Developmental Biology 412 (2016) 160–170

Contents lists available at ScienceDirect

Developmental Biology

journal homepage: www.elsevier.com/locate/developmentalbiology

Genomes & Developmental Control

Mutagenesis of GATA motifs controlling the endoderm regulator *elt-2* reveals distinct dominant and secondary *cis*-regulatory elements

Lawrence Du, Sharon Tracy, Scott A. Rifkin *

Section of Ecology, Behavior, and Evolution, Division of Biological Sciences, University of California, San Diego, 9500 Gilman Drive, La Jolla, CA 92093-0116, United States

ARTICLE INFO

Article history:
Received 1 December 2015
Received in revised form
3 February 2016
Accepted 10 February 2016
Available online 16 February 2016

Keywords:
Endoderm
Cis-regulation
Promoter
smFISH
Transcription
GATA
C. elegans

ABSTRACT

Cis-regulatory elements (CREs) are crucial links in developmental gene regulatory networks, but in many cases, it can be difficult to discern whether similar CREs are functionally equivalent. We found that despite similar conservation and binding capability to upstream activators, different GATA *cis*-regulatory motifs within the promoter of the *C. elegans* endoderm regulator *elt-2* play distinctive roles in activating and modulating gene expression throughout development. We fused wild-type and mutant versions of the *elt-2* promoter to a *gfp* reporter and inserted these constructs as single copies into the *C. elegans* genome. We then counted early embryonic *gfp* transcripts using single-molecule RNA FISH (smFISH) and quantified gut GFP fluorescence. We determined that a single primary dominant GATA motif located 527 bp upstream of the *elt-2* start codon was necessary for both embryonic activation and later maintenance of transcription, while nearby secondary GATA motifs played largely subtle roles in modulating postembryonic levels of *elt-2*. Mutation of the primary activating site increased low-level spatiotemporally ectopic stochastic transcription, indicating that this site acts repressively in non-endoderm cells. Our results reveal that CREs with similar GATA factor binding affinities in close proximity can play very divergent context-dependent roles in regulating the expression of a developmentally critical gene *in vivo*.

1. Introduction

For the nematode *C. elegans* and its close relatives, early embryonic development is characterized by a tight link between cell lineage and cell fate that is largely determined by transcriptional gene regulatory networks (GRNs). Determining how transcription factors activate their respective targets within a GRN at the *cis*-regulatory level is key to understanding how multicellular organisms develop robustly.

However, understanding *cis*-regulation has been complicated by the fact that the vast majority of eukaryotic transcription factors have very short DNA binding domains, often leading to vastly more potential cognate binding sites than real functional targets (Mirny and Wunderlich, 2008). Additionally, individual transcription factors can co-occur with one or more paralogous factors from the same family leading to multiple transcription factors sharing individual *cis*-regulatory sites. Although recent years have seen vast advances in the mapping of transcription factors to their binding sites though techniques such as ChIP-seq (Gerstein et al., 2010), such techniques do not necessarily reveal whether bound sites are functionally equivalent. Many transcription factors are known to act as both activators and repressors depending on context.

Low target specificity, gene duplications, and contextual role switching have the potential to play a role in advancing developmental robustness. During early embryonic development, transcriptional networks must be robust to extrinsic insults as well as intrinsic variability at the molecular level. Cell divisions need to be spatially and temporally coordinated in the face of environmental variability and stochastic fluctuations of key molecules.

The transcriptional regulation of the *C. elegans* endoderm specifying gene *elt-2* is a good model for studying how *cis*-regulatory mechanisms impact developmental robustness. The gene *elt-2* is an essential switch for the endoderm cell fate decision and a fundamental developmental bottleneck: failure to activate *elt-2* results in a lethal absence of endoderm. The major *trans*-activators of *elt-2* are well characterized and have been demonstrated to contribute to developmental robustness at the *trans*-level. END-3, END-1, and ELT-7 are closely related GATA transcription factors that redundantly activate *elt-2* during early embryonic development (Lowry et al., 2009; Raj et al., 2010; Sommermann et al., 2010; Zhu et al., 1997, 1998), and ELT-2 maintains its own transcription through larval and adult stages by autoregulation

* Corresponding author.
E-mail address: sarifkin@ucsd.edu (S.A. Rifkin).

http://dx.doi.org/10.1016/j.ydbio.2016.02.013
0012-1606/

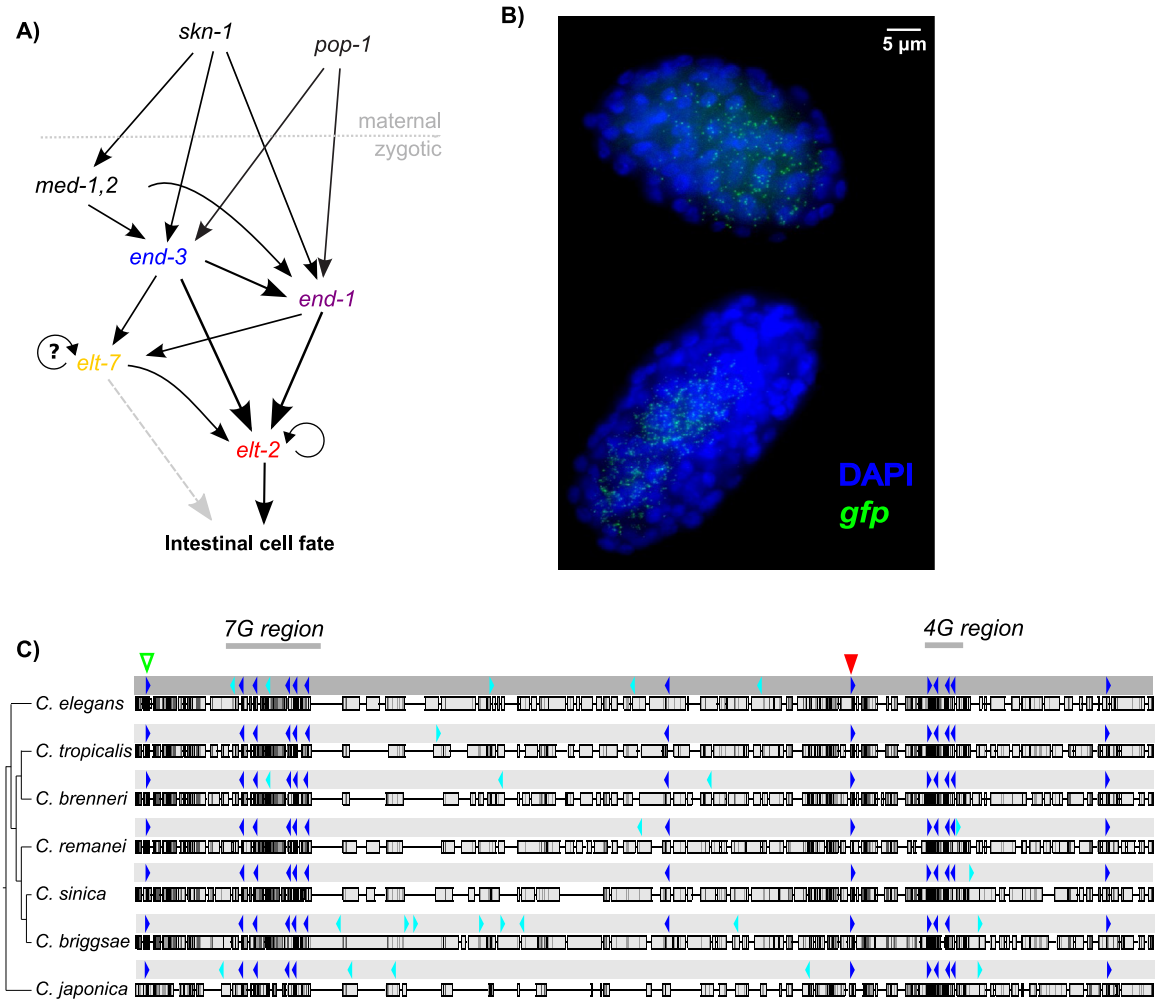


Fig. 1. The *elt-2* activation network and upstream promoter region. (A) Network diagram of *elt-2* activators. (B) Maximum Z-stack projection of an smFISH image from strain 1879:WT. (C) 1879 bp promoter region upstream of *elt-2* aligned to orthologous regions in *Elegans* supergroup members. Dark blue indicates a heavily conserved HGATAR motif and cyan indicates a weakly conserved HGATAR motif. The red, filled triangle indicates the position of the −527 bp ACTGATAAGA "A-site" motif. The green, open triangle indicates position of a −1857 bp ACTGATAAGG motif. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

(Fukushige et al., 1999) (Fig. 1A). Single null mutants of END-1, END-3, or ELT-7 are largely viable-with only transient developmental anomalies occurring in *end-1* and *end-3* single null mutants and a low (5–9%) rate of developmental failure in *end-3* mutants (Boeck et al., 2011; Maduro et al., 2005; Sommermann et al., 2010). A paralogous pair of redundant and nearly identical GATA factors-*med-1,2*-also helps to activate *end-3* and *end-1* (Maduro et al., 2001, 2007). Raj et al. (2010) demonstrated that particular mutations in the upstream maternal activating factor *skn-1* result in failure to activate *med-1,2* and *end-3* and highly variable expression of *end-1*. Noisy *end-1* expression, in turn, leads to bimodal *elt-2* expression states (Raj et al., 2010). The presence of redundant *trans*-activating factors effectively buffers the activation of *elt-2* from variability in levels of any single activator.

Despite our good understanding of *elt-2*'s *trans*-activators, little is known about how these *trans*-activators operate at the *cis*-regulatory level. The exact sequences and relative positions of the *cis*-regulatory motifs necessary for driving *elt-2* expression have not been determined, nor are there apparent TATA box (GTATAWWAG) or Sp1 core promoter motifs in the immediate region upstream of the *elt-2* transcriptional start site (TSS) (WormBase release WS220) (Harris et al., 2010; Saito et al., 2013). Furthermore, we could not identify any sequence similarities to the known basal promoter fragment of the *pes-10* gene. Based on the fact that all the known embryonic activators of *elt-2* are GATA transcription factors, we can narrow down the candidate *cis*-regulatory sequences considerably. The ~2 kb region upstream of the *elt-2* start codon in *C. elegans* contains 18 of the consensus GATA factor binding motif HGATAR, with 13 conserved in sequence and relative spacing throughout the *Elegans* supergroup (sequence data from *C. elegans, C. tropicalis* , *C. brenneri*, *C. remanei*, *C. sinica*, *C. briggsae*, and *C. japonica*) (Félix et al., 2014; Huang et al., 2014) (Fig. 1C).

