redundancy model, removal of two genes that function at alternate steps in the endoderm cascade does not result in the strong synergistic effect that we observed with removal of pairs of sequentially expressed genes (Figure 1A, E, F; Supplementary Figure 3). We observed normal expression of IFB-2 and AJM-1 along the length of the gut region in all such double mutants (Supplementary Figure 3). Further, while the *end-3(-); med-1(-)* double mutant, in which sequential genes in the cascade are removed, shows severe embryonic lethality, most (89.5%; n = 219) of the *end-1(-); med-1(-)* double mutants, in which alternate genes in the cascade are removed, develop into fertile adults (Figure 1A). This

observation is also consistent with the reported divergent roles of END-1 and -3 in endoderm specification (Boeck et al., 2011). Moreover, the *elt-7(-) end-3(-)* double mutants are largely viable, in contrast to the *end-3(-); med-1(-)* sequential double mutants, which show a strongly penetrant embryonic lethal phenotype (Figure 1A). Of the hatched L1 larvae, 31% of *end-3(-); med-1(-)* sequential double mutant animals exhibit no overt signs of gut differentiation, while only 2% of *elt-7(-) end-3(-)* alternate double mutants completely lack gut (Fisher's exact test; $p < 0.001$) (Figure 1D, G). However, many *elt-7(-) end-3(-)* mutants contain a partially differentiated gut (Figure 1G). We reasoned that the developmental defects observed in the most affected animals are the result of suboptimal expression of *end-1* in the absence of *end-3* (Maduro et al., 2007).

While neither single mutant shows a discernible phenotype, animals lacking both END-1 and ELT-7 are 100% inviable, and the arrested L1 larvae show a striking gut differentiation defect (Figure 1A, Supplementary Figure 2). The mutant worms contain patches of apparently differentiated gut as evidenced by expression of gut granules and immunoreactive IFB-2, similar to *elt-7(-); elt-2(-)* double mutants, which exhibit an all-or-none block to differentiation event along the length of the animals (Figure 1B, C; Supplementary Figure 2) (Sommermann et al., 2010). However, the differentiation defect observed in *elt-7(-) end-1(-)* double mutants appears to be somewhat milder from that in *elt-7(-); elt-2(-)*: unlike the latter, we observed defined, albeit sporadic, lumen and brush border, and the undifferentiated patches are more frequently interspersed with differentiated patches (Figure 1B, C; Supplementary Figure 2A) (more below).

Analysis of most alternate and sequential double mutant combinations show similar effects (Fig. 1A): sequential double mutants are invariably much more severely affected than alternate double mutants. Together, our data support a sequentially redundant cascade, comprising a recursive series of feedforward regulatory steps (Figure 1H). It is conceivable that such a feedforward system creates a strongly forward-driven, rapidly deployed switch that

ensures timely and robust cell fate commitment and robust lockdown of endoderm cell fate during embryogenesis.

Computational model predicts phenotypes of sequential and alternative double mutants

We took a complementary approach to testing the sequential feedforward model through a computational strategy. We constructed a mathematical model based on the network topology, in which the interactions between the GATA factors, as well as the additional POP-1-dependent activation of *end-3* and *end-1*, were written as a series of ordinary differential equations and the model parameters (Supplementary File 1) were determined by fitting to published transcriptomics data (Baugh et al., 2003; Tintori et al., 2016) using a custom algorithm that follows an iterative least-squares method. We then carried out *in silico* perturbations of the feedforward circuits (see Methods and Materials and Supplementary Materials) to provide predictions of these effects on the relative timing and levels of *elt-2* activation and therefore the final output of the endoderm GRN. Our computed results revealed that *elt-2* expression is predicted to occur, but with delayed onset in all single mutants and in mutants lacking alternate pairs of GATA factors (Figure 1I; Supplementary File 1). However, as observed with the experimental outcomes, *elt-2* expression is predicted to be completely abrogated in most double mutant combinations in which sequential members of the endoderm cascade are removed (Figure 1I), consistent with their pronounced developmental defects (Figure 1A). The predicted *elt-2* expression level strikingly correlates (Spearman's Rank Correlation $\rho = 0.96$; $p < 0.001$) with the experimentally observed penetrance of the phenotypes of the single and multiple mutants (Figure 1I): single and alternate double mutant combinations, which show no, or weak phenotypes, are predicted to express high levels of *elt-2* early, while sequential double mutants, which show strong defects in gut development and inviability, are predicted not to express *elt-2* at substantial levels at its normal time of onset. These findings, based on modeling with gene

expression data, bolster the proposed recursive feedforward structure of the GATA factor cascade.

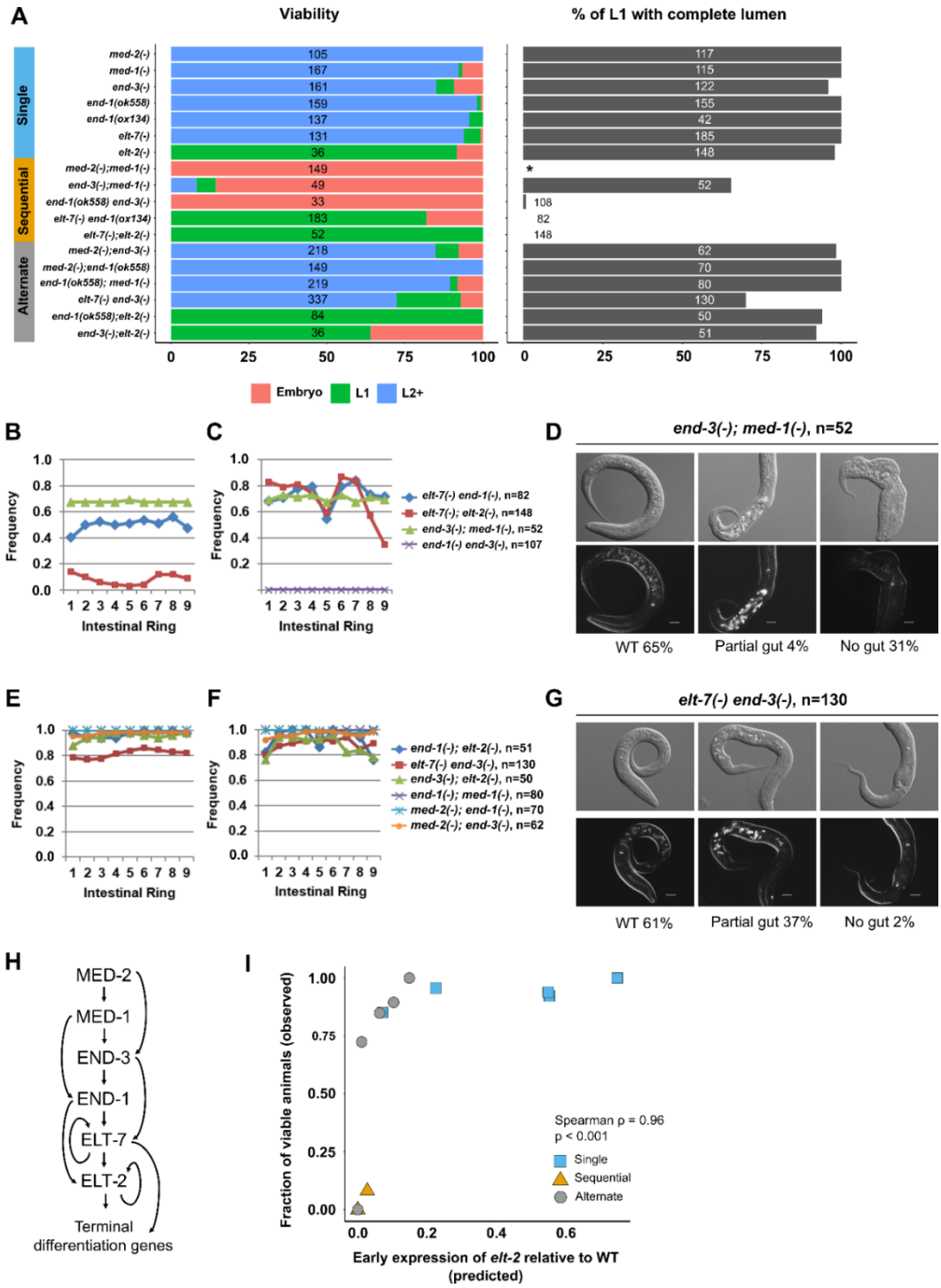


Figure 1: Evidence for recursive feedforward loops in the endoderm GRN.

(A) Comparison of viability and gut lumen morphology in mutant combinations lacking one or two GATA factors in the endoderm regulatory cascade. Mutants missing sequential members of the endoderm cascade exhibit more penetrant developmental defects than mutants with alternate steps disrupted. We note the two loss-of-function alleles of *end-1*, *ox134* (14822 bp deletion; also removes the adjacent *ric-7* gene) and *ok558* (879 bp deletion), do not result in significant loss in viability (Fisher's exact test; p = 0.31) or other discernible phenotypes. We used *ox134* in subsequent analyses unless stated otherwise (see Supplementary Table 1). The total number of animals scored for each genotype is indicated. No L1 animals were scored for *med-2(-); med-1(-)* (indicated by *) as all double mutants arrest as embryos owing to misspecification of MS (and E in majority but not all of the embryos). The frequencies of (B, E) visible lumen and (C, F) gut granules are dramatically reduced in double mutants missing (B, C) sequential GATA pairs compared to those lacking (E, F) alternate members of the endoderm cascade. (D, G) *end-3(-); med-1(-)* shows a more severe defect in intestinal differentiation than *elt-7(-) end-3(-)*. 31% of *end-3(-); med-1(-)* and 2% of *elt-7(-) end-3(-)* show no overt signs of endoderm differentiation. (G) Many *elt-7(-) end-3(-)* larvae contain a partial gut as observed by DIC microscopy and birefringent gut granules, presumably due to sub-threshold levels of END-1-dependent activation of *elt-2*. Scale bar = 10 µm. (H) The sequential redundancy model of the endoderm GRN. (I) Computed *elt-2* expression levels predict the phenotype severity of the mutant combinations. All *med-2(-); med-1(-)* mutants show severe embryonic lethality despite *elt-2* activation owing to a fully penetrant MS→C misspecification. We excluded *med-2(-); med-1(-)* from this analysis.

Variation in temporal expression explains distinct functions for MED-1 and -2

It has been shown previously that *end-3* expression is activated slightly earlier than *end-1* (Supplementary Figure 4) (Maduro et al., 2007; Tintori et al., 2016), as is also consistent with our genetic analyses (Figure 1). Additionally, the END paralogues have diverged considerably with variable DNA-binding domains that perform overlapping but distinct functions (Boeck et al., 2011). Unlike the ENDs, MED-1 and -2 are 98% identical (Maduro et al., 2001). Nonetheless, our genetic evidence demonstrates distinguishable contributions of the two nearly identical paralogues, with MED-2 functions preceding MED-1 (Figure 1; Supplementary Figure 2, 3). By examining published lineage-resolved single-cell RNA-seq data (Tintori et al., 2016), we observed that *med-2* transcripts are undetectable by the 8-cell stage, while *med-1* expression persists briefly in the endoderm precursors (Supplementary Figure 4A). The differential temporal expression of the MEDs was recapitulated using protein-fusion reporters: unlike MED-1, which is expressed in 16E embryos, MED-2 protein expression is largely diminished by the 8E embryonic stage (Supplementary Figure 4B). Together, our data suggest *med-1* and *-2* genes are differentially regulated and MED-2 functions upstream of MED-1, further supporting the feedforward structure at the beginning of the endoderm regulatory cascade.

Synergistic requirements and cross-regulatory interactions of END-1, ELT-7, and ELT-2

As described above, while both *elt-7(-)* and *elt-2(-)* single mutants contain a fully differentiated gut with a contiguous lumen from the pharynx to the rectum surrounded by cells of normal differentiated morphology, *elt-7(-); elt-2(-)* double mutants invariably lack both a defined gut lumen as well as some gut cells, and show a sporadic, all-or-none, block to gut differentiation along the length of the animals (Figure 1A-C; Figure 2A; Supplementary Figure 2, 5) (Sommermann et al., 2010). Although differentiation is highly defective in the absence of ELT-2 and -7, patches of well-differentiated gut are nonetheless evident. Moreover, many terminal differentiation genes remain activated in the absence of ELT-2 and -7 (Dineen et al.,

2018). For example, eliminating the functions of ELT-2 or -7 has little effect on the expression of *act-5*, a gene encoding an actin isoform required for microvilli formation (Dineen et al., 2018; MacQueen et al., 2005). These observations suggest that at least one additional factor, in addition to the ELTs, may activate gut differentiation. One such candidate is END-1, which acts in specification of E cell identity immediately upstream of the *elt* genes. Indeed, although *end-1(-)* and *elt-7(-)* mutants are both phenotypically silent, we found that the *elt-7(-) end-1(-)* double mutant shows extensive gut differentiation defects, with sporadic expression of rhabditin granules (Figure 1A-C; Figure 2B) and *ifb-2* (Figure 2C, D; Supplementary Figure 2), as well as reduced number of differentiated gut cells (Figure 2E; Supplementary Figure 6), reminiscent of *elt-7(-); elt-2(-)* double mutant animals.

We further tested whether END-1 could account for the residual gut-promoting activity by removing it from animals also lacking ELT-2/7. Strikingly, simultaneously eliminating END-1, ELT-7, and ELT-2 results in an apparent elimination of intestinal differentiation (Figure 2F-I). We found that 19.6% of *elt-7(-) end-1(-); elt-2(-)* mutants undergo embryonic arrest, and the remainder die as L1 larvae (n = 143). The triple mutant animals exhibit no morphological signs of gut differentiation, showing no gut granules (Figure 2F, G, H), no detectable intestinal brush border or lumen (Figure 2F, G, I), and no differentiated gut nuclei expressing a gut-specific peptidase transporter, OPT-2 (Figure 2J; Supplementary Figure 7), suggesting a near total block to gut differentiation. We found that knocking out *end-1* strongly reduces the expression level of *act-5* (Figure 2K, L), and *act-5* expression is further downregulated in *elt-7(-) end-1(-)* and *elt-7(-) end-1(-); elt-2(RNAi)* animals (Figure 2K; Supplementary Figure 8), suggesting that END-1, ELT-7 and ELT-2 act collaboratively and redundantly to mediate *act-5* expression, and that END-1 may compensate for the loss of ELT-2 and -7 inputs.

A challenge to the notion of a role for END-1 as the gut-promoting factor acting in the absence of the ELTs is that its expression in wildtype embryos is transient, such that its product

is largely undetectable by the 16E embryonic stage (Li et al., 2019; Zhu et al., 1997); yet gut differentiation seen in the differentiated patches in the *elt-7(-); elt-2(-)* double mutant appears strong and robust late in development. These findings led us to hypothesize that ELT-2 and/or ELT-7 may normally repress *end-1* transcription through negative feedback and that in the absence of the ELTs, *end-1* may be upregulated and drive differentiation. Indeed, we found that the intensity of an *end-1* endogenous protein fusion reporter is modestly elevated in 8E embryos when *elt-2* is knocked down by RNAi (Figure 3A). Moreover, the expression level of *end-1* is upregulated in both 4E and 8E embryos when ELT-7 and -2 are simultaneously depleted (Figure 3B). While these findings provide evidence for the hypothesized feedback inhibition, the effect is rather weak, as we did not observe obvious longer-term perdurance of *end-1* expression (see Discussion).