This conglomeration of potential GATA factor binding sites in the *elt-2* upstream region suggests several possible ways that these motifs might interact with the *trans*-factors END-3, END-1, ELT-7, and ELT-2 itself to control *elt-2* transcriptional activation. Perhaps many independently dispersed transcription start sites, driven or aided by GATA factor binding, contribute additively and redundantly to overall gene expression levels and noise (Juven-Gershon and Kadonaga, 2010). Under this dispersed promoter scenario, mutation of single HGATAR motifs might be expected to reduce transcription activation proportionate to the number of motifs mutated (Davidson, 2001; Flores et al., 2000). Alternatively, but not exclusively, *elt-2 cis*-activation could also be driven by a combinatorial code involving binding of different GATA factors with different specificities. Under a combinatorial control scenario, mutation of any single HGATAR motif in a larger combinatorial code should result in an equivalent impact on gene expression as mutating any single motif in the same code. Finally, the contributions of different sites may instead be unequal, with one or a

few key sites responsible for the majority of expression and the rest playing minor supporting roles.

In addition, few studies have deeply examined the relationship between between *cis*-regulatory elements (CREs) and gene expression variability. For a gene like *elt-2*, variability in early embryonic activation might be expected to have severe consequences for viability.

In this study, we dissect the *elt-2* promoter in a reporter transgene context to determine how conserved transcription factor binding sites drive activation and maintenance of a key developmental regulator. Using a technique for labeling individual mRNAs, we also precisely capture the relationship between conserved features of the *elt-2* promoter with gene expression variability in early embryonic development.

2. Materials and methods

2.1. Strain generation

To control for copy number variations that could affect gene expression levels, we generated single copy insertions of promoter *gfp* reporter lines using MosSCI (Frøkjaer-Jensen et al., 2008; Frøkjær-Jensen et al., 2012). To generate the reporter, we used PCR fusion to join the *elt-2* 3′ UTR to a *gfp* fragment amplified from pPD95.81 (gift from Andrew Fire; Addgene kit # 1000000001). We generated mutant promoter variants by PCR fusion of synthetic oligonucleotides containing mutated HGATAR motifs to PCR amplified *elt-2* promoter sequence. GIBSON assembly was used to fuse wild-type and mutant promoter constructs to the *gfp::elt-2* 3′ UTR DNA fragment and a pCFJ350 MosSCI backbone vector (gift from Erik Jorgenson; Addgene plasmid #34866). All *Pelt-2::gfp::elt-2 3′UTR* reporters were integrated at site ttTi5605 (strain EG6699) with stable integration confirmed by PCR screening of insert junctions into the *C. elegans* genome.

2.2. In vitro binding assays

For EMSAs, DNA binding domain sequences for GATA transcription factors were cloned into pET His6 TEV LIC (1B) for N-terminal His6 tagging (gift from Scott Gradia; Addgene plasmid # 29653). Constructs were transformed into Rosetta 2(DE3) pLysS competent cells, and fusion protein expression was induced by addition of 0.1 mM IPTG at 18 °C for 12 h. His6 tagged proteins were purified by use of TALON metal affinity beads (Clontech). Binding assays were done at room temperature in EMSA buffer (10 mM HEPES, 200 mM NH_4 OAc, 30 mM NaCl, 1.5 mM $MgCl_2$, 0.5 uM $Zn(OAc)_2$, 0.2 mg/ml BSA, 1 mM DTT, 8% glycerol, pH 7.0). EMSAs were run on a 8% 29:1 Acrylamide/Bis-acrylamide 8% glycerol 0.5 × TBE gel and stained with Sybr Green Nucleic acid gel stain (Molecular Probes S-7563) (see Supplemental methods for more detail).

2.3. smFISH data collection

For imaging of embryonic development, we collected and fixed embryos collected from adults 52–53 h post-synchronization in 4% formaldehyde. Collected embryos were hybridized to Cy3::*gfp*, ATTO 647N::*elt-2*, and Alexa 594::*end-1* smFISH probes (Raj et al., 2010, 2008) (Stellaris). Nuclei were labeled with DAPI. For each smFISH data set, we generated Z-stack images of at least 170 embryos with a maximum spacing of 0.4 uM at 100 × magnification on an epifluorescence microscope. For quantifying spot counts, we used an in-house developed machine learning spot classification tool, AroSpotFinding Suite, to automatically collect spot data counts (Rifkin, 2011; Wu and Rifkin, 2015). Nuclei were counted by hand and used as a measure of developmental stage.

2.4. Gut fluorescence quantification

To obtain consistent gut reporter gene fluorescence data, we determined that it was important to control for food density. We found that different food concentrations could affect both mean gut fluorescence intensity and overall size. To control for food levels, we grew all strains in liquid culture at controlled food concentrations. We aliquoted and pelleted 2 L of overnight OP50 grown in LB culture and aspirated off all excess liquid. We then weighed the mass of each pellet, and froze each pellet at −80 °C. For each data collection day, an OP50 food pellet would be thawed and resuspended in S-medium at a concentration of 40 mg/ml. Worms synchronized overnight S-medium would then be placed in a solution of OP50 and S-medium for 22 h at a concentration of approximately 500 worms per ml. For imaging, the synchronized worms were mounted onto 3% agarose pad slides, and anaesthetized with 10 mM levamisole. Images were taken on an AxioImager R1 at 10 × magnification. Quantification was performed using a custom MATLAB script (available upon request).

2.5. Statistical analysis of smFISH data

To determine whether the trajectories of gene expression for different strains were significantly different across time, we calculated the difference between smoothing splines fit to the data for each strain. We constructed a null distribution for this difference trajectory by randomly assigning the strain labels for each datapoint (permuting within a developmental stage: 0–1E, 2E, 4E, 8E), fitting splines to the shuffled dataset, and calculating the difference between these shuffled splines. We repeated this 10,000 times for each pair of trajectories to form null distributions. Because we were looking for significant differences in expression at any point along the trajectories, we adjusted the significance level cutoff to account for multiple testing. We estimated the number of parameters used in the splines and used the Dunn-Šidák method to determine a conservative single-test significance level that would yield an experiment-wise significance level of 0.05 (Ury, 1976). These adjusted cutoffs varied slightly between comparisons but were around 0.006.

3. Results

Since *elt-2* is an essential gene, we used a reporter construct to investigate the relationship between conserved and putatively functional features of the *elt-2* promoter and *elt-2* promoter-driven gene expression levels and noise. We integrated wild-type and mutant *elt-2* promoter–reporter constructs (*Pelt-2::gfp::elt-2 3′ UTR*; see Methods) into the *C. elegans* genome at a defined location using *Mos1*-mediated single copy insertion (MosSCI) (Frøkjaer-Jensen et al., 2008; Frøkjær-Jensen et al., 2012). We then measured *gfp* transcript levels across early embryonic development using single-molecule RNA Fluorescence *In Situ* Hybridization (smFISH) (Raj et al., 2008) (Fig. 1B). We also confirmed that transgene insertion did not impact endogenous *elt-2* levels significantly by labeling *elt-2* as a control (Supplemental Fig. S10).

Previous research had shown that a 5.1 kb region upstream of the *elt-2* start codon is sufficient to drive gut reporter expression when introduced to wild-type worms in the form of an extrachromosomal array (Fukushige et al., 1998', 1999). We found that this large region contained a putative coding region (C39B10.7) and non-coding RNA (C33D3.6). In order avoid duplicating these potential *trans*-regulatory elements, we initially looked at the shorter 1879 bp region between the putative non-coding RNA

Table 1
Table of wild-type and HGATAR mutant strains mutated for this study. Dark blue indicates a heavily conserved HGATAR motif and cyan indicates a weakly conserved HGATAR motif. Red indicates an artificially mutated HGATAR motif. *(Asterisk) indicates low-level transcription observed.

Strain	Promoter Description	Embryo *gfp*	Adult GFP	Graphic
1879:WT	1879 bp; wild-type	+	+	
1879:4G	1879 bp; 4 HGATAR mutations	+	+	
1879:11G	1879 bp; 11 HGATAR mutations	+	+	
1879:A	1879 bp; 1 ACTGATAAG mutation at -527 bp	-*	-	
422:WT	422 bp; wild-type	-	-	
613:WT	613 bp; wild-type	+	+	
613:A	613 bp; 1 ACTGATAAG mutation at -527 bp	-*	-	
613:4G	613 bp; 4 HGATAR mutations	+	+	
613:A4G	613 bp; 1 ACTGATAAG and 4 HGATAR mutations	-	-	
61:WT	61 bp; wild-type	-	-	

C33D3.6 and the *elt-2* start codon. This 1879 bp region contains 18 motifs matching to the GATA transcription factor binding consensus sequence HGATAR with 10 of the 18 motifs matching to TGATAA-a more specific motif often found in the promoters of gut expressed genes (McGhee et al., 2009).

Fluorescence microscopy of the 1879:WT reporter strain showed that the 1879 bp upstream fragment was sufficient to drive production of *gfp* transcripts during embryonic development (smFISH of *gfp* transcripts) as well as the remainder of the worm's life span (GFP fluorescence) (Table 1, Supplemental Fig. S1). Although this reporter produced lower levels of expression than endogenous *elt-2* in worms with between approximately 70 and 120 nuclei, the overall expression trajectory was similar to endogenous *elt-2* (Supplemental Fig. S1) (Nair et al., 2013; Raj et al., 2010), suggesting that many if not most of the critical CREs for the majority of *elt-2* expression fall within this region and that targeted mutation of the reporter's promoter would give valuable insights into the logic and functional organization of *elt-2 cis*-regulation.

3.1. *The promoter region from −613 to −422 bp upstream of the elt-2 start codon is necessary for gene expression*

To cut down on the number of candidate HGATAR motifs that could be responsible for *elt-2* promoter-driven gene activation, we tested whether a more minimal promoter could drive gut *gfp* expression. We first generated a reporter construct encompassing 613 bp upstream of the *elt-2* start codon (Table 1 *613:WT*). This shortened promoter produced *gfp* expression levels only about 50 transcripts lower than the 1879 bp promoter during the 4E stage, suggesting that this fragment contains the primary CREs needed for driving endoderm gene expression (Figs. 2 and S2).