Taken together, our results suggest END-1 is poised in the cascade at the interface between specification and differentiation, placing it at the crux of this key transition. END-1, acting with END-3, regulates *specification* of the E lineage, whereas END-1, acting with ELT-7 and ELT-2, controls *differentiation* of the intestine.

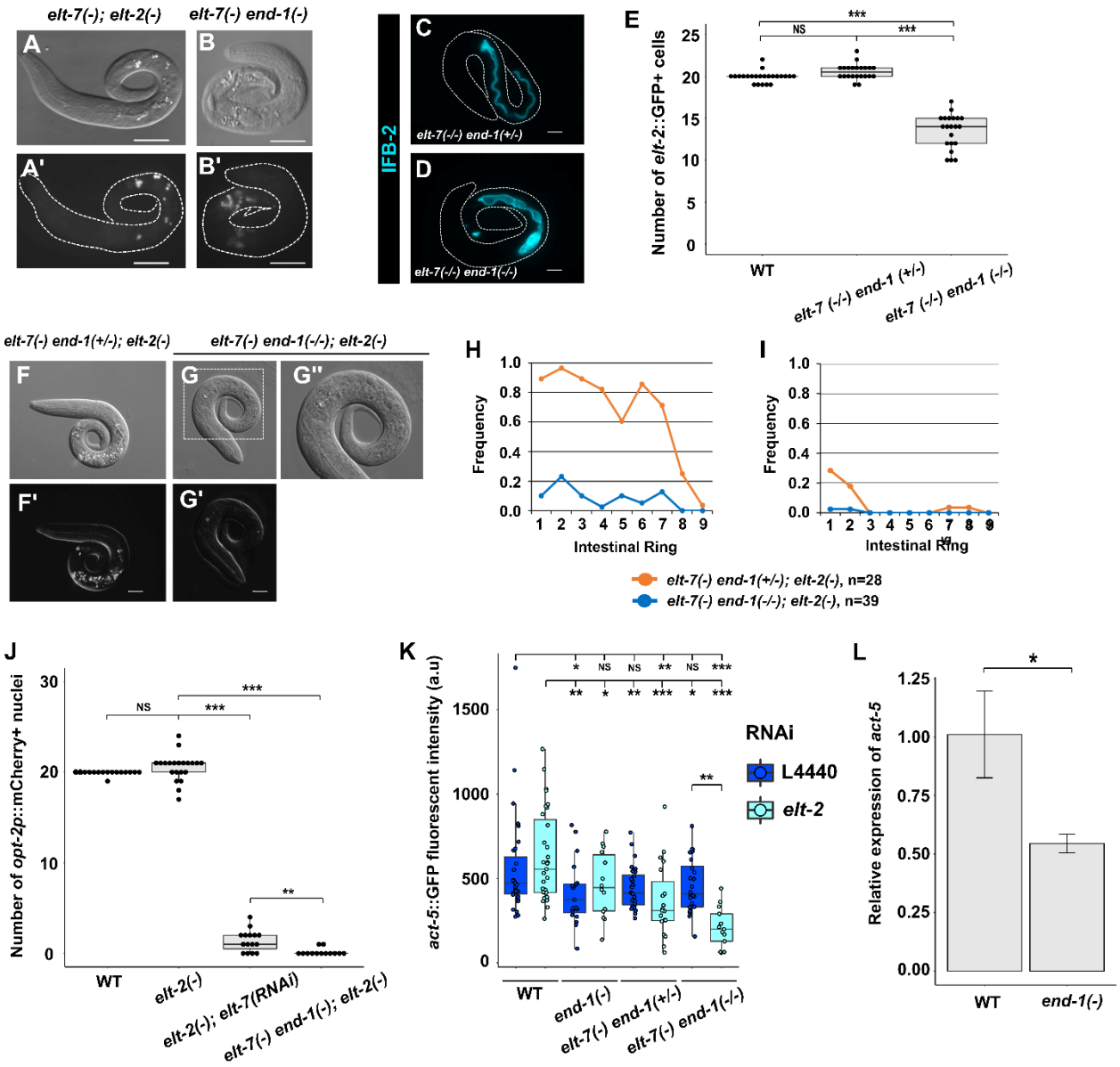


Figure 2: Synergistic actions of END-1, ELT-7 and ELT-2 mediate morphological differentiation of endoderm.

(A, B) *elt-7(-) end-1(-)* contains a grossly defective gut with sporadic patches of rhabditin granules interspersed with apparently undifferentiated regions, similar to that observed in *elt-7(-); elt-2(-)*. Scale bars = 20 µm. (C, D) Balanced *elt-7(-/-) end-1(-/+)* larva shows uniform expression of IFB-2 (*kcIs6* transgene) along the length of the animal, while *elt-7(-/-) end-1(-/-)* shows sporadic expression of the transgene. Scale bars = 10 µm. (E) Wildtype and *elt-7(-/-) end-1(-/+)* animals contain an average of 20 intestinal cells, marked by *elt-2*::GFP reporter. The number of differentiated gut cells is markedly reduced in *elt-7(-/-) end-1(-/-)* larvae (mean = 13.5

cells). (F, G) Representative micrographs of *elt-7(-) end-1(+/-); elt-2(-)* and *elt-7(-) end-1(-/-); elt-2(-)* triple mutants. (F, F') *elt-7(-) end-1(+/-); elt-2(-)* lacks a visible lumen and contains sporadic birefringent granules. (G, G') *elt-7(-) end-1(-/-); elt-2(-)* exhibits absolutely no signs of endoderm differentiation. (G") Magnified view of the animal present in (G). Scale bars = 10 µm. (H, I) The average frequencies of (H) gut granule and (I) lumen are dramatically reduced in *elt-7(-) end-1(-/-); elt-2(-)* compared to *elt-7(-) end-1(+/-); elt-2(-)* animals. (J) Wildtype and *elt-2(-)* show similar number of differentiated intestinal cells, although the variance in *elt-2(-)* is significantly increased (F-test $p < 0.001$). The number of *opt-2p::mCherry* (*irSi24*) expressing cells is strongly reduced in *elt-2(-); elt-7(RNAi)*. No gut cells were detected in vast majority of *elt-7(-) end-1(-); elt-2(-)* triple mutants. (K) The expression of *act-5::GFP* translational reporter *(jyIs13*) in various mutant combinations. END-1, ELT-7, and ELT-2 appear to act together to regulate *act-5* expression in the intestine. (L) *act-5* is downregulated in *end-1(-)* compared to wildtype as detected by RT-qPCR. *act-1* was used as the internal reference. Three replicates were performed for each genotype. Error bars represent standard deviation. * $p \leq 0.05$ by two-tail t-test. For panels E and J, NS $p > 0.05$, ** $p \leq 0.01$, *** $p \leq 0.001$ by non-parametric Kruskal-Wallis test and pairwise Wilcoxon tests with Benjamini & Hochberg correction. For panel K, NS $p > 0.05$, * $p \leq 0.05$, ** $p \leq 0.01$, *** $p \leq 0.001$ by parametric one-way ANOVA followed by pairwise t-tests with Benjamini & Hochberg correction.

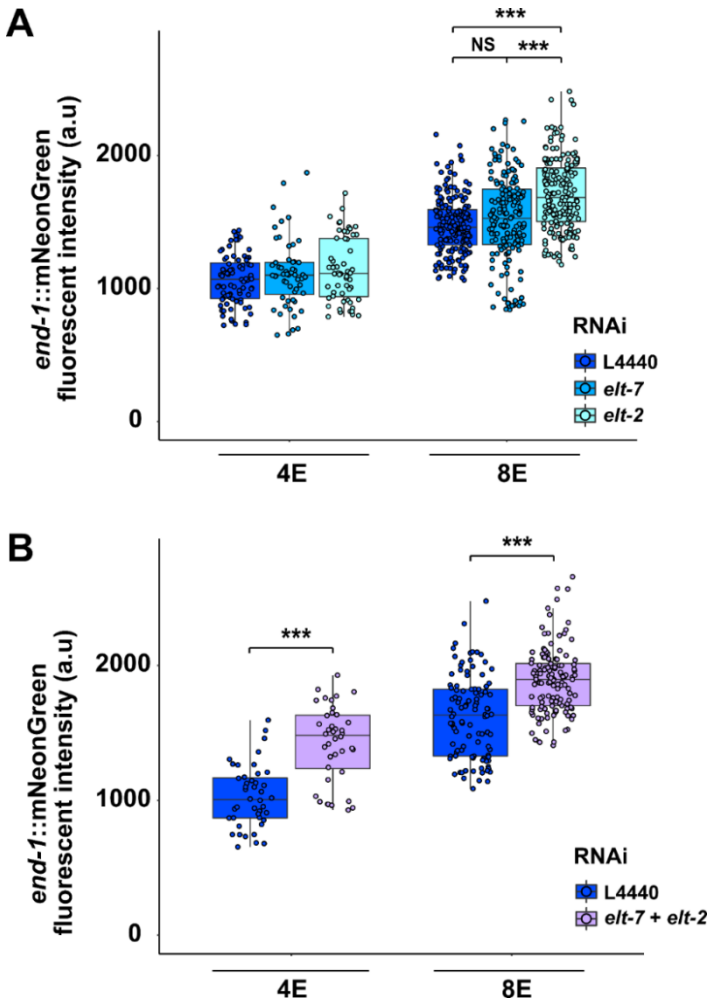


Figure 3: ELT-2 antagonizes *end-1* expression.

(A) Expression of endogenously tagged *end-1* increases in 8E embryos upon *elt-2* RNAi treatment. NS $p > 0.05$, *** $p \leq 0.001$ by Kruskal-Wallis test and pairwise Wilcoxon tests with Benjamini & Hochberg correction. (B) Knocking down both *elt-7* and *elt-2* further elevates END-1 expression in both 4E and 8E embryos. *** $p \leq 0.001$ by Wilcoxon tests.

ELT-2 and ELT-7 collaborate to safeguard intestinal cell fate

When improperly specified, E has been shown to undergo wholesale lineage conversion to either a C-like mesectodermal fate (when SKN-1, the MEDs, or END-1/3 are removed), thereby inappropriately generating epidermis, or an MS-like mesodermal fate (for example, in the absence of Wnt signaling), thereby inappropriately giving rise to pharyngeal tissue (details described in Figure 4A). Given that END-1 straddles the transition from specification of the endoderm progenitor and differentiation of the gut, we sought to investigate whether specification and differentiation involve distinct regulatory events by examining non-endodermal (epidermal and pharyngeal) gene activity when differentiation is impaired after this transition is thought to occur. In *end-1(-) end-3(-)* double mutants, misspecification of the E cell and its conversion to a C-like mesectodermal progenitor, results in severe morphological defects as a result of the supernumerary epidermal cells and consequent deformation of the epidermis (Figure 4B; Supplementary Figure 9) (Maduro et al., 2005a). In contrast, the *elt-7(-) end-1(-); elt-2(-)* triple mutant animals that, like the *end-1(-) end-3(-)* double mutants, do not make a discernible gut, are not substantially defective for overall body morphogenesis (Supplementary Figure 9), suggesting that E is initially specified. Consistent with this observation, unlike *end-1(-) end-3(-)* double mutant animals, *elt-7(-) end-1(-); elt-2(-)* triple mutant larvae contain a wildtype number of epidermal cells, implying that E→C misspecification does not occur (Figure 4B). These results imply that END-3 alone is sufficient to promote endoderm specification over C-like mesectodermal cell fate. However, we were surprised to observe mis-expression of the pharyngeal muscle-specific myosin gene, *myo-2*, in the gut region of many *elt-7(-); elt-2(-)* and *elt-7(-) end-1(-); elt-2(-)* mutants (Figure 4C, D). Additionally, knocking down *elt-7* in *end-1(-); elt-2(-)* results in ectopic expression of *ceh-22*, which encodes a pharynx-specific NK-2-type homeodomain protein, in the otherwise undifferentiated gut (Supplementary Figure 10). As the Wnt/POP-1-dependent polarization of EMS is unperturbed in the mutants, we reasoned that

inappropriate expression of pharyngeal genes is unlikely to be the result of wholesale E→MS transformation, but may reflect later errors in the fidelity of differentiation.

PHA-4/FoxA is the organ selector gene that specifies pharyngeal identity, regulating *myo-2* and *ceh-22* among thousands of other targets in the pharynx (Mango et al., 1994; Zhong et al., 2010). We found that the ectopic expression of *myo-2* in *elt-7(-); elt-2(-)* is at least partially suppressed in *pha-4(RNAi)* animals (Figure 4E). In wildtype animals, PHA-4 is expressed at high levels in the pharynx and rectum and at low levels in the intestine (Figure 4F, F') (Horner et al., 1998; Kalb et al., 1998). It was previously found that ELT-2 positively regulates *pha-4* as forced expression of *elt-2* causes widespread activation of *pha-4* (Kalb et al., 1998). Paradoxically, we found that *pha-4* is upregulated in the intestine in *elt-2(-)* animals (Figure 4G, G'), and depleting *elt-7* in *elt-2(-)* animals further enhances this effect (Figure 4H, H'). These findings raise the possibility that ELT-2 serves dual roles as both an activator and a repressor depending on its expression level. Thus, it appears that upregulation of *pha-4* in the midgut in the absence of ELT-2 and -7 activates sporadic ectopic pharyngeal gene activity. Supporting our model, PHA-4 target genes have been shown to be regulated in part by PHA-4 binding affinity and occupancy (Fakhouri et al., 2010; Gaudet et al., 2004). Taken together, our results show that the boundaries of regulator state domains along the digestive tract of *C. elegans* are established, at least partly, by transcriptional repression mediated by ELT-2 and -7 in the intestine (Figure 4I).