To determine whether presence of the *4G* motif cluster was *sufficient* for driving reporter expression, we generated a 422 bp un-mutated *elt-2* promoter–reporter strain (Table 1 *422:WT*). The 422:WT strain failed to produce any reporter expression during embryogenesis (assessed by smFISH) or in adults (assessed by GFP

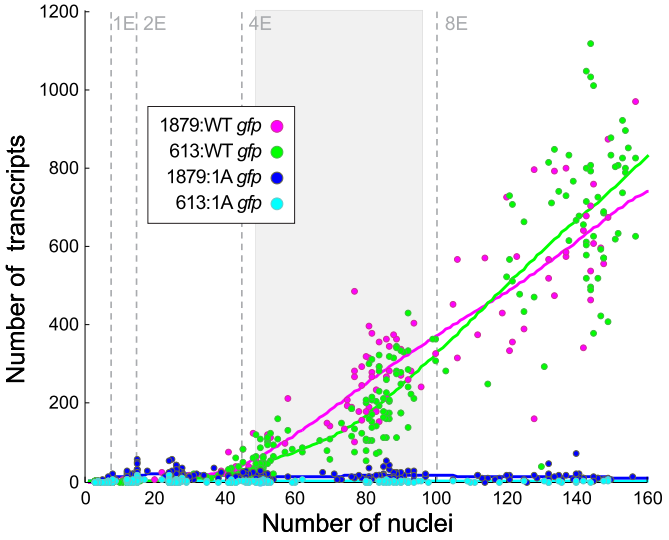


Fig. 2. The majority of *elt-2* promoter-driven expression is dependent upon a single ACTGATAAG motif at −527 bp. (A) *gfp* expression levels of 1879:WT ($n=187$), 613:WT ($n=459$), 1879:A ($n=253$), and 613:A ($n=217$). Lines represent smoothing splines of data. The shaded gray area represents the part of the developmental trajectories that are significantly different between the longer and shorter WT promoters (Supplemental Fig. S2).

fluorescence). These results demonstrate that one or more CREs located between −613 bp and −422 bp upstream of the *elt-2* start codon are necessary for promoter-driven gene expression. Within this region, we identified a single highly conserved ACTGATAAGA motif at −527 bp (Fig. 1C; *red arrow*).

3.2. *An ACTGATAAG sequence −527 bp upstream of the elt-2 start codon is necessary for elt-2 promoter-driven gene expression*

The −527 bp ACTGATAAGA sequence matched perfectly to the highest scoring sequence found to be associated with intestinally

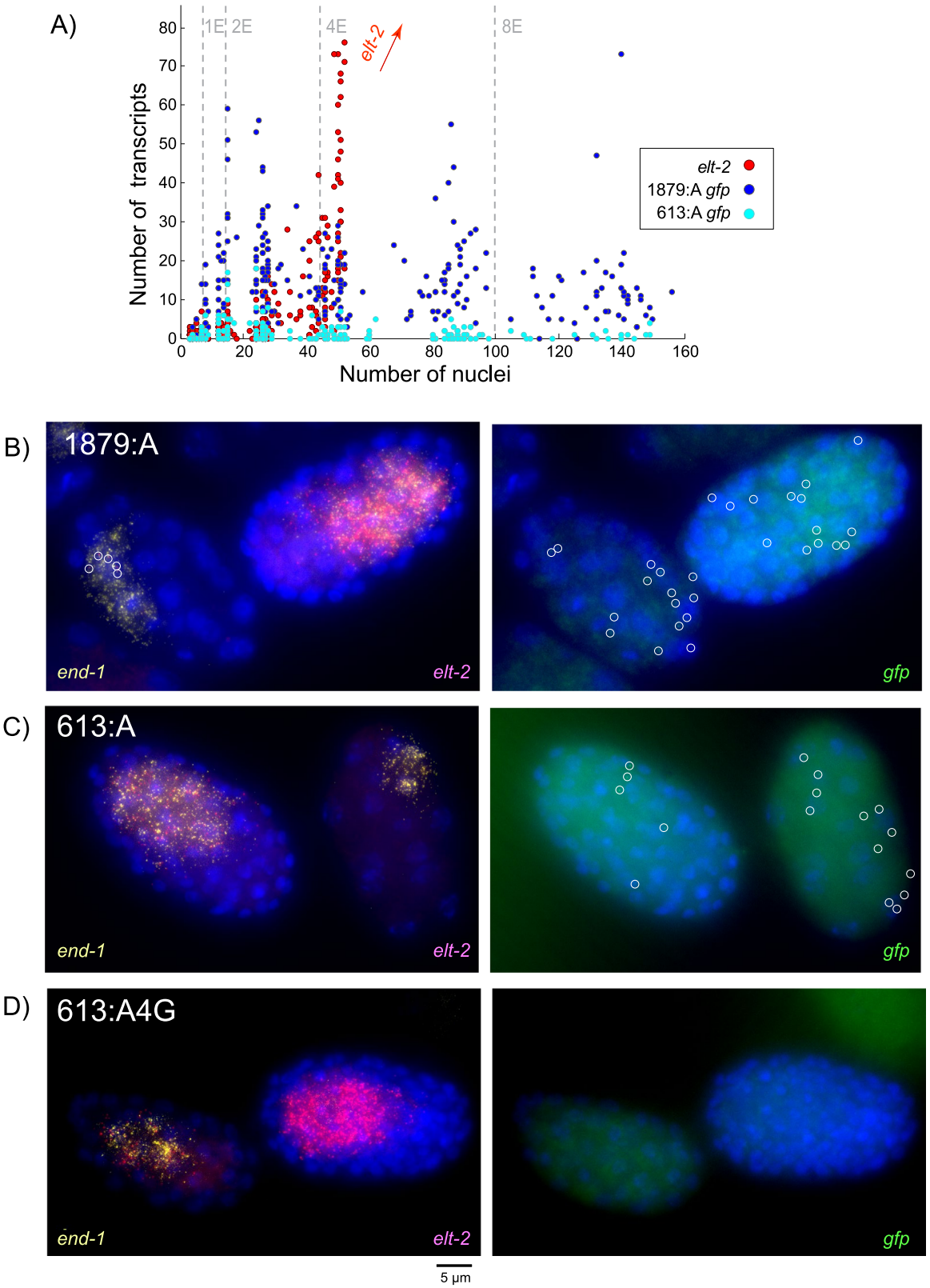


Fig. 3. Stochastic low-level ectopic transcription occurs in the absence of the −527 bp ACTGATAAG element with an increase in transcription during the 1E and 2E stages. (A) *gfp* transcript counts for A-site mutants 1879:A ($n=253$) and 613:A ($n=217$) compared to *elt-2* in the same embryos (*elt-2* data pooled from 1879:A and 613:A ($n=470$); *elt-2* transcript counts greater than 80 not shown). (B–D) Left column: *end-1* in yellow *elt-2* in red. Right column: *gfp* in green; all *gfp* transcripts circled for ease of visualization. (B) Ectopic expression of *gfp* outside of *end-1* and *elt-2* expressing E-cells in strain 1879:A (five early *elt-2* transcripts circled in left column). (C) Ectopic early *gfp* expression in 613:A. (D) Mutation of 4G HGATAR sites eliminates low-level stochastic *gfp* reporter expression. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

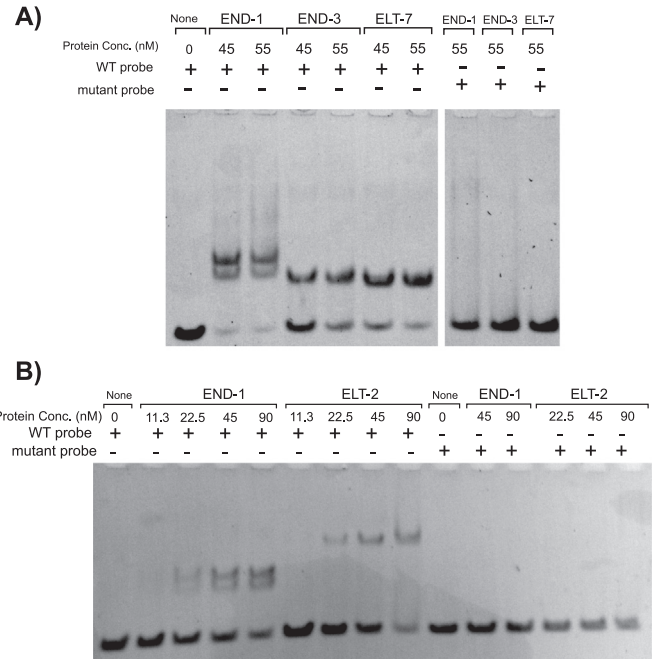


Fig. 4. END-1, END-3, ELT-2, and ELT-7 DNA binding domains all bind *in vitro* to an ACTGATAAG motif. (A, B) All EMSAs were performed using 50 bp synthetic oligonucleotide probe corresponding to the sequence −549 bp to −499 bp upstream of the *elt-2* start codon containing an ACTGATAAG motif (WT probe) or a ACTCTGTAG (mut probe). Probe concentration in all lanes is 70 nM. (A) END-1, END-3, and ELT-7 binding. Non-adjacent lanes from the same gel. (B) END-1 and ELT-2 binding.

expressed genes in McGhee et al. (2009). Furthermore, the ACTGATAAG portion of this sequence is almost perfectly conserved in the *Elegans* supergroup (in *C. remanei* the sequence is A**G**TGATAAG) (Fig. 1C; *red arrow*). Close examination of TSS data from Saito et al. (2013) revealed both dispersed low-level transcription along the 1879 bp *elt-2* upstream region and a distinct peak of transcriptional activity at −482 bp (Fig. S3).

To test whether the −527 bp ACTGATAAG motif (hereafter *A-site*) is necessary for reporter activity, we generated mutant 613 bp and 1879 bp reporters with the ACT**GATA**AG motif mutated to ACT**CTGT**AG (Table 1 *613:A, 1879:A*). Embryonic transcript expression revealed *near* total loss of embryonic reporter expression in both mutant strains (Figs. 2, 3, and S4). Additionally, we did not observe any embryonic, larval, or adult GFP expression distinguishable from regular gut autofluorescence in either mutant strain (Table 1).

These extreme drops in gene expression demonstrate that a single, non-redundant ACTGATAAG at −527 bp is necessary for the vast majority of gene expression driven by the *elt-2* promoter during and after embryonic development.

3.3. Mutation of the −527 bp ACTGATAAG element results in precocious and spatially ectopic stochastic low-level transcription

Although reporter strains with mutations in the *A-site* motif exhibited greatly reduced *gfp* expression, in many embryos reporter mRNA expression was not zero. Instead, strains 613:A and 1879:A exhibited low-level stochastic transcription during early embryonic development—an observation made possible by our highly sensitive smFISH assay. The 613:A mutant produced variable numbers of transcripts ranging from 0 to 18 during early embryonic development. Similarly, the 1879:A mutant exhibited stochastic, low-level transcription ranging from 0 to 73 transcripts over the same time span (Figs. 3A and S4). Total absence of *gfp* expression in strains 422:WT and 613:A4G confirmed that low-level expression in strains 613:A and 1879:A was not the result of leaky expression due to insertion site effects (Table 1 and Fig. 3D).

Surprisingly, the peak expression in strains 613:A and 1879:A exceeded both endogenous *elt−2* levels and *gfp* levels in wild-type reporter constructs 613:WT and 1879:WT during the 1E and 2E stages, (Figs. 3A and S4A–C). Peak expression in 1879:A also exceeded peak expression in 613:A across embryonic development, indicating that regions of the promoter upstream of −613 bp play a role in driving low-level transcription independent of the *A-site* (Fig. S4D).

Close examination of embryo images at the 1E and 2E stages revealed that in many cases, precocious stochastic transcription could occur in both E cells and other lineages. By using *elt-2, end-1*, and *gfp* smFISH probes in the same embryos, we were able to determine that *gfp* expression occurs in cells not expressing *elt-2* or *end-1* during very early development (Fig. 3B and C). This ectopic expression suggests that the *A-site* not only serves as a primary activator, but also as a transcriptional repressor against stochastic low-level transcription in non-E-cells during very early embryonic development.

We roughly estimated the level of ectopic expression at the 2E stage in both wild-type and mutant reporter strains by quantifying the amount of *gfp* not directly overlapping *end-1* and *elt-2* expression (Supplemental Fig. S5). Under our estimates, 1879:A was found to have significantly greater levels of ectopic expression (p-value < 0.0001) than its wild-type counterpart.

3.4. ELT-2, END-1, END-3, and ELT-7 have an in vitro binding affinity for the A-site as well as secondary GATA site regions

To assess *in vitro* binding of the END-1, END-3, ELT-7, and ELT-2 DNA binding domains to the *A-site* we performed EMSAs using DNA probes corresponding to the 50 bp region containing the *A-site* (corresponding to sequence −549 bp to −499 bp upstream of the endogenous *elt-2* start codon). The END-1, END-3, ELT-7, and ELT-2 DNA binding domains all showed *in vitro* binding affinity for the *A-site* containing region which could be eliminated by mutating the core motif from ACT**GATA**AG to ACT**CTGT**AG (Fig. 4). This mutation was identical to the mutations we generated to create mutant strains 1879:A and 613:A. The capability of essentially all known embryonic upstream activators of *elt-2* to bind to the *A-site* is consistent with the largely redundant roles these upstream activators play in driving *elt-2* expression. In a separate paper, we report that END-1, END-3, and ELT-7 also all bind with greatest affinity to a TGATAA sequence – essentially the core component of the *A-site* (Tracy and Rifkin; not yet published).

Interestingly, we found END-1/DNA complexes running as a doublet at all protein concentrations tested when using an A-site wild-type probe. This double banding requires the ACTGATAAG motif and can be due to END-1 fragment multimer binding on the A-site, conformational changes to DNA structure induced by END-1 binding to the A-site, or END-1 binding at the A-site potentiating binding to non-A-site sequences on the DNA probe.

3.5. Conserved secondary GATA motif clusters have weak impact on promoter-driven gene expression during early embryogenesis

The abundance of conserved HGATAR motifs within the *elt-2* upstream region initially suggested that *elt-2* activation might dependent on multiple GATA motifs in either a combinatorial or additive manner. In a combinatorial activation scenario for a focused promoter, transcription would be dependent upon one or more clusters of *cis*-regulatory motifs, and mutation of cluster motifs should result in failure to activate gene expression. In an additive scenario, each binding site or cluster of sites would contribute towards the overall rate of transcription, and mutation of

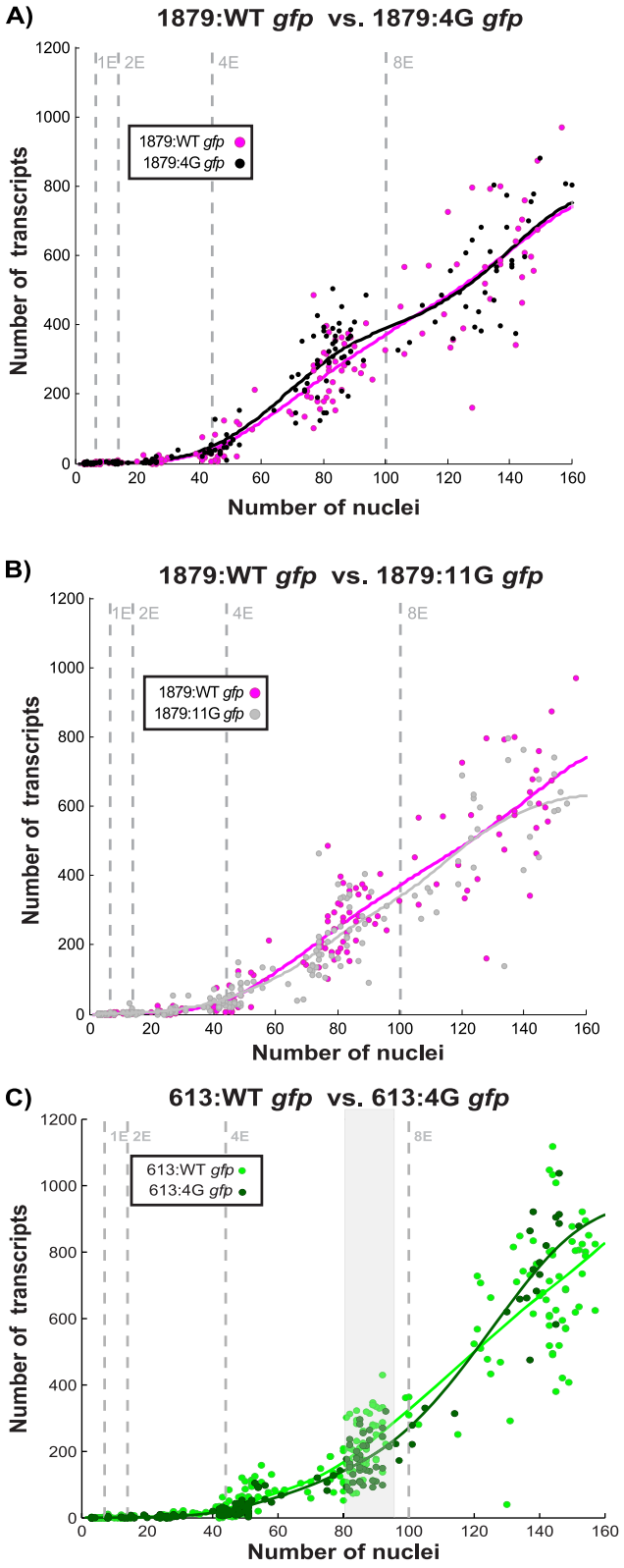


Fig. 5. Mutation of secondary 4G and 7G HGATAR motifs have minimal impact on embryonic gene expression levels or variability. (A and B) Mutation of 4G and 7G sites in a 1879 bp promoter fragment does not significantly alter embryonic expression (Fig. S7). *gfp* transcript counts shown for 1879:WT ($n=187$), 1879:4G ($n=217$), 1879:11G ($n=214$). (C) Mutation of 4G sites in a minimal 613 bp promoter does not significantly alter embryonic expression (Supplemental Fig. S8). *gfp* transcript counts shown for 613:WT ($n=459$) and 613:4G ($n=217$). The shaded gray area represents timepoints where mean expression trajectory of 613:WT is significantly different from that of 613:4G (Supplemental Fig. S7).

some cluster motifs would lower, but not necessarily eliminate, *elt-2* expression. Partial redundancy within this additive scenario would mean that more sites would need to be mutated before an effect would be apparent.

Within the 1879 bp *elt-2* upstream element we identified two clusters of HGATAR motifs hereafter "secondary GATA sites" that show conservation in both sequence and spacing in the *Elegans* supergroup. Between −400 bp and −338 bp there are four HGATAR motifs (the *4G region*) that are heavily conserved between *C. elegans* and five other members of the *Elegans* supergroup, including *C. japonica* (Fig. 1C). Between −1679 bp and −1525 bp upstream of the *elt-2* start codon there are seven HGATAR motifs (the *7G region*), with four conserved between *C. elegans* and six other members of the *Elegans* supergroup (Fig. 1C).

To determine whether these HGATAR motif cluster play a role in *elt-2* promoter-driven gene activation, we performed site directed mutagenesis on HGATAR motifs within the 4G and 7G regions while keeping the *A-site* intact (Fig. 5). Surprisingly, mutating as many as eleven HGATAR motifs across two clusters had insignificant impacts on gene expression levels and noise during embryonic development, despite the heavy conservation of motifs within these clusters (Supplemental Figs. S6 and S7, Fig. 1C). The only strain exhibiting statistically significant differences from its wild-type counterpart was 613:4G, and this was only observed as a seemingly minor temporary drop in mean expression during the 4E stage (Fig. 5C, Supplemental Fig. S7).

We confirmed that known upstream activators of *elt-2* could bind *in vitro* to the 4G region by performing an EMSA using a DNA probe corresponding to an 80 bp region encompassing the 4G region GATA sites (located −410 to −330 bp upstream of the *elt-2* start codon) (Supplemental Fig. S8). We observed distinct band shifts for ELT-7, END-3, and END-1 binding domains, confirming that *elt-2* activators can bind to regions containing secondary GATA motifs. Our results also indicate that the END-1, END-3, and ELT-7 DNA binding domains have distinct binding characteristics to the 80 bp region probed. At high protein concentrations, we observed three band shifts for ELT-7, four for END-3, and more than five for END-1. The band shift pattern for ELT-7 suggests that the ELT-7 binding domain binds to three out of the four GATA motifs within the 4G region at the protein concentrations tested. Interestingly, although five primary band shifts were visible for END-1, we clearly observed double banding on the three fastest migrating primary bands similar to the double banding observed in Fig. 4. These doublets are consistent with multimerization of the END-1 protein fragment we used on individual DNA motifs, and suggest that *in vitro* END-1 behavior seen in Fig. 4 is seen with GATA motifs in general rather than specific to the A-site motif.

These results indicate that although secondary GATA motif clusters have a minimal impact on early embryonic gene activation *in vivo*, they exhibit similar binding behavior to early embryonic *elt-2* activators *in vitro*.

3.6. Reporter GFP fluorescence in larval and adult worms reveal secondary GATA motifs modulate post-embryonic expression

To investigate whether secondary GATA site mutations had impacts on *elt-2* expression later in life, we imaged GFP fluorescence in larval and adult worms. The expression of *elt-2* persists though larval stages and adult life via autoregulation (Fukushige et al., 1999) and positive feedback from *elt-2* targets (Zhang et al., 2013). We found that only strains with an intact *A-site*, except for the super minimal promoter strain 61:WT, expressed gut GFP during larval and adult stages (Table 1). The *A-site* was indispensable for both early embryonic activation as well as a larval and adult maintenance of gene expression.