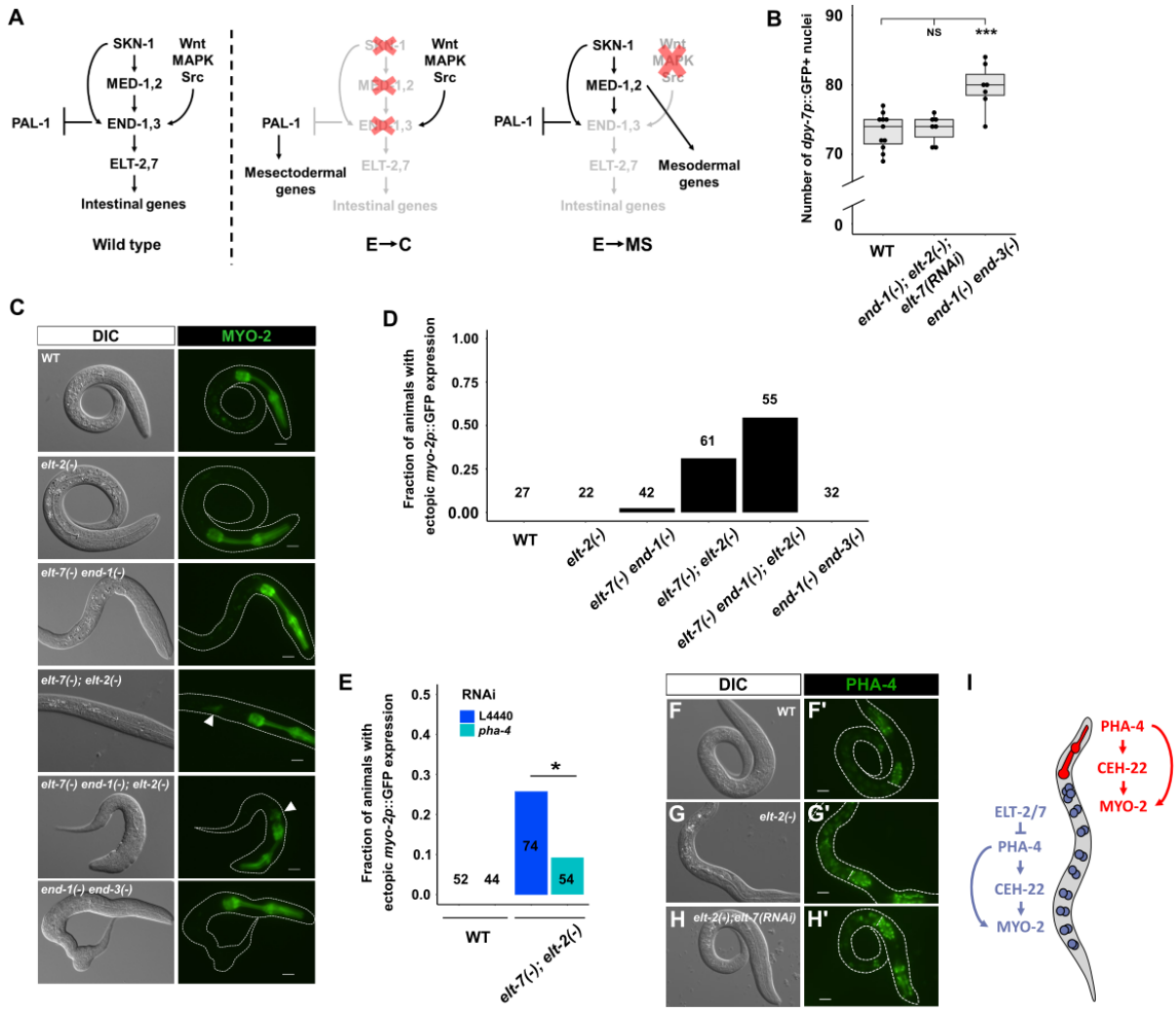


Figure 4: ELT-2 and ELT-7 repress pharyngeal fate in the intestine.

(A) E/C or E/MS binary fate choice during early embryogenesis. The Caudal homolog PAL-1 is required for the specification of mesectodermal C blastomere, which gives rise to epidermis and body wall muscles (Hunter and Kenyon, 1996). Maternally provided *pal-1* is specifically translated in the EMS and P_2 cells. In MS and E, PAL-1 activity is blocked by TBX-35 and END-1/3, respectively (Broitman-Maduro et al., 2006). Thus, depleting TBX-35, END-1/3, or their upstream activators, MED-1/2 and SKN-1, causes excess skin and muscle owing to the misspecification of MS and/or E as C, the somatic descendant of P_2 (Hunter and Kenyon, 1996). As EMS divides, Wnt, MAPK, and Src signaling from P_2 polarizes EMS, leading to the phosphorylation and nucleocytoplasmic redistribution of POP-1, and activation of *end-1/3* in E, but not MS. When the polarizing signal from P_2 is disrupted, POP-1 is unphosphorylated, and MED-1 and -2, instead, activate the development of mesoderm (MS), which gives rise to the posterior pharynx and body wall muscles (Maduro and Rothman, 2002; Maduro et al., 2002;

Rocheleau et al., 1999; Shin et al., 1999). (B) Wildtype and *end-1(-); elt-2(-); elt-7(RNAi)* larvae contain ~73 epidermal cells, while *end-1(-) end-3(-)* larvae contain ~80 epidermal cells marked by *dpy-7p::GFP* expression. NS $p > 0.05$, *** $p \leq 0.001$ by parametric one-way ANOVA followed by pairwise t-tests with Benjamini & Hochberg correction. (C) Representative DIC and fluorescent micrographs showing different mutation combinations expressing *myo-2p::GFP*. Ectopic expression of *myo-2* is evident in *elt-7(-); elt-2(-)* and *elt-7(-) end-1(-); elt-2(-)* mutants (arrowheads). (D) The frequency of animals showing mis-expression of *myo-2* as shown in (C). Number of animals scored for each genotype is indicated. (E) Knocking down *pha-4* partially rescues ectopic expression of *myo-2* in *elt-7(-); elt-2(-)* animals. * $p \leq 0.05$ by Fisher's exact test. (F-H) The expression of endogenously tagged *pha-4* reporter in (F, F') wildtype, (G, G') *elt-2(-)*, and (H, H') *elt-2(-); elt-7(RNAi)* worms. The white horizontal lines in (F'-H') mark the posterior end of the pharynx. Exposure time = 195 ms. All scale bars = 10 μm. (I) Model of spatial repression and fate exclusion in the digestive tract. ELT-2 and -7 repress *pha-4* and prevent the expression pharyngeal genes in the midgut.

END-1 and ELT-7 establish the boundary between the valve and intestinal tubes

The foregoing results suggest that the core regulators involved in gut differentiation (END-1, ELT-2, and ELT-7) regulate faithful differentiation of cells in the digestive tract. The pharynx and the intestine are linked by the pharyngeal-intestinal valve (vpi), which consists of six cells arranged into three rings (Rasmussen et al., 2013) (Figure 5A). In wildtype worms, *ajm-1*::*GFP* (*jcIs1* transgene) is strongly expressed through the adherens junctions lining the lumen of the pharynx and vpi, and the expression drops off sharply to low levels starting at the anteriormost ring of the intestine, and continuing throughout the entire length of the animal (Köppen et al., 2001; Sommermann et al., 2010). However, while *end-1(-)* and *elt-7(-/-) end-1(+/-)* animals show wildtype *ajm-1* expression pattern, *ajm-1* signal is markedly elevated in the anterior intestinal terminus of *elt-7(-/-) end-1(-/-)* animals (Figure 5B-E). Additionally, we observed ectopic expression of two vpi markers, *cdf (cation diffusion facilitator)-1* (Figure 5F-H) and *hum (heavy chain, unconventional myosin)-1* (Figure 5I-K), in the anterior terminus of *end-1(-/-) elt-7(-/-)* larvae.

It has been previously shown that worms lacking *elt-2*, like *elt-7(-/-) end-1(-/-)* animals, exhibit striking caudal extension of the valve cell markers (Sommermann et al., 2010). We found that ELT-2 is expressed at wildtype levels in *elt-7(-/-) end-1(-/-)* larvae, owing to its positive autoregulation (Figure 5L). This suggests that the expansion of vpi gene expression in the intestine that we observed in *elt-7(-/-) end-1(-/-)* animals may be independent of ELT-2 function. It is currently unclear whether the expression of vpi reporters in the intestine reflects *bona fide* transformation of gut cells into valve-like cells or aberrant development of vpi and mispositioning of excess valve cells. Regardless, our results demonstrate the important roles of intestinal GATA factors in the development of a properly patterned digestive tract.

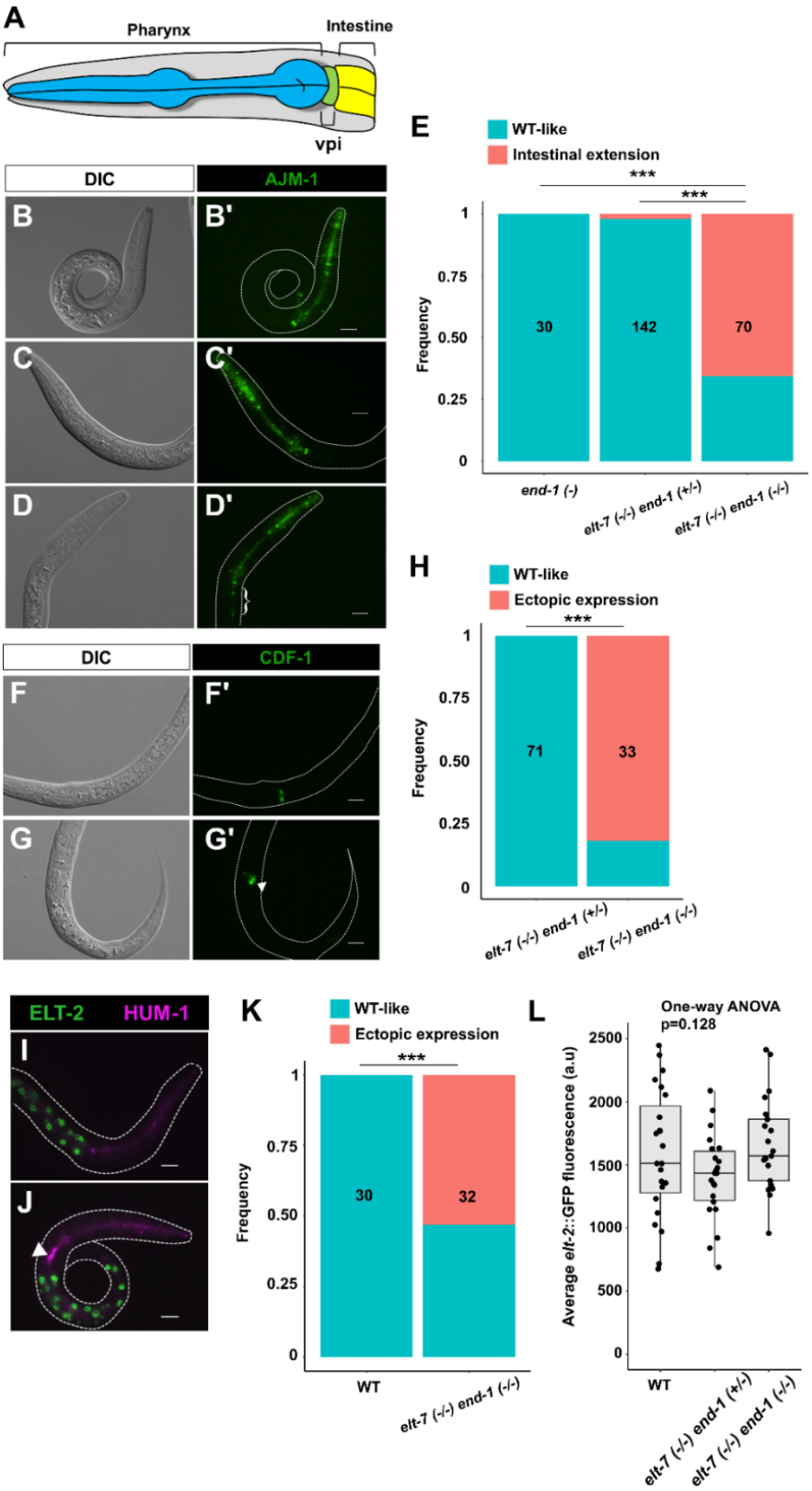


Figure 5: ELT-7 and END-1 function synergistically to repress ectopic expression of valve cell markers in the anterior gut.

(A) Drawing highlighting the anatomy of pharynx, vpi, and intestine. The expression of *jcIs1[ajm-1::GFP]* transgene appears wildtype in (B, B') *end-1(-)* and (C, C') balanced *elt-7(-/-) end-1(+/-)* mutants. (D, D') *elt-7(-) end-1(-)* shows caudal expansion of *jcIs1* expression into the anterior gut (bracket). (E) The frequency of animals exhibiting intense *jcIs1* reporter expression in the anterior gut. (F, F') *cdf-1::GFP* expression is restricted to the vpi in *elt-7(-/-) end-1(+/-)*. (G, G') Ectopic expression of *cdf-1* reporter is observed in *elt-7(-/-) end-1(-/-)* animals (arrowhead). (H) The frequency of animals with ectopic *cdf-1::GFP* expression in the anterior gut. (I) HUM-1 is highly expressed in the vpi in wildtype animal as revealed by an endogenously tagged reporter. (J) Ectopic expression of *hum-1* is observed in *elt-7(-/-) end-1(-/-)* (arrowhead).

The intestinal cells are marked by *elt-2::GFP*. (K) The frequency of animals with ectopic *hum-1::RFP* expression in the anterior gut. (L) The expression of ELT-2 is not altered in *elt-7(-/-) end-1(+/-)* and *elt-7(-/-) end-1(-/-)*, compared to wildtype L1 larvae. All scale bars = 10µm. For panels E, H, K, the number of animals scored is indicated in each graph. *** $p \leq 0.001$ by Fisher's exact test.

Discussion

The development of *C. elegans* endoderm provides a powerful system to study the regulatory logic underlying cell specification and differentiation. In this study, we reported four major findings: (1) the hierarchical organization and feedforward regulation of GATA factors promote robust and rapid lockdown of endodermal cell fate during *C. elegans* early embryogenesis. (2) END-1 participates in both specification and differentiation and mediates a smooth regulatory state transition. (3) ELT-2 and -7 repress expression of *pha-4* in the midgut to establish regulatory state boundary between the pharynx and the intestine. (4) END-1, ELT-7, and ELT-2 repress the characteristics of pharyngeal valve cell fate at the anterior gut terminus, further defining the spatial domains of the foregut and midgut. Our study therefore provides an important insight into the regulatory circuits that direct specification-to-differentiation transition and subsequent restriction and maintenance of differentiation pattern during development.

Architecture of the *C. elegans* Endoderm Regulatory Cascade

This study, together with findings reported in other publications, provided a comprehensive view of the core endoderm regulatory cascade, with six GATA factors acting through reiterated sequential feedforward loops (Figure 6). At the top of the cascade, maternally provided SKN-1 binds to and turns on the *meds* and the *ends* (Maduro et al., 2001; Maduro et al., 2005b). Although MED-1 and MED-2 protein sequences are nearly identical, we found distinguishable contributions between the two paralogs, with MED-2 acting upstream of MED-1. Indeed, embryos lacking MED-1 show a weaker loss-of-gut phenotype than those lacking MED-2, when SKN-1 function is debilitated (Maduro et al., 2007). Moreover, *med-2* is expressed slightly earlier than *med-1* (Maduro et al., 2007; Tintori et al., 2016), suggesting they are differentially regulated as we have observed in this study.

In the E blastomere, SKN-1 and the MEDs collaboratively activate END-3. MED-1/2 and END-3 in turn activate END-1, the next step in the cascade. While the two endoderm-specifying factors, END-1 and -3, perform largely overlapping functions, our data suggest that END-3 alone is sufficient to direct endoderm specification and suppress inappropriate mesectodermal development, consistent with the lack of detectable phenotype of in *end-1(-)* single mutants. Interesting, we found that END-1, poised at the interface between specification and differentiation, works synergistically with ELT-2 and -7 to activate the endoderm differentiation program (Figure 6). Supporting our model, in vitro gel-shift assays demonstrate the binding of END-1, ELT-7, and ELT-2 to TGATAA sites, which are highly enriched in the promotors of intestinal genes (Du et al., 2016; McGhee et al., 2009; Wiesenfahrt et al., 2016). Remarkably, END-1 is able to initiate endoderm differentiation in *Xenopus* embryos, demonstrating its role as a potent organ selector (Shoichet et al., 2000). Hence, it appears that specification and differentiation involve a *bona fide* handoff of regulatory events by END-1.

Regulatory Logic of Developmental GRN

Our results support the conclusion that the *C. elegans* endoderm GRN comprises a recursive series of feedforward steps, culminating with rapid terminal differentiation (Figure 6). Each GATA factor in the entire cascade receives redundant activating inputs acting through an "OR" logic gate. Consequently, any single mutation in the regulatory cascade is largely phenotypically silent, with the exception of *elt-2(-)*; however, even *elt-2(-)* animals contain what appears to be a well-differentiated, albeit dysfunctional, intestine. Coherent feedforward motifs of the type we observe reiteratively in the endoderm GRN are ubiquitous in developmental GRNs. Such a network configuration appears to result in a rapid response to an activating signal and a delayed response when the inputs are removed (termed sign-sensitive delay), thereby prolonging the effect of a transient activator (Mangan and Alon, 2003). Additionally, such feedforward loops are effective at buffering the system against stochastic noise, thereby

ensuring developmental robustness (Chepyala et al., 2016; Gui et al., 2016; Maduro, 2015). This design principal appears to be crucial to ensure timely and robust activation of *elt-7* and *-2*, as we and others have shown (Boeck et al., 2011; Dineen et al., 2018; Maduro et al., 2015). Timely onset of *elt-2* appears to be critical, as its delay in early embryos has been shown to cause sustained metabolic defects in larvae despite later reattainment of wildtype ELT-2 levels (Maduro et al., 2015).

Our results suggest that END-1 mediates an all-or-none switch in the differentiation program in the absence of ELT-2 and -7. END-1 straddles both specification and differentiation, being buttressed by END-3 upstream and ELT-2/7 downstream. The only difference in these END-1 functions appears to be its timing of action and partnership with another regulatory factor. Regulatory nodes in early specification can indeed directly control morphogenetic events in various contexts (Davidson, 2010; Zhu and Rosenfeld, 2004). Interestingly, although END-1 expression decays shortly after gastrulation, we observed a persistent reduction in *act-5* expression in postembryonic larvae lacking *end-1*. One possible explanation for this observation, and the finding that patches of well-differentiated gut arise in *elt-7(-); elt-2(-)* double mutants long after END-1 becomes undetectable, might be that END-1 targets an unidentified factor (X; Figure 6) that directs differentiation in at least a subset of endodermal progenitors. Like ELT-2 and -7, expression of the putative X factor would be expected to be maintained through a positive autoregulatory loop, thereby continuing to promote gut differentiation after the expression of END-1 has subsided (Figure 6) (Sommermann et al., 2010). Alternatively, END-1 may directly activate differentiation gene batteries in early embryos, and that regulatory state might be sustained through the propagation of epigenetic memory. This priming mechanism has recently been demonstrated in the specification of the ASE sensory neurons in *C. elegans* (Charest et al., 2020). Two transiently expressed T-box factors, TBX-37 and -38, lock their target, *lsy-6*, in a transcriptionally active state during early embryogenesis, priming it for

activation in restricted neuronal lineage (Charest et al., 2020). In mammals, Pax-7 (paired-box-7) initiates myogenic specification and differentiation. Interestingly, many enhancers of Pax-7 target genes retain epigenetic signatures and remain active even in the absence of the initial activator (Zhang et al., 2020).

We found that knocking down *elt-2* causes a slight but significant increase in *end-1* expression in early embryos; however, we did not observe obvious perdurance of END-1 when ELT-2 and ELT-7 are depleted, though we cannot rule out the possibility that the modest negative feedback we observed resulted from incomplete RNAi penetrance. Nevertheless, we propose that this modest feedback inhibition may function to facilitate the transition of regulatory states and ensure that developmental process moves inexorably forward. For example, repression of an early cardiac specification factor, Bmp2, by the Nkx2-5 homeodomain factor is necessary for proper morphological development of the heart in mice (Prall et al., 2007). Such transcriptional repression is also frequently used to install spatial subdivision of regulatory states (Davidson, 2010). As we have shown above, ELT-2, ELT-7 and END-1 may repress alternate cell fates in the midgut and define the boundaries of the digestive tract. It is noteworthy that structurally similar regulatory circuits are repeatedly deployed in biological networks of different contexts while performing similar functions. Thus, the functional output of a GRN depends not only on the specificity of the transcription factors, but also the underlying circuit architecture (Davidson, 2010; Peter, 2020).

Rapid Rewiring of the Endoderm GRN in *Caenorhabditis*

How might a regulatory system of the type described here evolve? Effectors acting on terminal differentiation gene batteries, such as ELT-2 and PHA-4, are widely conserved across the animal kingdom, while the upstream inputs into GRNs appear to be recent innovations that arose during the radiation of the Elegans supergroup within the *Caenorhabditis* genus (Maduro, 2020). The *ends* and *meds* have been proposed to have arisen from the duplication of *elt-2*.

Hence, gene duplication, coupled with *cis*-regulatory changes, may have led to the emergence of new circuitry and rewiring of the endoderm GRN in nematodes through intercalation of these factors in the regulatory network.

Intercalation of the MEDs and ENDs in the cascade may serve to buffer the system against environmental variation and developmental noise by freeing ELT-2 from the direct control of SKN-1, which has been shown to play conserved pleiotropic roles in stress response and lifespan regulation (reviewed in Ewe et al., 2020; Ewe et al., 2021). As we and others have shown, robust induction of *elt-2* is critical to ensure the viability and fitness of the animals (Maduro et al., 2015; Raj et al., 2010). Moreover, the deployment of MEDs and ENDs in the sequential hierarchy may allow canalization of the endodermal lineage by rapidly establishing its regulatory state in the E blastomere (Peter and Davidson, 2011). Consequently, this structure may enable increased developmental speed and early specification of the founder cells in *Caenorhabditis* species.

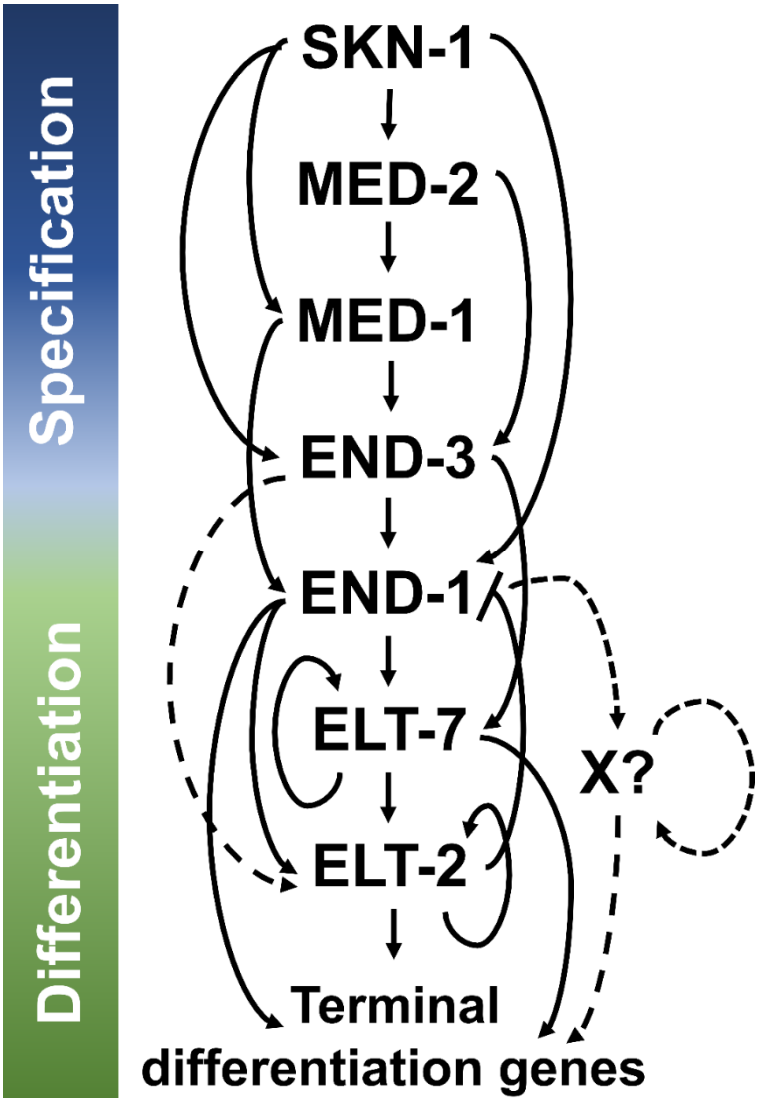


Figure 6: Current model for the *C. elegans* endoderm GRN.

Solid lines indicate known interactions identified biochemically or implied genetically, while dashed lines represent proposed interactions. Maternal factor SKN-1 initiates the endoderm cascade which consists of GATA transcription factors arranged in interlocking feedforward circuits. END-1 and -3 are also regulated by non-GATA transcriptional factors, including SPTF-3 (Sullivan-Brown et al., 2016), PAL-1 (Maduro et al., 2005b), PLP-1 (Witze et al., 2009), and POP-1 (Maduro et al., 2005b), which are omitted from this model for simplicity. The observation of residual gut differentiation in *elt-7(-) end-(-)* larvae suggests that END-3 may directly act on ELT-2 and initiate gut differentiation program in some, but not all, of the endodermal progenitors in the double mutants. END-1 may activate an unidentified transcription factor (X) which

functions to promote gut differentiation after END-1 decays at ~16E embryonic stage (see text for details).

Acknowledgments

We thank members of Rothman lab for helpful advice and feedback, and Cricket Wood for excellent technical support. We thank Dr. Irene Chen (University of California Los Angeles) for her helpful inputs and Dr. Zhuo Du (University of Chinese Academy of Sciences) for providing strain SYS549. We would also like to thank Dr. Rubayn Goh, Dr. Francesca Zappa, Nerea Muniozguren, and Soham Chowdhury for technical assistance. We acknowledge the use of the UCSB NRI-MCDB Microscopy Facility supported by the NIH Grant S10OD010610-01A1, as well as WormBase. Some nematode strains used in this work were provided by the Caenorhabditis Genetics Center (CGC), which is funded by NIH Office of Research Infrastructure Programs Grant P40 OD010440.

Competing Interests

The authors declare no competing or financial interests.

Author contributions

Conceptualization: CKE, EMS, JHR; Resources: CKE, EMS, SEF, MFM; Formal analysis: CKE, EMS, JK; Investigation: CKE, EMS, JK; Writing - original draft: CKE; Writing – review & editing: CKE, EMS, JK, SEF, MFM, JHR; Supervision: JHR; Funding acquisition: JHR

Funding

This work is supported by National Institute of Health ((#1R01HD082347 and # 1R01HD081266) to JHR and NSF (IOS#1258054) to MM.

Data Availability

All data generated or analyzed during this study are included in this article and accompanied supplementary materials. The source code for the GRN computational model is available on https://github.com/RothmanLabCode/endoderm_GRN_model.

References

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

Bei, Y., Hogan, J., Berkowitz, L. A., Soto, M., Rocheleau, C. E., Pang, K. M., Collins, J. and Mello, C. C. (2002). SRC-1 and Wnt signaling act together to specify endoderm and to control cleavage orientation in early C. elegans embryos. *Dev. Cell* **3**, 113–25.

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.

Boveri, T. (1899). *Die Entwickelung von Ascaris megalocephala mit besonderer Rücksicht auf die Kernverhältnisse*. Jena, Gustav Fischer (Opitz Library).

Bowerman, B., Eaton, B. A. and Priess, J. R. (1992). skn-1, a maternally expressed gene required to specify the fate of ventral blastomeres in the early C. elegans embryo. *Cell* **68**, 1061–75.

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

Broitman-Maduro, G., Lin, K. T.-H., Hung, W. W. K. and Maduro, M. F. (2006). Specification of the C. elegans MS blastomere by the T-box factor TBX-35. *Development* **133**, 3097–106.

Charest, J., Daniele, T., Wang, J., Bykov, A., Mandlbauer, A., Asparuhova, M., Röhsner, J., Gutiérrez-Pérez, P. and Cochella, L. (2020). Combinatorial Action of Temporally Segregated Transcription Factors. *Dev. Cell* **55**, 483-499.e7.

Chepyala, S. R., Chen, Y. C., Yan, C. C. S., Lu, C. Y. D., Wu, Y. C. and Hsu, C. P. (2016). Noise propagation with interlinked feed-forward pathways. *Sci. Rep.* **6**, 1–15.

Choi, H., Broitman-Maduro, G. and Maduro, M. F. (2017). Partially compromised specification causes stochastic effects on gut development in C. elegans. *Dev. Biol.* **427**, 49–60.

Clokey, G. V and Jacobson, L. A. (1986). The autofluorescent “lipofuscin granules” in the intestinal cells of Caenorhabditis elegans are secondary lysosomes. *Mech. Ageing Dev.* **35**, 79–94.

Davidson, E. H. (2010). Emerging properties of animal gene regulatory networks. *Nature* **468**,

911–920.

Davidson, E. H. and Levine, M. S. (2008). Properties of developmental gene regulatory networks. *Proc. Natl. Acad. Sci. U. S. A.* **105**, 20063–6.

Davidson, E. H., Rast, J. P., Oliveri, P., Ransick, A., Calestani, C., Yuh, C. H., Minokawa, T., Amore, G., Hinman, V., Arenas-Mena, C., et al. (2002). A genomic regulatory network for development. *Science (80-.).* **295**, 1669–1678.

Dineen, A., Osborne Nishimura, E., Goszczynski, B., Rothman, J. H. and McGhee, J. D. (2018). Quantitating transcription factor redundancy: The relative roles of the ELT-2 and ELT-7 GATA factors in the C. elegans endoderm. *Dev. Biol.* **435**, 150–161.

Du, L., Tracy, S. and Rifkin, S. A. (2016). Mutagenesis of GATA motifs controlling the endoderm regulator elt-2 reveals distinct dominant and secondary cis-regulatory elements. *Dev. Biol.* **412**, 160–170.

Ewe, C. K., Torres Cleuren, Y. N. and Rothman, J. H. (2020). Evolution and developmental system drift in the endoderm gene regulatory network of Caenorhabditis and other nematodes. *Front. Cell Dev. Biol.* **8**, 170.

Ewe, C. K., Alok, G. and Rothman, J. H. (2021). Stressful development: integrating endoderm development, stress, and longevity. *Dev. Biol.* **471**, 34–48.

Fakhouri, T. H. I., Stevenson, J., Chisholm, A. D. and Mango, S. E. (2010). Dynamic Chromatin Organization during Foregut Development Mediated by the Organ Selector Gene PHA-4/FoxA. *PLOS Genet.* **6**, e1001060.

Fay, D. S. (2013). Classical genetic methods. *WormBook* 1–58.

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.

Gaudet, J., Muttumu, S., Horner, M. and Mango, S. E. (2004). Whole-Genome Analysis of Temporal Gene Expression during Foregut Development. *PLOS Biol.* **2**, e352.

Gilleard, J. S. and McGhee, J. D. (2001). Activation of Hypodermal Differentiation in the Caenorhabditis elegans Embryo by GATA Transcription Factors ELT-1 and ELT-3 . *Mol. Cell. Biol.* **21**, 2533–2544.