To quantify this expression we looked at L3 larvae 22 h post-

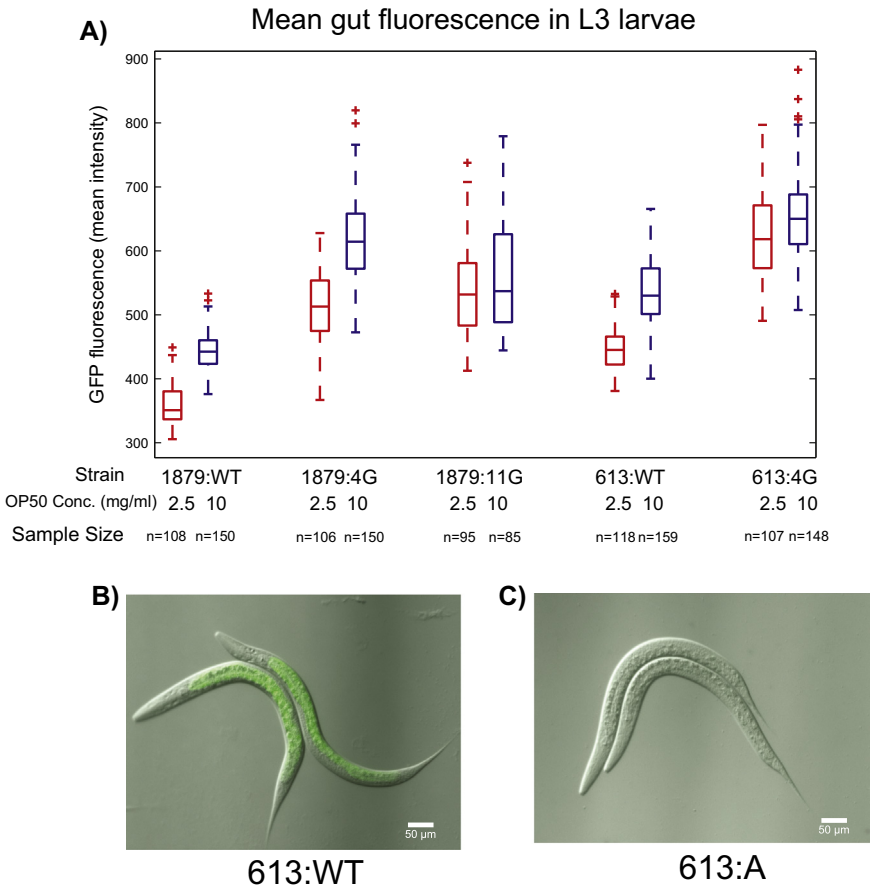


Fig. 6. Secondary HGATAR motifs play a role in modulating post-embryonicgene expression. (A) Mean gut GFP fluorescence levels for wild-type and mutant *elt-2* promoter–reporter strains 22 h post-synchronization. Red boxes (left in each pair) represent worms grown at 2.5 mg/ml OP50 liquid media and blue boxes (right in each pair) represent worms grown at 10 mg/ml OP50 liquid media. Red crosses represent outliers. (B) Gut GFP fluorescence in 613:WT. (C) Absence of gut fluorescence in 613:A mutants. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

synchronization (later stages introduced gut autofluorescence complications). We found significant differences in mean gut fluorescence between wild-type promoter–reporter strains and strains with mutations in the 4G and 7G regions (Fig. 6 and Supplemental Fig. S9). Mean fluorescence and worm size were strongly coupled with food concentrations in some strains, indicating that diet has a positive effect on *elt-2* promoter driven gene expression. In 1879:WT, 613:WT, 1879:4G, and 613:4G increasing the concentration of OP50 food from 2.5 mg/ml to 10 mg/ml significantly increased mean gut fluorescence (Supplemental Fig. S9). This response is consistent with ELT-2's role as the activator of genes involved in digestion (McGhee et al., 2009).

Mutating the 4G GATA motifs immediately downstream of the A-site in both long and short promoter contexts (1879:4G, 613:4G), resulted in significant increases in mean gut fluorescence and variability across different dietary contexts compared to wild-type counterparts (Supplemental Fig. S9). These increases demonstrate that some or all of the GATA motifs within the 4G region play a transcriptionally repressive role during post-embryonic development. The location of these sites immediately downstream of the *A-site* and TSS (Supplemental Fig. S3) suggest a steric mechanism.

We found that the apparent function of 7G region GATA motifs is dependent on context, and may be related to dietary response. Strains 1879:11G and 613:4G both lack many GATA motifs and both exhibit greater variability and less distinction between dietary regimes in post-embryonic reporter expression. In contrast to all other strains, 1879:11G did not show a statistically significant difference in mean gut fluorescence between groups raised at 10 mg/ml OP50 and groups raised at 2.5 mg/ml OP50 ($p=0.0451$; $\alpha_{SID}=0.00465$). 1879:11G also experiences a slight but significant drop in mean gut intensity at 10 mg/ml compared to 1879:4G ($p=1.78E-10$; $\alpha_{SID}=0.00465$), suggesting 7G region GATA sites may be activating under certain contexts. 613:WT, however, maintains a significant dichotomy in dietary response ($p=2.85E-36$; $\alpha_{SID}=0.00465$) despite loss of many upstream GATA motifs. These results indicate that 7G and 4G clusters (upstream and downstream of the *A-site* respectively) may not necessarily function independently of each other.

Overall, many of the 7G and 4G GATA sites modulate post-embryonic *elt-2* levels in both repressive and activating ways, dependent upon dietary context. Although many of these GATA motifs have similar sequence and conservation to the *A-site*, our functional dissection reveals that the roles of secondary motifs can be highly divergent and not necessarily related to activation. In contrast to the *A-site*, the secondary GATA motifs we examined play dispensable yet important roles in setting *elt-2* expression levels in different environmental contexts.

3.7. An ACTGATAAG motif is not sufficient for driving gene expression to near-wild-type levels

Because mutation of the *A-site* produced a disproportionate impact on gene expression levels, we hypothesized that the presence of an ACTGATAAG motif might be sufficient for driving gene expression. In the 1879:A mutant strain, we kept a second naturally occurring ACTGATAAGG motif at −1857 bp intact (Fig. 1C *green arrow*). Despite the presence of an intact ACTGATAAG sequence at

−1857 bp, this construct could not rescue reporter gene expression to wild-type promoter levels in *A-site* mutants (Figs. 2 and 3B).

To determine whether a more minimal promoter element could be generated from the *A-site*, we generated a promoter–reporter construct consisting of a 61 bp region containing the *A-site* (Table 1 *61:WT*). This construct failed to produce either embryonic *gfp* transcripts as assessed by smFISH or adult gut GFP fluorescence. This failure to drive even low-level reporter expression demonstrates that an ACTGATAAG motif alone is not sufficient for driving gene expression, and that a secondary CRE or some kind of positional information is still necessary to trigger transcription. Our results indicate that such a secondary CRE is not one of the GATA sites mutated in this study, and is unlikely to be a GATA motif.

3.8. 4G region HGATAR motifs play a role in driving low-level ectopic expression

Having identified the 613 bp *elt-2* upstream region as a minimal promoter fragment, we sought to determine if mutation of *4G region* sites would have a more obvious impact on gene expression in a minimal promoter context. Mutation of *4G region* HGATAR sites in strain 613:4G resulted in a slightly altered, but near-wild-type embryonic expression profile (Fig. 5C and Supplemental Fig. S7).

Mutation of the *A-site* along with 4G HGATAR motifs (Table 1 *613:A4G*) resulted in total loss of gene expression, effectively bringing the low-level stochastic expression observed in strain 613:A to zero for all embryos sampled (Fig. 3C). This indicates that low-level stochastic transcription resulting from mutation of the *A-site* was dependent upon one or more of the 4G HGATAR motifs.

4. Discussion

In this study, we performed the first *in vivo* single-molecule resolution investigation of the relation between CREs and transcriptional output during *C. elegans* development. We found that CREs in close proximity with similar sequences, similar conservation, and capable of binding to the same transcription factors can have wildly different functional roles in development.

Rather than multiple HGATAR sites in the *elt-2* promoter serving as redundant transcriptional activators, a single critically positioned key motif is necessary for both early END-1/END-3 driven embryonic activation as well as later ELT-2/ELT-7 driven larval and adult maintenance of gene expression. This key GATA site is used for earliest activation, later maintenance, and, surprisingly, repression of spatiotemporally ectopic early embryonic expression. Our results suggest that the critically positioned *A-site* acts as a GATA responsive “core” or “proximal” promoter element distinct from nearby conserved motifs with similar HGATAR sequences.

In contrast, the majority of GATA motifs mutated in this study appear to be dispensable for embryonic activation, but important for tuning postembryonic expression levels. We found that in wild-type promoter contexts, gene expression levels were dependent upon dietary food concentrations, with less food leading to less *elt-2* promoter activation. Furthermore, despite binding capability to upstream activators (Supplemental Fig. S8, Tracy and Rifkin, unpublished), mutation of the HGATAR motifs within the 4G region paradoxically resulted in elevated gene expression levels and variability in both long and short promoter contexts (1879:4G, 613:4G) compared to wild-type reporters. This implies that some or all of these GATA sites play repressive roles in postembryonic development. One possible explanation for this behavior is the formation of repressive or simply non-activating multimers on 4G region sites by otherwise activating GATA factors. Mammalian GATA factors have been reported to form homo- and hetero-oligmers (Chen et al., 2012), and the conservation of spacing and positioning of three of the 4G region sites within the *Elegans* supergroup suggest that orientation is important for this region of the promoter. Another possibility is that these sites may bind post-embryonically expressed non-GATA transcription factors (*see below*).

Although the extended DNA motif sequence of each individual GATA CRE may contribute to establishing the distinctive roles of the dominant primary *A-site* versus secondary auxiliary GATA CREs, we found that the extended sequence motif does not completely explain the relative importance of the *A-site.* The *A-site's* extended sequence is ACTGATAAGA, but mutation of the key *A-site* motif at −527 bp could not be rescued to near-wild-type levels by any of 9 remaining TGATAA or 17 remaining HGATAR motifs present in strain 1879:A (Figs. 2, 3, S2, and S4). Furthermore, one of the alternate TGATAA motifs is a very similar ACTGATAAGG located at −1857 bp that remains intact in strain 1879:A (Fig. 1C; *green arrow*). However, this similar sequence is insufficient to activate promoter–reporter gene expression. The presence of an *A-site* is also not sufficient for driving even low-level expression in a 61 bp minimal promoter (Table 1 *61:WT*) context suggesting that more *cis*-information may be needed for gene activation. This additional information could come in the form of a second motif, the local chromatin context, or even the relative abundance of T-rich sequence (Grishkevich et al., 2011).

Despite this insufficiency, evidence from a wide array of studies indicate that the ACTGATAAG sequence can act as a binding site for multiple gut-associated transcription factors, some of which are not even GATA factors. Several studies previously demonstrated that similar or identical sequences are present and functionally necessary in the promoters of *elt-2* regulated gut genes (Egan et al., 1995; Fukushige et al., 2005; MacMorris et al., 1992', 1994; McGhee et al., 2009). Furthermore, previous research into the FOXO transcription factor DAF-16 revealed that the promoters of DAF-16 targets are enriched for the sequence TGATAAG (also known as “Daf-16 associated element” or DAE) and that DAF-16 can bind to the TGATAAG motif *in vitro* (Murphy et al., 2003; Zhang et al., 2013). A recent study, Mueller et al. (2014), also showed that TGATAAG sequences, particularly the specific sequence ACTGATAAGA, are heavily over-represented in the promoters of genes that are upregulated in response to ultraviolet light exposure, and downregulated in response to starvation. That study implicated the GATA factor EGL-27 as the primary effector of UV stress-responsive gene activation.