Gui, R., Liu, Q., Yao, Y., Deng, H., Ma, C., Jia, Y. and Yi, M. (2016). Noise decomposition

principle in a coherent feed-forward transcriptional regulatory loop. *Front. Physiol.* **7**, 600.

Hahn, G. D. (1991). A modified Euler method for dynamic analyses. *Int. J. Numer. Methods Eng.* **32**, 943–955.

Hashimshony, T., Feder, M., Levin, M., Hall, B. K. and Yanai, I. (2015). Spatiotemporal transcriptomics reveals the evolutionary history of the endoderm germ layer. *Nature* **519**, 219–222.

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

Hunter, C. P. and Kenyon, C. (1996). Spatial and temporal controls target pal-1 blastomere-specification activity to a single blastomere lineage in C. elegans embryos. *Cell* **87**, 217–226.

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**, 2171–80.

Kamath, R. S., Fraser, A. G., Dong, Y., Poulin, G., Durbin, R., Gotta, M., Kanapin, A., Le Bot, N., Moreno, S., Sohrmann, M., et al. (2003). Systematic functional analysis of the Caenorhabditis elegans genome using RNAi. *Nature* **421**, 231–237.

Köppen, M., Simske, J. S., Sims, P. A., Firestein, B. L., Hall, D. H., Radice, A. D., Rongo, C. and Hardin, J. D. (2001). Cooperative regulation of AJM-1 controls junctional integrity in Caenorhabditis elegans epithelia. *Nat. Cell Biol.* **3**, 983–991.

Li, X., Zhao, Z., Xu, W., Fan, R., Xiao, L., Ma, X. and Du, Z. (2019). Systems Properties and Spatiotemporal Regulation of Cell Position Variability during Embryogenesis. *Cell Rep.* **26**, 313-321.e7.

Livak, K. J. and Schmittgen, T. D. (2001). Analysis of relative gene expression data using real-time quantitative PCR and the 2-ΔΔCT method. *Methods* **25**, 402–408.

MacQueen, A. J., Baggett, J. J., Perumov, N., Bauer, R. A., Januszewski, T., Schriefer, L. and Waddle, J. A. (2005). ACT-5 is an essential Caenorhabditis elegans actin required for intestinal microvilli formation. *Mol. Biol. Cell* **16**, 3247–3259.

Maduro, M. F. (2015). Developmental robustness in the *Caenorhabditis elegans* embryo. *Mol. Reprod. Dev.* **82**, 918–931.

Maduro, M. F. (2017). Gut development in C. elegans. *Semin. Cell Dev. Biol.* **66**, 3–11.

Maduro, M. F. (2020). Evolutionary dynamics of the Skn-1/Med/end-1,3 regulatory gene cascade in Caenorhabditis endoderm specification. *G3 Genes, Genomes, Genet.* **10**, 333–356.

Maduro, M. F. and Rothman, J. H. (2002). Making Worm Guts: The Gene Regulatory Network of the Caenorhabditis elegans Endoderm. *Dev. Biol.* **246**, 68–85.

Maduro, M. F., Meneghini, M. D., Bowerman, B., Broitman-Maduro, G. and Rothman, J. H. (2001). Restriction of mesendoderm to a single blastomere by the combined action of SKN-1 and a GSK-3beta homolog is mediated by MED-1 and -2 in C. elegans. *Mol. Cell* **7**, 475–85.

Maduro, M. F., Lin, R. and Rothman, J. H. (2002). Dynamics of a Developmental Switch: Recursive Intracellular and Intranuclear Redistribution of Caenorhabditis elegans POP-1 Parallels Wnt-Inhibited Transcriptional Repression. *Dev. Biol.* **248**, 128–142.

Maduro, M. F., Hill, R. J., Heid, P. J., Newman-Smith, E. D., Zhu, J., Priess, J. R. and Rothman, J. H. (2005a). Genetic redundancy in endoderm specification within the genus Caenorhabditis. *Dev. Biol.* **284**, 509–522.

Maduro, M. F., Kasmir, J. J., Zhu, J. and Rothman, J. H. (2005b). The Wnt effector POP-1 and the PAL-1/Caudal homeoprotein collaborate with SKN-1 to activate C. elegans endoderm development. *Dev. Biol.* **285**, 510–523.

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.-Y. and Rifkin, S. A. (2015). MED GATA factors promote robust development of the C. elegans endoderm. *Dev. Biol.* **404**, 66–79.

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

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

Martindale, M. Q., Pang, K. and Finnerty, J. R. (2004). Investigating the origins of triploblasty: "mesodermal" gene expression in a diploblastic animal, the sea anemone Nematostella vectensis (phylum, Cnidaria; class, Anthozoa). *Development* **131**, 2463–74.

McGhee, J. (2007). The C. elegans intestine. *WormBook* 1–36.

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.

Meneghini, M. D., Ishitani, T., Carter, J. C., Hisamoto, N., Ninomiya-Tsuji, J., Thorpe, C. J., Hamill, D. R., Matsumoto, K. and Bowerman, B. (1999). MAP kinase and Wnt pathways converge to downregulate an HMG-domain repressor in Caenorhabditis elegans. *Nature* **399**, 793–797.

Murakami, R., Okumura, T. and Uchiyama, H. (2005). GATA factors as key regulatory molecules in the development of Drosophila endoderm. *Dev. Growth Differ.* **47**, 581–589.

Nakanishi, N., Sogabe, S. and Degnan, B. M. (2014). Evolutionary origin of gastrulation: Insights from sponge development. *BMC Biol.* **12**,.

Okumura, T., Matsumoto, A., Tanimura, T. and Murakami, R. (2005). An endoderm-specific GATA factor gene, dGATAe, is required for the terminal differentiation of the Drosophila endoderm. *Dev. Biol.* **278**, 576–586.

Owraghi, M., Broitman-Maduro, G., Luu, T., Roberson, H. and Maduro, M. F. (2010). Roles of the Wnt effector POP-1/TCF in the C. elegans endomesoderm specification gene network. *Dev. Biol.* **340**, 209–21.

Peter, I. S. (2020). The function of architecture and logic in developmental gene regulatory networks. In *Current Topics in Developmental Biology*, pp. 267–295. Academic Press Inc.

Peter, I. S. and Davidson, E. H. (2011). Evolution of gene regulatory networks controlling body plan development. *Cell* **144**, 970–85.

Prall, O. W. J., Menon, M. K., Solloway, M. J., Watanabe, Y., Zaffran, S., Bajolle, F., Biben,

C., McBride, J. J., Robertson, B. R., Chaulet, H., et al. (2007). An Nkx2-5/Bmp2/Smad1 Negative Feedback Loop Controls Heart Progenitor Specification and Proliferation. *Cell* **128**, 947–959.

Raj, A., Rifkin, S. A., Andersen, E. and van Oudenaarden, A. (2010). Variability in gene expression underlies incomplete penetrance. *Nature* **463**, 913–918.

Rasmussen, J. P., Feldman, J. L., Reddy, S. S. and Priess, J. R. (2013). Cell Interactions and Patterned Intercalations Shape and Link Epithelial Tubes in C. elegans. *PLoS Genet.* **9**, 1003772.

Rocheleau, C. E., Yasuda, J., Shin, T. H., Lin, R., Sawa, H., Okano, H., Priess, J. R., Davis, R. J. and Mello, C. C. (1999). WRM-1 Activates the LIT-1 Protein Kinase to Transduce Anterior/Posterior Polarity Signals in C. elegans. *Cell* **97**, 717–726.

Rodaway, A. and Patient, R. (2001). Mesendoderm: An Ancient Germ Layer? *Cell* **105**, 169–172.

Rual, J.-F., Ceron, J., Koreth, J., Hao, T., Nicot, A.-S., Hirozane-Kishikawa, T., Vandenhaute, J., Orkin, S. H., Hill, D. E., van den Heuvel, S., et al. (2004). Toward Improving Caenorhabditis elegans Phenome Mapping With an ORFeome-Based RNAi Library. *Genome Res.* **14**, 2162–2168.

Ruhe, A. (1979). Accelerated Gauss-Newton algorithms for nonlinear least squares problems. *BIT Numer. Math. 1979 193* **19**, 356–367.

Shin, T. H., Yasuda, J., Rocheleau, C. E., Lin, R., Soto, M., Bei, Y., Davis, R. J. and Mello, C. C. (1999). MOM-4, a MAP kinase kinase kinase-related protein, activates WRM-1/LIT-1 kinase to transduce anterior/posterior polarity signals in C. elegans. *Mol. Cell* **4**, 275–80.

Shoichet, S. A., Malik, T. H., Rothman, J. H. and Shivdasani, R. A. (2000). Action of the Caenorhabditis elegans GATA factor END-1 in Xenopus suggests that similar mechanisms initiate endoderm development in ecdysozoa and vertebrates. *Proc. Natl. Acad. Sci. U. S. A.* **97**, 4076–81.

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.

Sullivan-Brown, J. L., Tandon, P., Bird, K. E., Dickinson, D. J., Tintori, S. C., Heppert, J. K.,

Meserve, J. H., Trogden, K. P., Orlowski, S. K., Conlon, F. L., et al. (2016). Identifying regulators of morphogenesis common to vertebrate neural tube closure and Caenorhabditis elegans gastrulation. *Genetics* **202**, 123–139.

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.

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

Tintori, S. C., Osborne Nishimura, E., Golden, P., Lieb, J. D. and Goldstein, B. (2016). A Transcriptional Lineage of the Early C. elegans Embryo. *Dev. Cell* **38**, 430–444.

Waddington, C. (1957). *The strategy of the genes*. London: Volume George Allen and Unwin.

Wiesenfahrt, T., Berg, J. Y., Osborne Nishimura, E., Robinson, A. G., Goszczynski, B., Lieb, J. D. and McGhee, J. D. (2016). The function and regulation of the GATA factor ELT-2 in the *C. elegans* endoderm. *Development* **143**, 483–491.

Witze, E. S., Field, E. D., Hunt, D. F. and Rothman, J. H. (2009). C. elegans pur alpha, an activator of end-1, synergizes with the Wnt pathway to specify endoderm. *Dev. Biol.* **327**, 12–23.

Yip, K. Y., Alexander, R. P., Yan, K.-K. and Gerstein, M. (2010). Improved Reconstruction of In Silico Gene Regulatory Networks by Integrating Knockout and Perturbation Data. *PLoS One* **5**, e8121.

Yuh, C. H., Dorman, E. R., Howard, M. L. and Davidson, E. H. (2004). An otx cis-regulatory module: a key node in the sea urchin endomesoderm gene regulatory network. *Dev. Biol.* **269**, 536–551.

Zhang, N., Mendieta-Esteban, J., Magli, A., Lilja, K. C., Perlingeiro, R. C. R., Marti-Renom, M. A., Tsirigos, A. and Dynlacht, B. D. (2020). Muscle progenitor specification and myogenic differentiation are associated with changes in chromatin topology. *Nat. Commun.* **11**, 1–18.

Zhong, M., Niu, W., Lu, Z. J., Sarov, M., Murray, J. I., Janette, J., Raha, D., Sheaffer, K. L., Lam, H. Y. K., Preston, E., et al. (2010). Genome-Wide Identification of Binding Sites Defines Distinct Functions for Caenorhabditis elegans PHA-4/FOXA in Development and

Environmental Response. *PLoS Genet.* **6**, e1000848.

Zhu, X. and Rosenfeld, M. G. (2004). Transcriptional control of precursor proliferation in the early phases of pituitary development. *Curr. Opin. Genet. Dev.* **14**, 567–574.

Zhu, J., Hill, R. J., Heid, P. J., Fukuyama, M., Sugimoto, A., Priess, J. R. and Rothman, J. H. (1997). end-1 encodes an apparent GATA factor that specifies the endoderm precursor in Caenorhabditis elegans embryos. *Genes Dev.* **11**, 2883–96.

Supplementary Materials

Description of computational model

The interactions between the GATA factors (Figure 6), as well as POP-1 activation of *end-3* and *end-1*, were written as a system of differential equations. Each factor is expressed according to the concentration of its activators multiplied by respective coefficients a, representing the strength of activation (see Supplementary File 1). The activation is not instant, but delayed by a time interval τ, which accounts for the time required for transcription, translation, and re-localization to the nucleus. SKN-1 and POP-1 activation occur as square waves in EMS and E blastomere, respectively. ELT-7 and ELT-2 self-activate via feedback loops of strength f_1 and f_2, respectively, after surpassing arbitrary concentration thresholds Φ_1 and Φ_2, respectively. Finally, all factors are degraded at the same rate δ. The resulting system of equations is as follows.

(1) $\frac{d[MED2]}{dt} = a_1[SKN1]_t - \delta[MED2]_t$

(2) $\frac{d[MED1]}{dt} = a_2[SKN1]_t + a_5[MED2]_{t-\tau} - \delta[MED1]_t$

(3) $\frac{d[END3]}{dt} = a_3[SKN1]_t + a_9[POP1]_t + a_6[MED2]_{t-\tau} + a_7[MED1]_{t-\tau} - \delta[END3]_t$

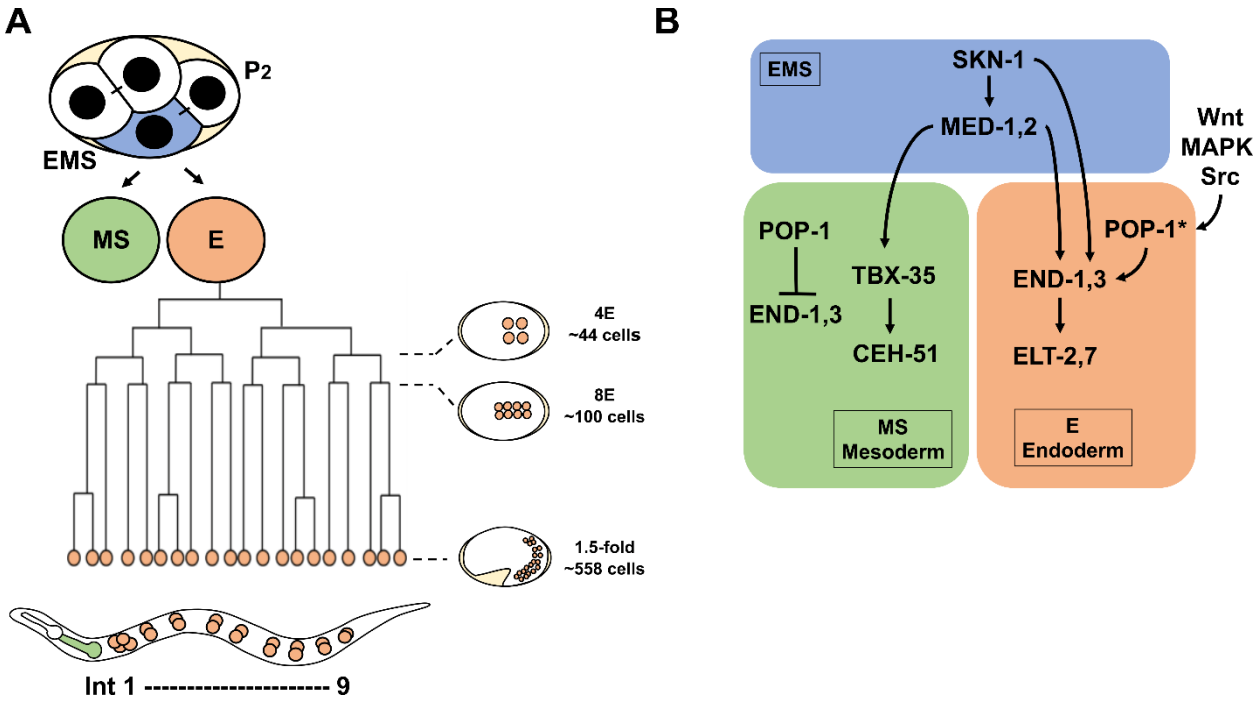
(4) $\frac{d[END1]}{dt} = a_4[SKN1]_t + a_{10}[POP1]_t + a_8[MED1]_{t-\tau} + a_{11}[END3]_{t-\tau} - \delta[END1]_t$