These studies combined with our results suggest that in the worm's adult life, regulatory signals from different pathways may converge on and compete for access of a single dominant primary *cis*-regulatory motif in the *elt-2* promoter, effectively forming a regulatory information bottleneck. Such a bottleneck can be useful in setting an upper limit on the level of transcriptional activation that can occur while simultaneously allowing responsiveness to signals related to processes such as aging (Zhang et al., 2013), stress (Schieber and Chandel, 2014), and disease resistance (Head and Aballay, 2014). Once the primary activation site is saturated with binding factors, further activation may be impeded.

The wide affinity of GATA-family motifs combined with the large number of GATA related transcription factors may partially explain our observations of low-level stochastic ectopic transcription in *A-site* mutants in both E cells and non-E-cells. The absence of transcriptional activators such as *med-1/2, end-3,* or *end-1* in non-E-cells implies that there may be other GATA-binding transcriptional activators present outside of the nascent endoderm during early embryogenesis (Budovskaya et al., 2008; Gilleard et al., 1999; Mueller et al., 2014; Murphy et al., 2003). We found that in a minimal 613 bp promoter context, secondary auxiliary HGATAR motif(s) in the *4G region* are necessary for producing low-level stochastic transcription (Table 1 613:A4G; Fig. 3D). Including

more proximal regions of the *elt-2* promoter (−1879 bp to −613 bp; contains 13 HGATAR motifs) in an *A-site* mutant background increases low-level transcription levels even further. Our analysis of deep sequencing data from Saito et al. (2013) revealed small numbers of TSS reads overlying proximal promoter regions including the *7G* and *4G* regions, suggesting many ectopic transcripts may originate from secondary GATA sites (Supplemental Fig. S3) (Saito et al., 2013).

These non-specific mini-activation events reveal a layer of transcriptional regulation not previously observed. For multicellular organisms, off target activation events can result in ectopic expression that can have negative consequences for cell fate specification, particularly if the gene being ectopically expressed is an autoregulatory transcription factor. Here we observe that an intact *A-site* plays a role in suppressing this ectopic expression in non-E-cells. The same site used for activation in the endoderm is used for repression in non-endoderm fated cells.

5. Conclusions

Our study shows that *C. elegans* can serve as a powerful model for single-molecule cell-type-specific *cis*-regulatory studies. We determined that a single key *cis*-regulatory site-in the midst of a host of similarly conserved sites-is used for the earliest activation and post-embryonic maintenance of the expression of an essential regulatory gene, and that this same CRE helps repress early stochastic expression in non-E cells. We identified diverse roles for secondary sites in tuning postembryonic gene expression. The wildly different roles served by CREs examined in this study illustrate the diverse functions that similar CREs can take on and reveal a distinction between early and post-embryonic *elt-2* activation and function.

Acknowledgments

Special thanks to George Kassavetis and Jim Kadonaga for help with *in vitro* assays, Jim McGhee for *elt-2* plasmids and discussion of the *elt-2* promoter, Emily Troemel for microscope time, and Allison Wu for help in generating smFISH probes and initial processing of the data. Some strains were provided by the Caenorhabditis Genetics Center, which is funded by NIH Office of Research Infrastructure Programs (P40 OD010440). This work was supported in part by US Department of Education Graduate Assistance in Areas of National Need (GAANN) Grant P200A100159, NIH Predoctoral Training Grant T32 GM008666, and NIH/NIGMS Grant R01GM103782 to SAR.

Appendix A. Supplementary material

Supplementary data associated with this article can be found in the online version at http://dx.doi.org/10.1016/j.ydbio.2016.02.013.

References

Boeck, M.E., Boyle, T., Bao, Z., Murray, J., Mericle, B., Waterston, R., 2011. Specific roles for the GATA transcription factors end-1 and end-3 during *C. elegans* E-lineage development. Dev. Biol. 358 (2), 345–355. http://dx.doi.org/10.1016/j.ydbio.2011.08.002.

Budovskaya, Y.V., Wu, K., Southworth, L.K., Jiang, M., Tedesco, P., Johnson, T.E., Kim, S.K., 2008. An elt-3/elt-5/elt-6 GATA transcription circuit guides aging in C. *elegans*. Cell 134 (2), 291–303. http://dx.doi.org/10.1016/j.cell.2008.05.044.

Chen, Y., Bates, D.L., Dey, R., Chen, P., Machado, A.C.D., Laird-Offringa, I.A., Rohs, R., Chen, L., 2012. DNA binding by GATA transcription factor suggests mechanisms of DNA looping and long-range gene regulation. Cell Reports 2 (5), 1197–1206. http://dx.doi.org/10.1016/j.celrep.2012.10.012.

Davidson, E.H., 2001. Genomic Regulatory Systems: Development and Evolution. Academic Press, San Diego.

Egan, C.R., Chung, M.A., Allen, F.L., Heschl, M.F.P., Van Buskirk, C.L., 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 (2), 397–419. http://dx.doi.org/10.1371/journal.pone.0094723.

Félix, M.-A., Braendle, C., Cutter, A.D., 2014. A streamlined system for species diagnosis in Caenorhabditis (Nematoda: Rhabditidae) with name designations for 15 distinct biological species. PLoS One 9 (4), e94723. http://dx.doi.org/10.1371/journal.pone.0094723.

Flores, G.V., Duan, H., Yan, H., Nagaraj, R., Fu, W., Zou, Y., Banerjee, U., 2000. Combinatorial signaling in the specification of unique cell fates. Cell 103 (1), 75–85. http://dx.doi.org/10.1016/S0092-8674(00)00106-9.

Frøkjær-Jensen, C., Davis, M.W., Ailion, M., Jorgensen, E.M., 2012. Improved Mos1-mediated transgenesis in *C. elegans*. Nat. Methods 9 (2), 117–118. http://dx.doi.org/10.1038/nmeth.1865.

Frøkjaer-Jensen, C., Davis, M.W., Hopkins, C.E., Newman, B.J., Thummel, J.M., Olesen, S.-P., Jorgensen, E.M., 2008. Single-copy insertion of transgenes in *Caenorhabditis elegans*. Nat. Genet. 40 (11), 1375–1383. http://dx.doi.org/10.1038/ng.248.

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

Fukushige, T., Hendzel, M.J., Bazett-Jones, D.P., 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 (21), 11883–11888. http://dx.doi.org/10.1073/pnas.96.21.11883.

Fukushige, Tetsunari, Goszczynski, B., Yan, J., McGhee, J.D., 2005. Transcriptional control and patterning of the pho-1 gene, an essential acid phosphatase expressed in the *C. elegans* intestine. Dev. Biol. 279 (2), 446–461. http://dx.doi.org/10.1016/j.ydbio.2004.12.012.

Gerstein, M.B., Lu, Z.J., Van Nostrand, E.L., Cheng, C., Arshinoff, B.I., Liu, T., Ikegami, K., 2010. Integrative analysis of the *Caenorhabditis elegans* genome by the modENCODE Project. Science 330 (6012), 1775–1787. http://dx.doi.org/10.1126/science.1196914.

Gilleard, J.S., Shafi, Y., Barry, J.D., McGhee, J.D., 1999. ELT-3: a *Caenorhabditis elegans* GATA factor expressed in the embryonic epidermis during morphogenesis. Dev. Biol. 208 (2), 265–280. http://dx.doi.org/10.1006/dbio.1999.9202.

Grishkevich, V., Hashimshony, T., Yanai, I., 2011. Core promoter T-blocks correlate with gene expression levels in *C. elegans*. Genome Res. 21 (5), 707–717. http://dx.doi.org/10.1101/gr.113381.110.

Harris, T.W., Antoshechkin, I., Bieri, T., Blasiar, D., Chan, J., Chen, W.J., Sternberg, P. W., 2010. WormBase: a comprehensive resource for nematode research. Nucleic Acids Res. 38 (Database issue), D463–D467. http://dx.doi.org/10.1093/nar/gkp952.

Head, B., Aballay, A., 2014. Recovery from an acute infection in *C. elegans* requires the GATA transcription factor ELT-2. PLoS Genet. 10 (10), e1004609. http://dx.doi.org/10.1371/journal.pgen.1004609.

Huang, R.-E., Ren, X., Qiu, Y., Zhao, Z., 2014. Description of *Caenorhabditis sinica* sp. n. (Nematoda: Rhabditidae), a nematode species used in comparative biology for *C. elegans*. PLoS One 9 (11), e110957. http://dx.doi.org/10.1371/journal.pone.0110957.

Juven-Gershon, T., Kadonaga, J.T., 2010. Regulation of gene expression via the core promoter and the basal transcriptional machinery. Dev. Biol. 339 (2), 225–229. http://dx.doi.org/10.1016/j.ydbio.2009.08.009.

Lowry, J.A., Gamsjaeger, R., Thong, S.Y., Hung, W., Kwan, A.H., Broitman-Maduro, G., Mackay, J.P., 2009. Structural analysis of MED-1 reveals unexpected diversity in the mechanism of DNA recognition by GATA-type zinc finger domains. J. Biol. Chem. 284 (9), 5827–5835. http://dx.doi.org/10.1074/jbc.M808712200.

MacMorris, M., Broverman, S., Greenspoon, S., Lea, K., Madej, C., Blumenthal, T., Spieth, J., 1992. Regulation of vitellogenin gene expression in transgenic *Caenorhabditis elegans*: short sequences required for activation of the vit-2 promoter. Mol. Cell. Biol. 12 (4), 1652–1662.

MacMorris, M., Spieth, J., Madej, C., Lea, K., Blumenthal, T., 1994. Analysis of the VPE sequences in the *Caenorhabditis elegans* vit-2 promoter with extrachromosomal tandem array-containing transgenic strains. Mol. Cell. Biol. 14 (1), 484–491.

Maduro, M.F., Broitman-Maduro, G., Mengarelli, I., Rothman, J.H., 2007. Maternal deployment of the embryonic SKN-1− > MED-1,2 cell specification pathway in *C. elegans*. Dev. Biol. 301 (2), 590–601. http://dx.doi.org/10.1016/j.ydbio.2006.08.029.

Maduro, M.F., Hill, R.J., Heid, P.J., Newman-Smith, E.D., Zhu, J., Priess, J.R., Rothman, J. H., 2005. Genetic redundancy in endoderm specification within the genus Caenorhabditis. Dev. Biol. 284 (2), 509–522. http://dx.doi.org/10.1016/j.ydbio.2005.05.016.

Maduro, M.F., Meneghini, M.D., Bowerman, B., Broitman-Maduro, G., Rothman, J.H., 2001. Restriction of mesendoderm to a single blastomere by the combined action of SKN-1 and a GSK-3β homolog is mediated by MED-1 and -2 in *C. elegans*. Mol. Cell 7 (3), 475–485. http://dx.doi.org/10.1016/S1097-2765(01)00195-2.

McGhee, J.D., Fukushige, T., Krause, M.W., Minnema, S.E., Goszczynski, B., Gaudet, J., Kalb, J.M., 2009. ELT-2 is the predominant transcription factor controlling differentiation and function of the *C. elegans* Intestine, from embryo to adult. Dev. Biol. 327 (2), 551–565. http://dx.doi.org/10.1016/j.ydbio.2008.11.034.

Mirny, L.A., Wunderlich, Z., 2008. Fundamentally different strategies for transcriptional regulation are revealed by information-theoretical analysis of binding motifs. Trends Genet. 25 (10), 4.

Mueller, M.M., Castells-Roca, L., Babu, V., Ermolaeva, M.A., Müller, R.-U., Frommolt, P., Schumacher, B., 2014. DAF-16/FOXO and EGL-27/GATA promote developmental growth in response to persistent somatic DNA damage. Nat. Cell. Biol. 16 (12), 1168–1179. http://dx.doi.org/10.1038/ncb3071.

Murphy, C.T., McCarroll, S.A., Bargmann, C.I., Fraser, A., Kamath, R.S., Ahringer, J., Kenyon, C., 2003. Genes that act downstream of DAF-16 to influence the lifespan of *Caenorhabditis elegans*. Nature 424 (6946), 277–283. http://dx.doi.org/10.1038/nature01789.

Nair, G., Walton, T., Murray, J.I., Raj, A., 2013. Gene transcription is coordinated with, but not dependent on, cell divisions during *C. elegans* embryonic fate specification. Development 140 (16), 3385–3394. http://dx.doi.org/10.1242/dev.098012.

Raj, A., Rifkin, S.A., Andersen, E., van Oudenaarden, A., 2010. Variability in gene expression underlies incomplete penetrance. Nature 463 (7283), 913–918. http://dx.doi.org/10.1038/nature08781.

Raj, A., van den Bogaard, P., Rifkin, S.A., van Oudenaarden, A., Tyagi, S., 2008. Imaging individual mRNA molecules using multiple singly labeled probes. Nat. Methods 5 (10), 877–879. http://dx.doi.org/10.1038/nmeth.1253.

Rifkin, S.A., 2011. Identifying fluorescently labeled single molecules in image stacks using machine learning. Methods Mol. Biol. 772, 329–348. http://dx.doi.org/10.1007/978-1-61779-228-1_20.

Saito, T.L., Hashimoto, S., Gu, S.G., Morton, J.J., Stadler, M., Blumenthal, T., Morishita, S., 2013. The transcription start site landscape of *C. elegans*. Genome Res. 23 (8), 1348–1361. http://dx.doi.org/10.1101/gr.151571.112.

Schieber, M., Chandel, N.S., 2014. TOR signaling couples oxygen sensing to lifespan in *C. elegans*. Cell Rep. 9 (1), 9–15. http://dx.doi.org/10.1016/j.celrep.2014.08.075.

Sommermann, E.M., Strohmaier, K.R., Maduro, M.F., 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 (1), 154–166. http://dx.doi.org/10.1016/j.ydbio.2010.08.020.

Ury, H.K., 1976. A comparison of four procedures for multiple comparisons among means (Pairwise Contrasts) for arbitrary sample sizes. Technometrics 18 (1), 89–97. http://dx.doi.org/10.2307/1267921.

Wu, A.C., Rifkin, S.A., 2015. Aro: a machine learning approach to identifying single molecules and estimating classification error in fluorescence microscopy images. BMC Bioinforma. 16 (1), 102. http://dx.doi.org/10.1186/s12859-015-0534-z.

Zhang, P., Judy, M., Lee, S.-J., Kenyon, C., 2013. Direct and indirect gene regulation by a life-extending FOXO protein in *C. elegans*: roles for GATA factors and lipid gene regulators. Cell Metab. 17 (1), 85–100. http://dx.doi.org/10.1016/j.cmet.2012.12.013.

Zhu, J., Fukushige, T., McGhee, J.D., Rothman, J.H., 1998. Reprogramming of early embryonic blastomeres into endodermal progenitors by a *Caenorhabditis elegans* GATA factor. Genes Dev. 12 (24), 3809–3814.

Zhu, J., Hill, R.J., Heid, P.J., Fukuyama, M., Sugimoto, A., Priess, J.R., Rothman, J.H., 1997. end−1 encodes an apparent GATA factor that specifies the endoderm precursor in *Caenorhabditis elegans* embryos. Genes Dev. 11 (21), 2883–2896.

SUPPLEMENTAL INFORMATION AND METHODS

Mutagenesis of GATA motifs controlling the endoderm regulator *elt-2* reveals distinct dominant and secondary *cis*-regulatory elements

Authors: Lawrence Du[1], Sharon Tracy[1], Scott A. Rifkin[1*]

1. Biological Sciences, University of California, San Diego 9500 Gilman Dr. La Jolla, CA 92093

*Corresponding Author: sarifkin@ucsd.edu

Supplemental Methods

In vitro binding assays detailed methods

For isolating His6 tagged fusion constructs, transformed Rosetta 2(DE3) pLysS cells were grown at 37 °C in LB supplemented with 50 ug/ml Kanamycin, 34 ug/ml Chloramphenicol, and 20 uM $ZnCl_2$ to an OD of 0.5. Cultures were cooled to 18°C and expression was induced with 0.1 mM IPTG at 18°C for 12 hours. Cells were pelleted, washed in ice cold 1X PBS, and resuspended in GATA lysis buffer (10mM HEPES, 200mM NH_4OAc, 300 mM NaCl, 10mM Imidazole, 0.5 uM ZnOAc, 10 % Glycerol, 5 mM, β-mercaptoethanol, 1mM PMSF, 1 tablet/10 ml EDTA-free protease inhibitor cocktail (Roche 11-836-170-001)) and sonicated. Lysates were clarified at 14,000 g for 20 minutes at 4 °C. His6 tagged proteins were bound to Talon metal affinity beads (Clontech #635501), washed once in GATA lysis buffer, washed a second time in modified GATA lysis buffer (100 mM NaCl/20 mM Imidazole), and eluted with modified GATA lysis buffer (100 mM NaCl/200 mM Imidazole). Gycerol was adjusted to 20% and purified proteins were aliquoted, snap frozen in liquid nitrogen, and stored at -80 °C. Binding assays were done at room temperature for 30 minutes in EMSA buffer (10mM HEPES, 200mM NH_4 OAc, 30 mM NaCl, 1.5mM $MgCl_2$, 0.5 uM $Zn(OAc)_2$, 0.2 mg/ml BSA , 1mM DTT, 8% glycerol, pH 7.0). Complexes were run on 8% 29:1 Acrylamide/Bis-acrylamide 8% glycerol gel in 0.5X TBE and stained with Sybr Green Nucleic acid gel stain (Molecular Probes S-7563).

The following DNA binding domains were used for in vitro assays (measured from start codon):
END-1: amino acids 114-221
END-3: amino acids 156-242
ELT-7: amino acids 118-198
ELT-2: amino acids 129-336

For Figure 4, the following probe sequences were used:
WT probe sequence:
AGCGACAGAGGTCGGGGCTGAA**ACT**GATA**AG**AATAGTCGACACTAACGCC
mutant probe sequence:
AGCGACAGAGGTCGGGGCTGAA**ACT**CTGT**AG**AATAGTCGACACTAACGCC

For Supplemental Figure S8, the following probe sequence was used:
ATTTTCTTTTT**T**GATA**A**AATCAGC**C**TATC**T**ATACTTCCCAATCATTTTTAGTC**T**T
ATC**G**TTGAACAG**C**TATC**G**AGGTGCCA

Quantifying Early Embryonic Ectopic Expression Estimates

Due to the physical flattening of samples prior to imaging, the early staging of most of our image data, and the absence of cell-specific membrane markers, we had to develop a heuristic method to demarcate the limits of E-cells and proto-E-cells and quantify ectopic expression. We settled upon a method using two-dimensional alpha shapes to roughly outline the limits of aggregate *end-1* and *elt-2* localization, and then counting the number of *gfp* transcripts that fell outside of the generated alpha shape. We adapted a custom MATLAB script for generating alpha shapes, (Lundgren 2010) and used a radius parameter of 75 pixels to produce concavity.

For statistical analysis of ectopic expression estimates, we binned embryos at the 2E stage (14-44 nuclei) and used a permutation test to determine if specified datasets were significantly different from each other.

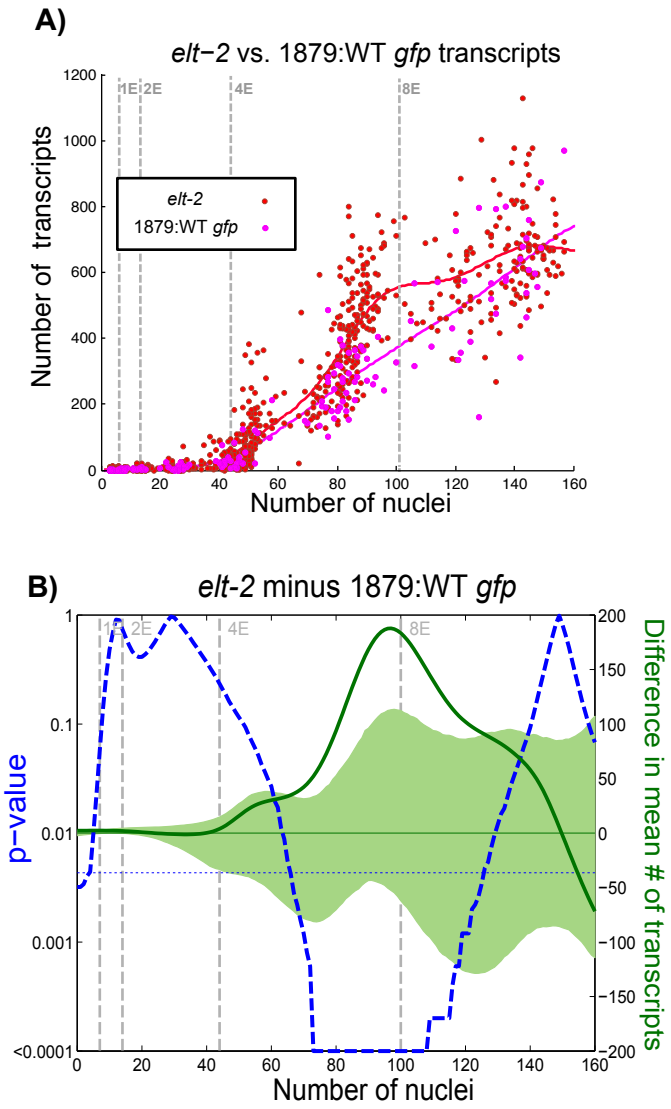
Primary Strain ID	**Alternate Strain ID**	**Promoter Description**	**Mutation coordinates (bp from endogenous *elt-2* start codon)**
1879:WT	**ELD005-1**	1879 bp wild-type	N/A
1879:4G	**ELD006-2**	1879 bp with 4 HGATAR>HCTTAR mutations	-400, -387, -358, -344
1879:11G	**ELD012-1**	1879 bp with 11 HGATAR>HCTTAR mutations	-1679, 1659, -1636, -1620, -1570, -1550, -1531,-400, -387, -358, -344
422:WT	**ELD016-1**	422 bp promoter wild-type	N/A
613:WT	**ELD018-3**	613 bp promoter wild-type	N/A
613:A	**ELD023-1**	613 bp promoter w/ ACTGATAAG>ACTCTGTAG mutation at -527 bp	-527
613:A4G	**ELD024-1**	613 bp promoter w/ ACTGATAAG>ACTCTGTAG mutation at -527 bp and 4 HGATAR>HCTTAR mutations	-527, -400, -387, -358, -344
613:4G	**ELD025-2**	613 bp promoter w/ 4 HGATAR>HCTTAR mutations	-400, -387, -358, -344
1879:A	**ELD026-2**	1879 bp promoter w/ ACTGATAAG>ACTCTGTAG mutation at -527 bp	-527
61:WT	**ELD027-3**	61 bp promoter w/ wild-type ACTGATAAG site at -527 bp	N/A (fragment from constructed from -539 to -478)

Supplemental Table 1. Specific strain mutations.