(5) $\frac{d[ELT7]}{dt} = a_{12}[END3]_{t-\tau} + a_{13}[END1]_{t-\tau} + f_1([ELT7]_{t-\tau} > \Phi_1)[ELT7]_{t-\tau} - \delta[ELT7]_t$

(6) $\frac{d[ELT2]}{dt} = a_{14}[END1]_{t-\tau} + a_{15}[ELT7]_{t-\tau} + f_2([ELT2]_{t-\tau} > \Phi_2)[ELT2]_{t-\tau} - \delta[ELT2]_t$

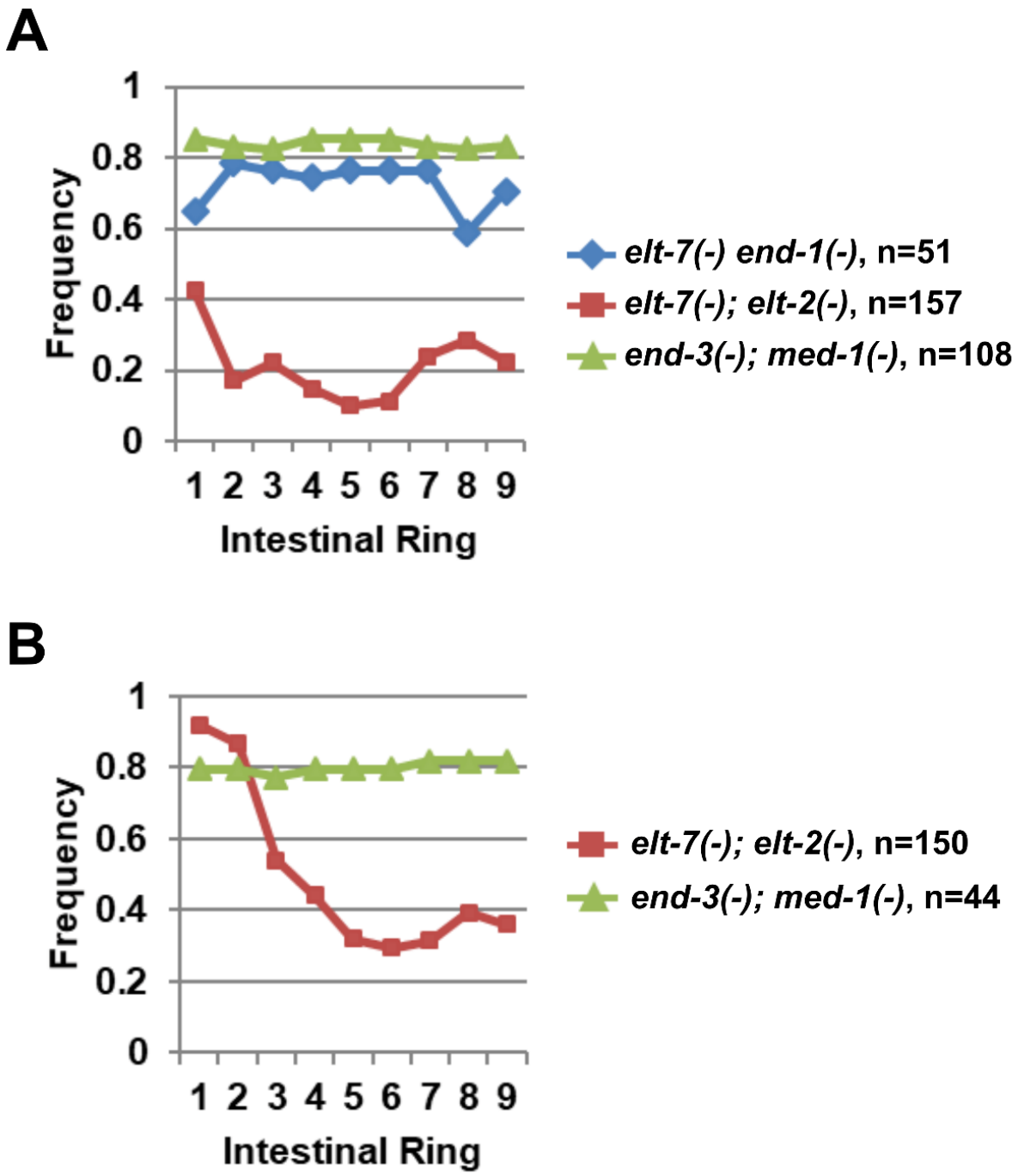
Model parameters (Supplementary File 1) were determined by fitting to transcriptomics data by a custom algorithm written in R following an iterative least-squares method. Parameter guesses were provided, and the model was run using an Euler approximation with 0.01 s time steps. The model fit was measured as the sum of squared differences between model predictions and transcriptomics data. Randomly selected parameter values were randomly changed, and the

model was run again. If the resulting sum of squares was lower, the new parameter values were kept. This process was iterated ~10^6 times. Algorithm-selected parameter values were validated by checking model predictions against the published phenotypes of the single mutants (Boeck et al., 2011; Dineen et al., 2018; Maduro et al., 2015). Two parameters (a_8 and a_{14}) were adjusted manually such that the predicted phenotype of *end-3(-)* matched that reported in published studies (Boeck et al., 2011; Maduro et al., 2005a). The source code is available on https://github.com/RothmanLabCode/endoderm_GRN_model.

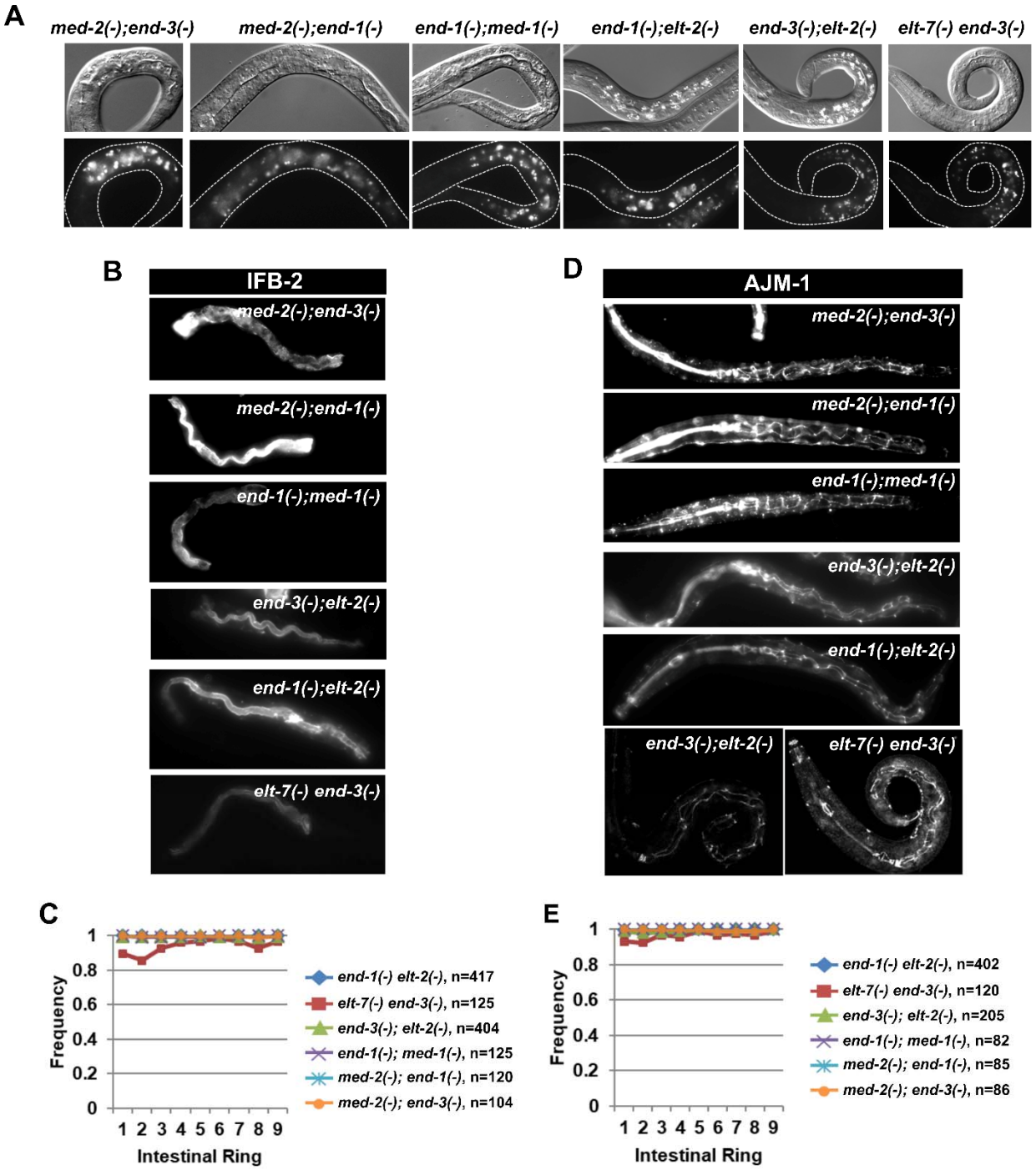


Supplementary Figure 1: *C. elegans* mesendoderm development.

(A) At the four-cell stage, EMS (blue) undergoes asymmetrical division to produce anterior MS (green) and posterior E (orange) blastomeres. Newly hatched L1s contain 20 intestinal cells arranged in nine rings (Ints). Int1 contains four cells, while the remaining eight rings contain two cells each. The MS cell gives rise mesodermal cell types, including the posterior pharynx. (B) Simplified mesendoderm GRN showing three sequential tiers of paired redundant GATA transcription factors (2 redundant MEDs → 2 redundant ENDs → 2 redundant ELTs). Maternally provided SKN-1 activates *med-1* and *-2*, which have both a maternal and zygotic component. In MS, POP-1 represses the expression of *end-1* and *-3* genes. MED-1 and -2 then directly activate the expression of mesoderm-specifying factor, TBX-35. In E, however, Wnt, MAPK and Src signaling from neighboring P2 cell ultimately leads to the phosphorylation of POP-1 (indicated by *), converting it from a repressor to an activator of gut fate. END-1 and -3 subsequently activate the expression of ELT-7 and -2, both of which promote gut morphological differentiation.

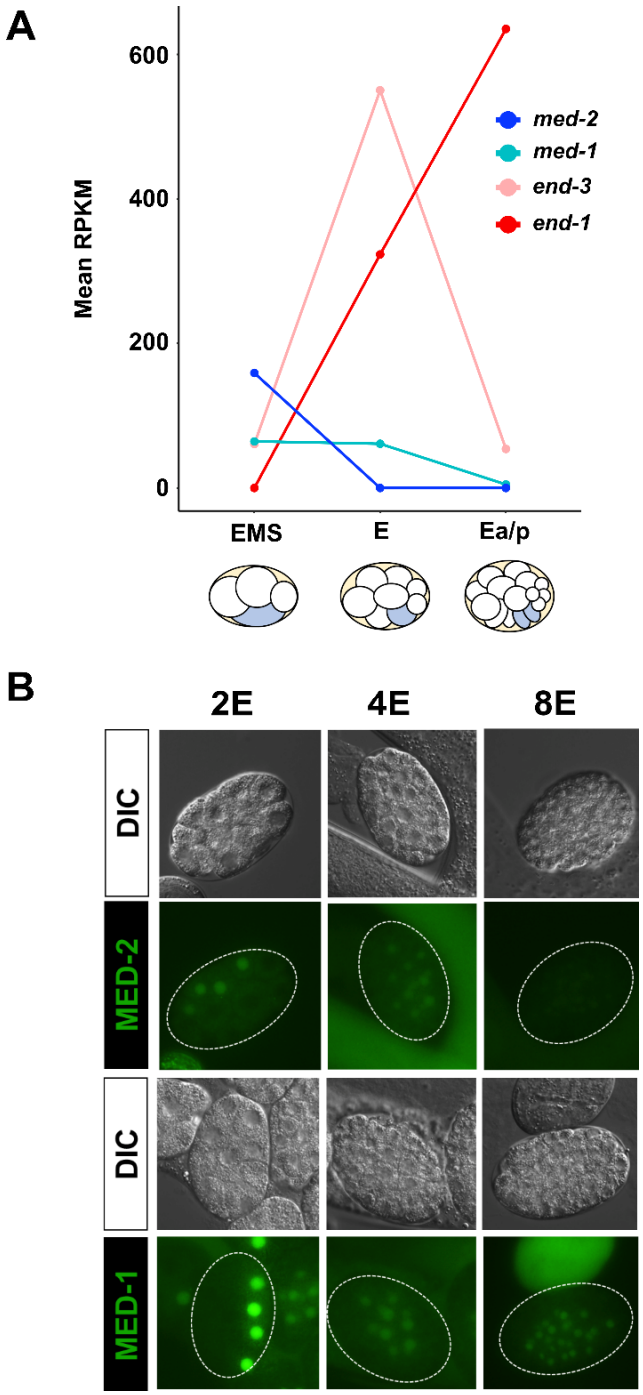


Supplementary Figure 2: Severe gut defects in mutants lacking sequential GATA pairs. Eliminating sequential GATA pairs causes impaired gut differentiation and aberrant expression of immunoreactive (A) IFB-2 and (B) AJM-1.

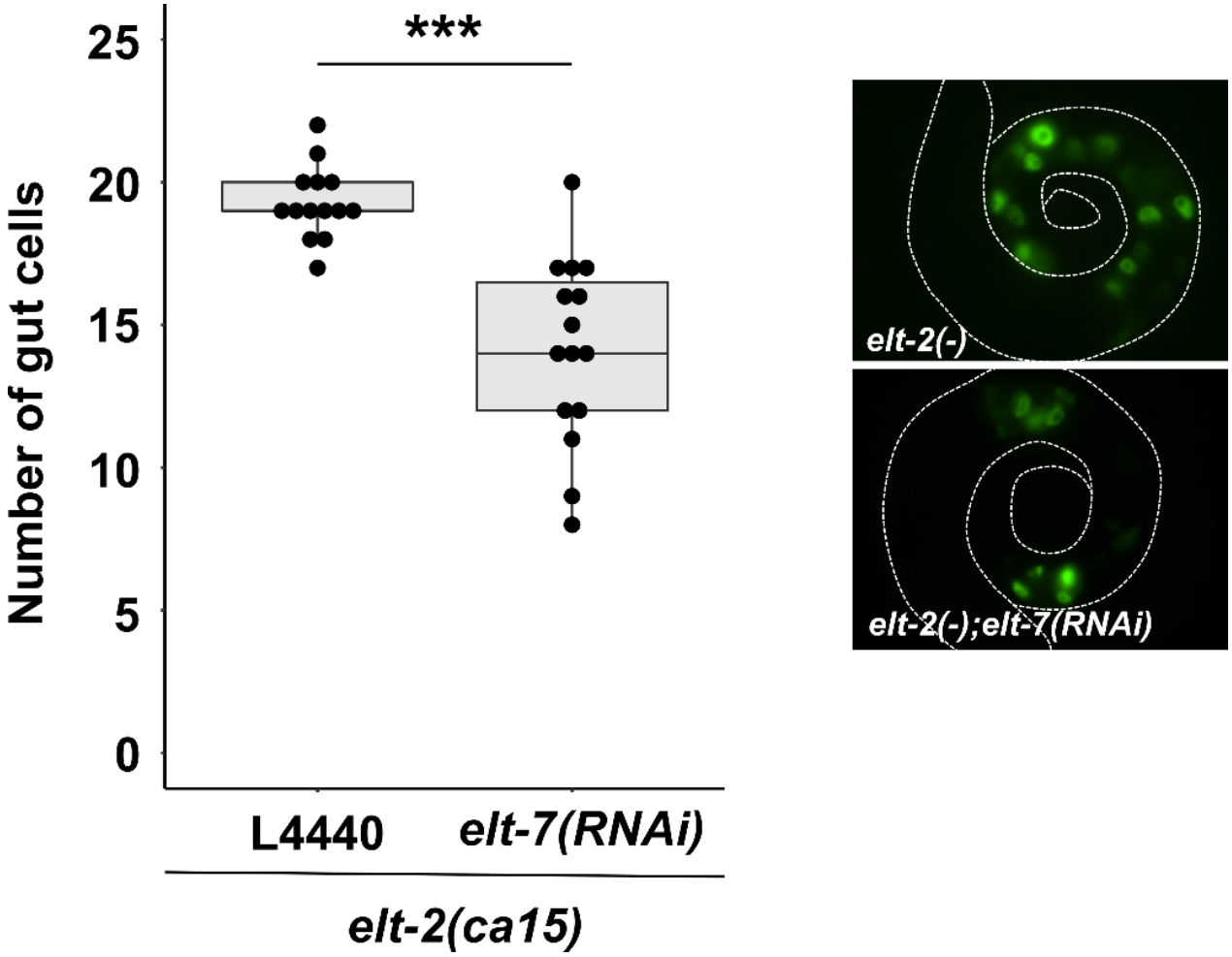


Supplementary Figure 3: Mutants lacking alternate GATA pairs do not show apparent gut defects.

(A) Mutants lacking alternate GATA pairs contain fully differentiated lumen (top row) and gut granules (bottom row) along the length of the animals. The same set of double mutants show wildtype expression of (B, C) immunoreactive IFB-2 and (D, E) AJM-1.

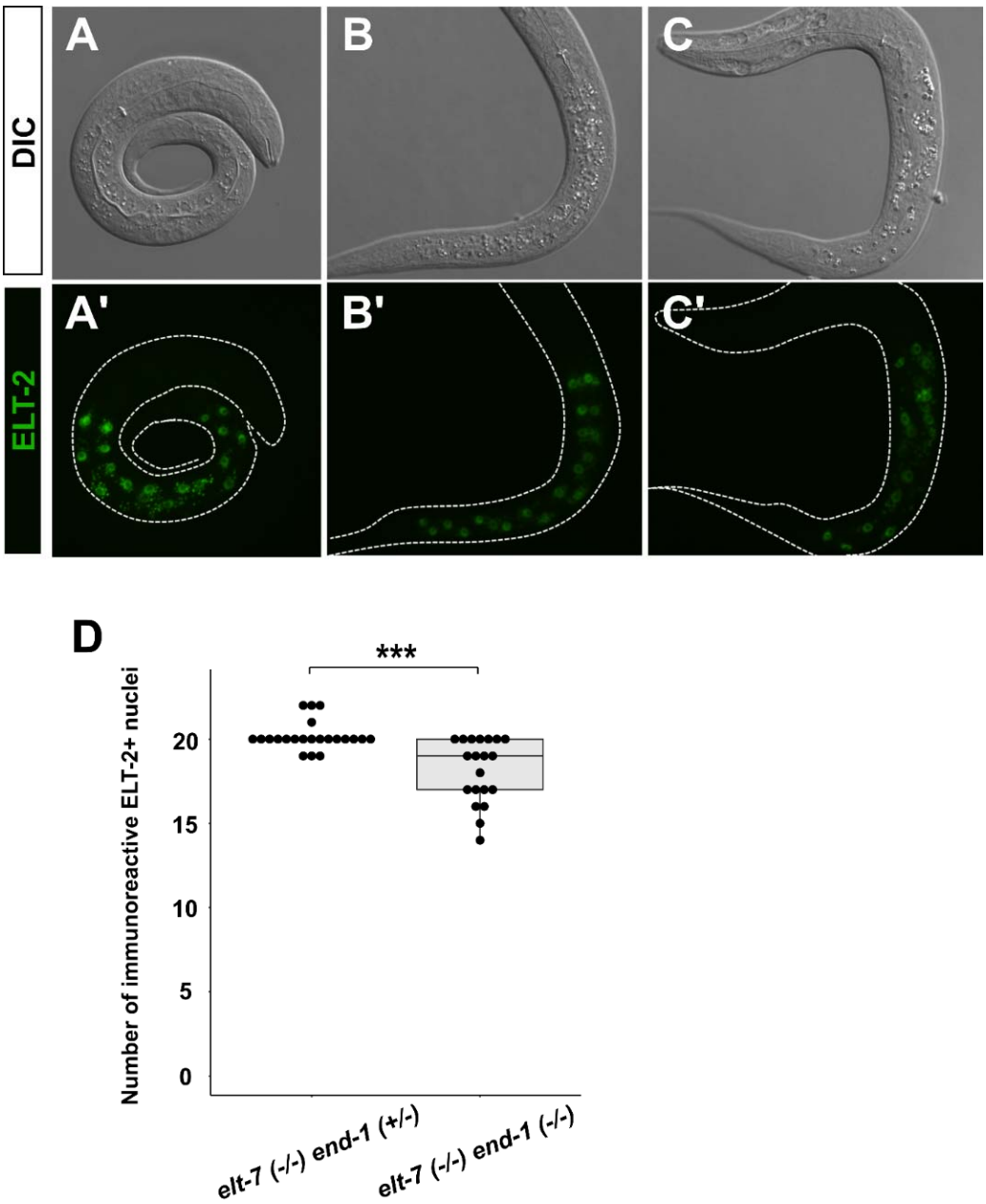


Supplementary Figure 4: The endoderm GATA factors are deployed in temporal order. (A) Temporal expression of *med-2*, *med-1*, *end-3,* and *end-1* revealed by single-cell transcriptomic analysis (http://tintori.bio.unc.edu/) (Tintori et al., 2016). (B) Expression of MED-2 and MED-1 protein-fusion reporters in staged embryos.



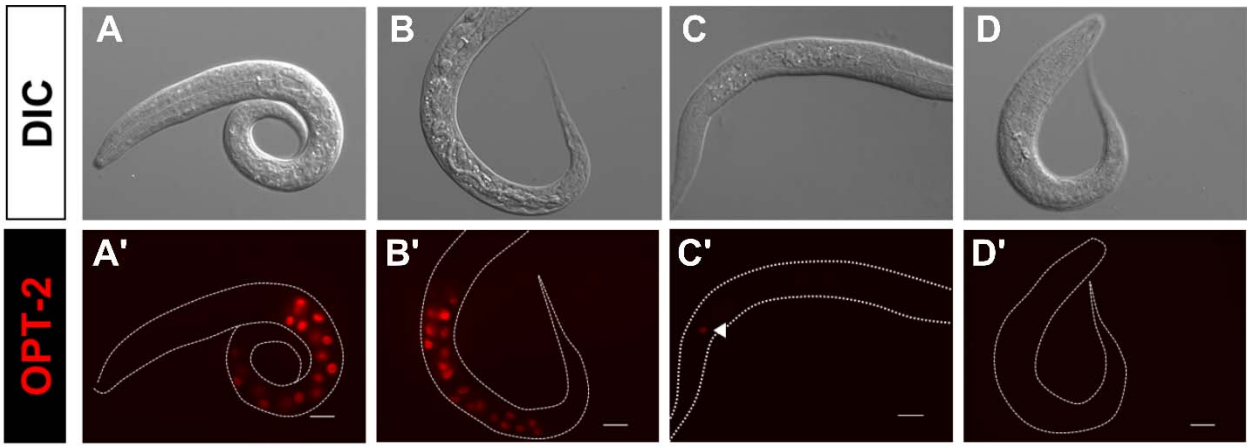
Supplementary Figure 5: Reduced number of differentiated intestinal cells *in elt-2(-); elt-7(RNAi)* animals.

On average, *elt-2(-)* animals contain 19.3 cells, while *elt-2(-); elt-7(RNAi)* animals contain 14.1 cells. The number of gut cells were scored by the expression of *elt-2p::GFP* transcriptional reporter *wIs84*. *** $p \leq 0.001$ by Wilcoxon tests.



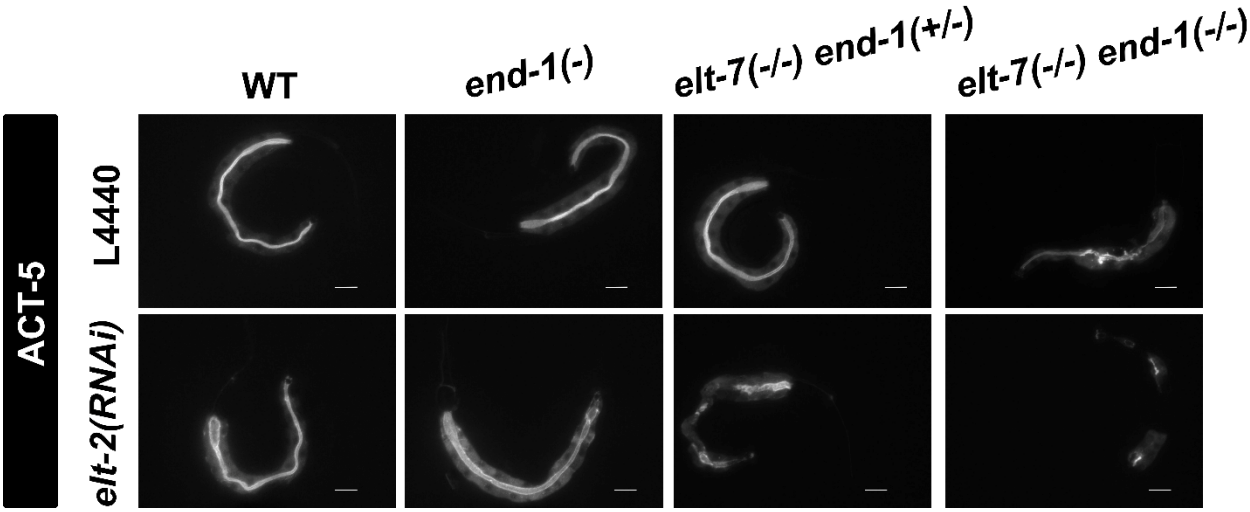
Supplementary Figure 6: Reduced number of differentiated intestinal cells in *elt-7(-) end-1(-)* animals.

(A-C) The expression of *elt-2::GFP* translational reporter in (A, A') wildtype, (B, B') *elt-7(-/-) end-1(+/-)*, and (C, C') *elt-7(-/-) end-1(-/-)* L1 larvae. (D) *elt-7(-/-) end-1(-/-)* double mutants contain fewer cells expressing immunoreactive ELT-2 than *elt-7(-/-) end-1(+/-)* animals. *** $p \leq 0.001$ by Wilcoxon tests.

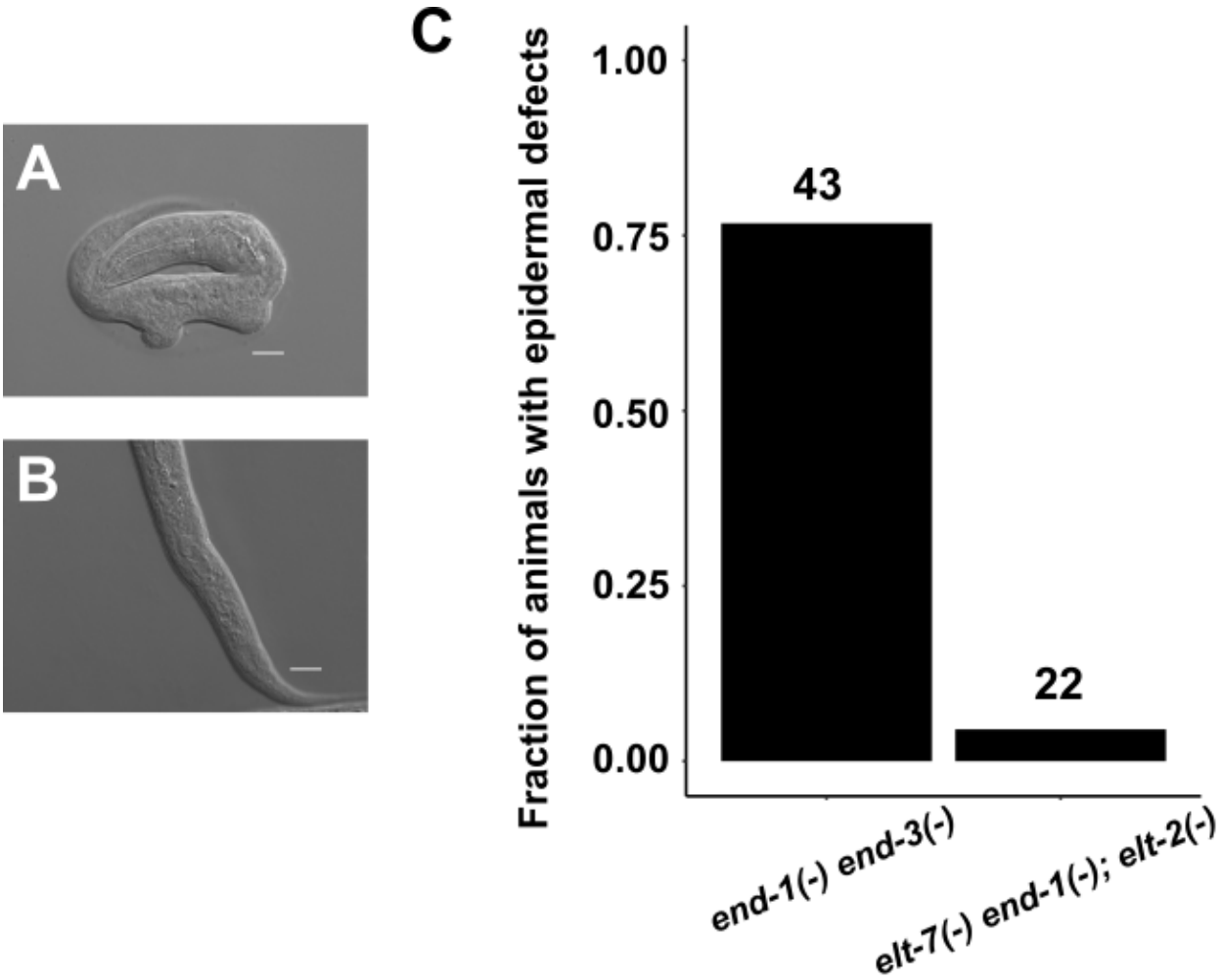


Supplementary Figure 7: Reduced number of differentiated gut cells in differentiation-defective mutants.

(A-D) DIC images of (A) wildtype, (B) *elt-2(-)*, (C) *elt-2(-); elt-7(RNAi),* and (D) *elt-7(-) end-1(-); elt-2(-)* animals. While wildtype and *elt-2(-)* worms contain a differentiated gut, *elt-2(-); elt-7(RNAi)* animals lack evident lumen and contains sporadic patches of gut granules. *elt-7(-) end-1(-); elt-2(-)* triple mutants show no apparent signs of differentiation. (A'-D') Fluorescent images of worms in (A-D) show expression of *opt-2p::mCherry*. The number of *opt-2*-expressing cells is markedly reduced in *elt-2(-); elt-7(RNAi)* (arrowhead). *opt-2* expression is completely abolished in *elt-7(-) end-1(-); elt-2(-).* Scale bars = 10 µm.

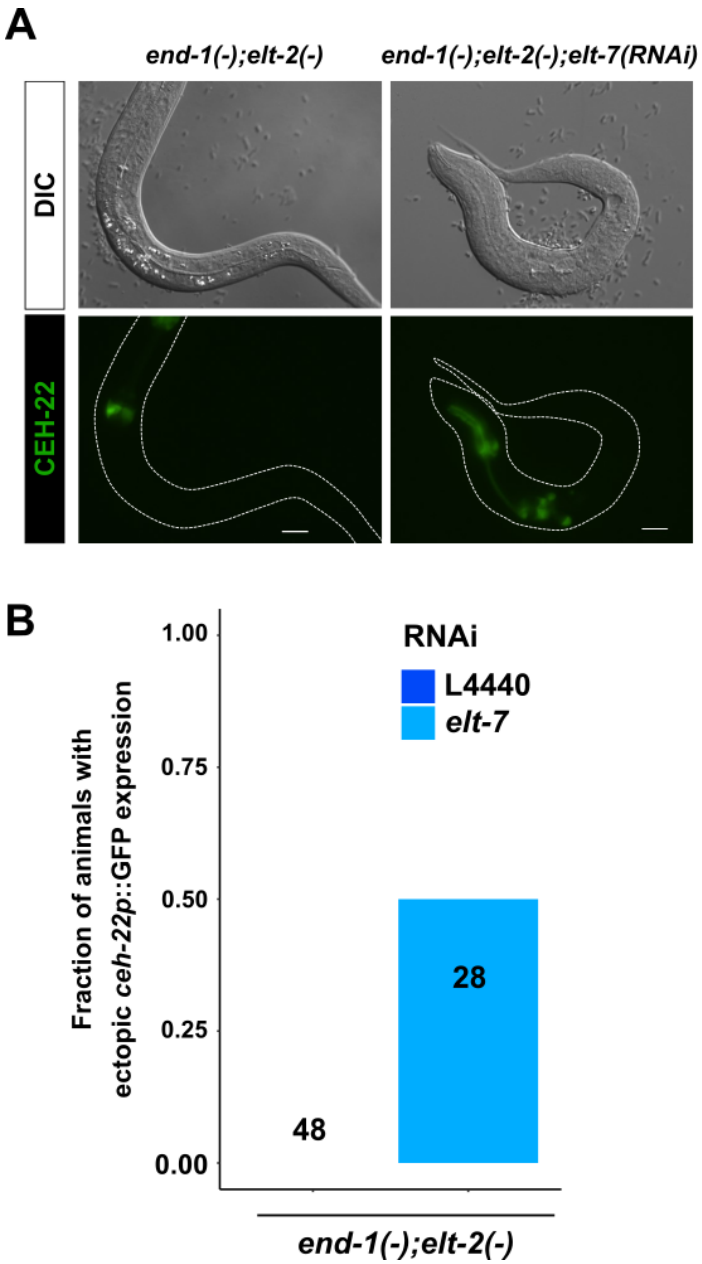


Supplementary Figure 8: END-1, ELT-7, and ELT-2 regulate *act-5* expression. *act-5* transgene (*jyIs13*) is expressed strongly in the intestine and weakly in the excretory canal cells. *act-5::GFP* signals appear sporadic in *elt-7(-/-) end-1(+/-); elt-2(RNAi)* and *elt-7(-/-) end-1(-/-)* animals, and are almost completely missing in *elt-7(-/-) end-1(-/-); elt-2(RNAi)* mutant. The residual *act-5* expression in *elt-7(-/-) end-1(-/-); elt-2(RNAi)* may be due to incomplete RNAi penetrance. Scale bars = 10 µm.



Supplementary Figure 9: Gross epidermal defects in GATA mutants.

(A) A representative *end-1(-) end-3(-)* arrested L1 shows gross epidermal defects due to E→C misspecification. (B) *elt-7(-) end-1(-); elt-2(-)* mutant does not show obvious epidermal defects, as observed by DIC microscopy. Scale bars = 10 µm. (C) A large fraction of *end-1(-) end-3(-)* mutants showed deformations of the epidermis, which was rarely observed in *elt-7(-) end-1(-); elt-2(-)* worms. Number of animals scored for each genotype is indicated.



Supplementary Figure 10: Eliminating *end-1*, *elt-7* and *elt-2* causes ectopic expression of *ceh-22* reporter.

(A) A representative *end-1(-); elt-2(-)* worm contains a differentiated gut with defined lumen (top) and wildtype expression pattern of *ceh-22* that is restricted to the pharynx (bottom). On the other hand, *end-1(-); elt-2(-); elt-7(RNAi)* mutant shows no sign of gut differentiation as observed by DIC microscopy (top), and ectopic expression of *ceh-22p::GFP* reporter (bottom). Scale bars = 10 µm. (B) Knocking out *elt-7* in *end-1(-); elt-2(-)* worms causes ectopic expression of *ceh-22p::GFP* marker, as shown in (A). Number of animals scored for each genotype is indicated.

Supplementary Table 1: Worm strains used in this study.

Strain	Genotype	Source
MS1893	*unc-119(ed9?) III; irSi24[opt-2p::mCherry::H2B, Cb-unc-119(+)] IV*	(Choi et al., 2017)
IA105	*unc-76(e911)V; ijIs12[dpy-7p::GFP::lacZ, unc-76(+)]*	(Gilleard and McGhee, 2001)
SYS549	*ujIs113[pie-1p::mCherry::H2B; nhr-2p::mCherry::HIS-24-let858UTR;unc119(+)] II; end-1(dev162[end-1::mNeonGreen]) V*	(Li et al., 2019)
JR1188	*wIs95[med-1p::GFP::MED-1+rol-6(su1006)]*	(Maduro et al., 2001)
MS1362	*med-2(cx9744) III; med-1(ok804) X; irEx581[T24D3, sur-5::dsRed]*	(Maduro et al., 2007)
MS386	*med-2(cx9744) III; dpy-11(e224) end-1(ok558) V*	(Maduro et al., 2007)
MS387	*dpy-11(e224) end-1(ok558) V; med-1(ok804) X*	(Maduro et al., 2007)
MS401	*end-3(ok1448) V; med-1(ok804) X*	(Maduro et al., 2007)
MS436	*med-2(cx9744) III; end-3(ok1448) V*	(Maduro et al., 2007)
MS1248	*end-1(ok558) end-3(ok1448) V; irEx568[K10F6, F58E10, pAS152(sur-5::dsRed)]*	(Owraghi et al., 2010)
JR3295	*elt-7(tm840) V; elt-2(ca15) X; irEx404[unc-119::CFP and elt-2(+)]*	(Sommermann et al., 2010)
MS851	*elt-2(ca15) X; irEx404[unc-119::CFP, elt-2(+)]*	(Sommermann et al., 2010)
RW10479	*unc-119(ed3) III; itIs37[pie-1p::mCherry::H2B::pie-1*	CGC

	3'UTR + unc-119(+)] IV; stIs10116[his-72(promoter)::his-24::mCherry::let-858 3'UTR + unc-119(+)]; stIs10426[med-2::TGF(6.2B3)::GFP::TY1::3xFLAG]	
AZ217	*unc-119(ed3) ruIs37 [myo-2p::GFP + unc-119(+)] III*	CGC
FX840	*elt-7(tm840) V*	CGC
MS123	*med-2(cx9744) II*	CGC
OP56	*unc-119(ed3) III; gaEx290[elt-2::TY1::EGFP::3xFLAG(92C12) + unc-119(+)]*	CGC
RB1331	*end-3(ok1448) V*	CGC
RB930	*med-1(ok804) X*	CGC
VC271	*end-1(ok558) V*	CGC
ERT60	*jyIs13 [act-5p::GFP::ACT-5 + rol-6(su1006)] II*	CGC
OH15876	*pha-4(ot946[pha-4::GFP]) V*	CGC
JR4405	*ruIs37 [myo-2p::GFP + unc-119(+)] III; him-5(e1490) V; elt-2(ca15) X; IrEx404[unc-119::CFP, elt-2(+)]*	This study
JR4384	*ruIs37 [myo-2p::GFP + unc-119(+)] III; elt-7(tm840) V; elt-2(ca15) X; IrEx404[unc-119::CFP, elt-2(+)]*	This study
JR4369	*hum-1(cv21[hum-1::RFP]) I; elt-7(tm840) end-1(ox134)/ elt-7(tm840) tmC12[egl-9(tmIs1197)] V; gaEx290[elt-2::TY1::EGFP::3xFLAG(92C12) + unc-119(+)]*	This study
JR4371	*hum-1(cv21[hum-1::RFP]) I; gaEx290[elt-2::TY1::EGFP::3xFLAG(92C12) + unc-119(+)]*	This study
JR4368	*ruIs37 [myo-2p::GFP + unc-119(+)] III; end-1(ok558) end-3(ok1448) V; irEx568[K10F6, F58E10, pAS152(sur-5::dsRed)]*	This study
JR2488	*him-8(e1489) IV; end-1(ox134) V*	This study
JR3341	*elt-7(tm840) end-1(ox134)/ dpy-11(e224) unc-76(e911) V*	This study

JR3342	*elt-7(tm840) end-3(ok1448) V*	This study
JR4037	*elt-2(ca15) X; ijIs12[dpy-7p::GFP::lacZ, unc-76(+)]; irEx404[unc-119::CFP, elt-2(+)]*	This study
JR4063	*him-8(e1489) irSi24[opt-2p::mCherry::H2B, Cb-unc-119(+)] IV; elt-2(ca15) X; irEx404[unc-119::CFP, elt-2(+)]*	This study
JR4097	*elt-7(tm840) end-1(ox134)/ elt-7(tm840) tmC12[egl-9(tmIs1197)] V*	This study
JR4194	*jcIs1[ajm-1p::AJM-1::GFP + unc-29(+) + rol-6(su1006)] IV; elt-7(tm840) end-1(ox134)/ elt-7(tm840) tmC12[egl-9(tmIs1197)] V*	This study
JR4195	*elt-7 (tm840) end-1(ox134)/ elt-7(tm840) tmC12[egl-9(tmIs1197)] V; elt-2(ca15) X; irEx404[unc-119::CFP, elt-2(+)]*	This study
JR4200	*him-8(e1489) irSi24 [opt-2p::mCherry::H2B, Cb-unc-119(+)] IV; elt-7(tm840) end-1(ox134)/ elt-7(tm840) tmC12[egl-9(tmIs1197)] V; elt-2(ca15) X; irEx404[unc-119::CFP, elt-2(+)]*	This study
JR4304	*elt-7 (tm840) end-1(ox134)/ elt-7(tm840) tmC12[egl-9(tmIs1197)] V; wEx1751[cdf-1p::CDF-1::GFP + rol-6(su1006)]*	This study
JR4313	*elt-7(tm840) end-1(ox134)/ elt-7(tm840) tmC12[egl-9(tmIs1197)] V; gaEx290[elt-2::TY1::EGFP::3xFLAG(92C12) + unc-119(+)]*	This study
JR4364	*ruIs37 [myo-2p::GFP + unc-119(+)] III; elt-7(tm840) end-1(ox134)/ elt-7(tm840) tmC12[egl-9(tmIs1197)] V; elt-2(ca15) X; irEx404[unc-119::CFP, elt-2(+)]*	This study
JR4367	*ruIs37 [myo-2p::GFP + unc-119(+)] III; elt-7(tm840) end-1(ox134)/ elt-7(tm840) tmC12[egl-9(tmIs1197)] V*	This study
MS880	*dpy-11(e224) end-1(ok558) V; elt-2(ca15) X;*	This study

	irEx404[unc-119::CFP, elt-2(+)]	
JR4407	*jcIs1[ajm-1p::AJM-1::GFP + unc-29(+) + rol-6(su1006)] IV; end-1(ox134) V*	This study
MS900	*end-3(ok1448) V; elt-2(ca15) X; irEx404[unc-119::CFP, elt-2(+)]*	This study
JR4412	*jyIs13 [act-5p::GFP::ACT-5 + rol-6(su1006)] II; elt-7(tm840) end-1(ox134)/ elt-7(tm840) tmC12[egl-9(tmIs1197)] V*	This study
JR4413	*jyIs13 [act-5p::GFP::ACT-5 + rol-6(su1006)] II; end-1(ox134) V*	This study
JR4439	*him-8(e1489) IV; cuIs1[ceh-22p::GFP + rol-6(su1006] end-1(ox134) V; elt-2(ca15); irEx404[unc-119::CFP, elt-2(+)]*	This study
JR4446	*elt-7(tm840) end-1(ox134)/ elt-7(tm840) tmC12[egl-9(tmIs1197)] V; kcIs6[ifb-2p::IFB-2::CFP]*	This study
JR4198	*end-1(ok558) end-3(ok1448) V; ijIs12[dpy-7p::GFP::lacZ, unc-76(+)]; irEx568[K10F6, F58E10, pAS152(sur-5::dsRed)]*	This study
JR4452	*end-1(ox134) V; elt-2(ca15); ijIs12[dpy-7p::GFP::lacZ, unc-76(+)]; irEx404[unc-119::CFP, elt-2(+)]*	This study
JR3315	*elt-2(ca15) X; wIs84[elt-2p::GFP+ rol-6(su1006)]; irEx404[unc-119::CFP, elt-2(+)]*	This study
JR4463	*pha-4(ot946[pha-4::GFP]) V; elt-2(ca15) X; irEx404[unc-119::CFP, elt-2(+)]*	This study

Supplementary File 1 (.xlsx): Model parameters and outputs of the endoderm GRN. Expression of each factor is determined by the concentration of its activators multiplied by a coefficient *a*, representing the strength of the inputs. SKN-1 expression follows a square wave in EMS blastomere. POP-1 expression follows a square wave in the E cell. Feedback coefficients *f* become nonzero once their respective factors surpass a certain threshold Φ.

1026