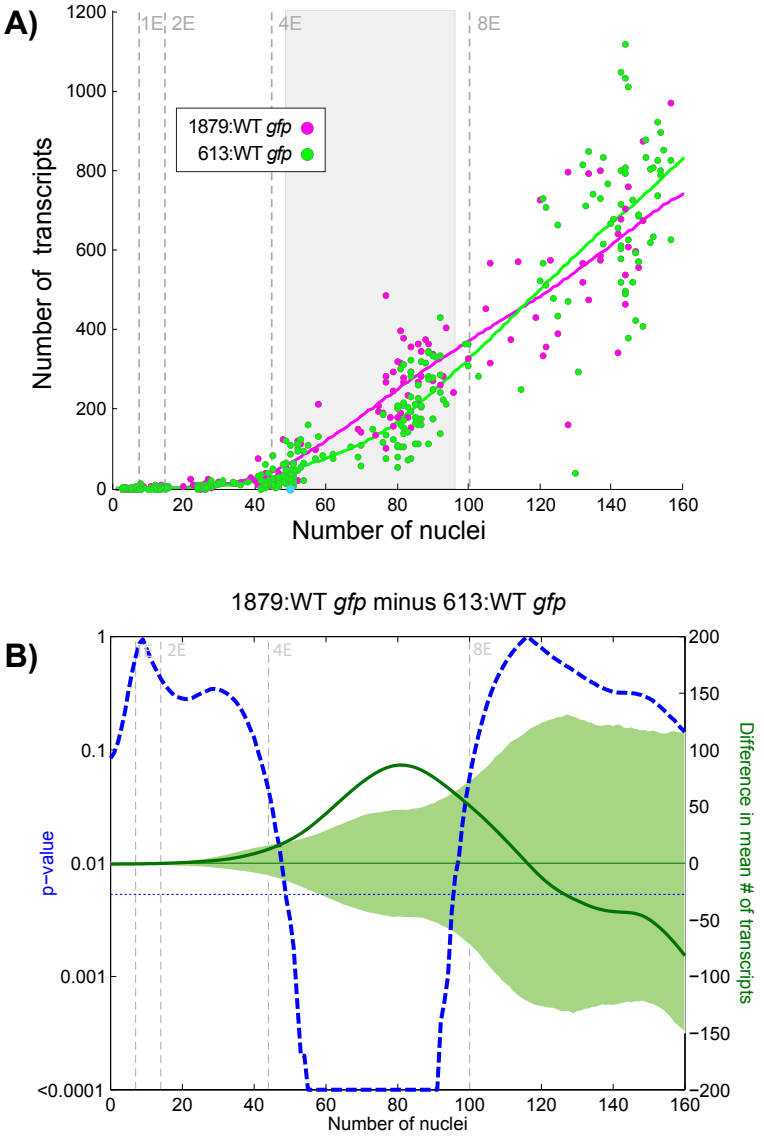
Supplemental References

Lundgren, J. 2010. Alpha Shapes (http://www.mathworks.com/matlabcentral/fileexchange/28851-alpha-shapes). *MATLAB Central File Exchange.* Accessed 9.15.15.

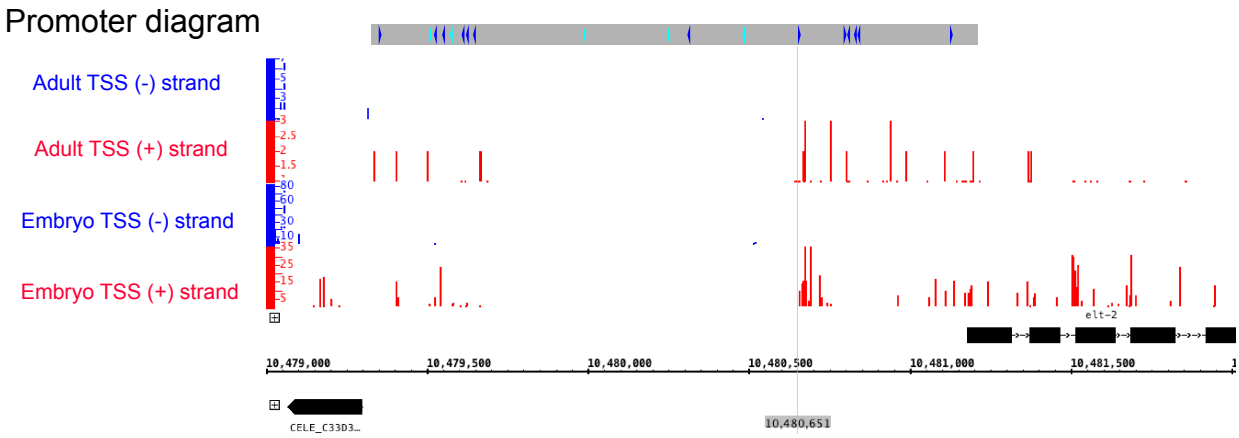
Nicol JW, Helt GA, Blanchard SG, Raja A, Loraine AE. 2009. The Integrated Genome Browser: free software for distribution and exploration of genome-scale datasets. *Bioinformatics* **25**: 2730–2731.



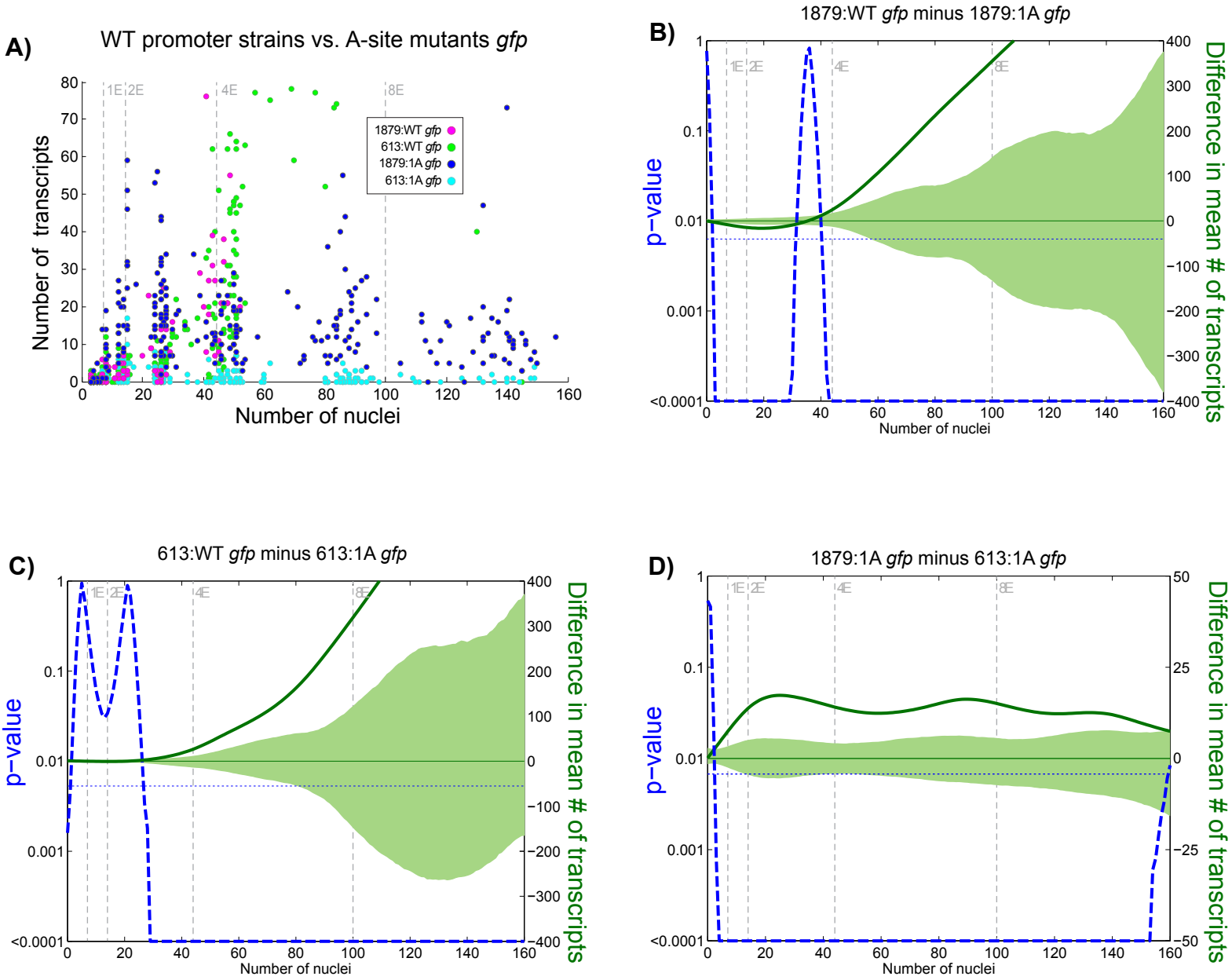
Supplemental Figure S1. Strain 1879:WT *gfp* reporter expression (n=187) compared to *elt-2* expression pooled from multiple experiments (n=1320). A) *elt-2* transcript count data collected from strains 613:WT, 1879:A, 613:A, 613:4G, and show comparable levels to wild-type *elt-2* levels observed in previous studies. Lines represent mean trajectory of aggregated *elt-2* data compared to 1879:WT. B) Permutation test. Dark green line: difference in the mean trajectories (right axis). The light green region represents the envelope of the null distribution with the Dunn-Šidák adjustment (see Methods). Trajectories are significantly different at timepoints where the green line is outside the light green region. The horizontal, fluctuating line tracks the p-value at each timepoint (log10 scale, left axis). There are significant differences between trajectories where this line dips below the dashed blue line.



Supplemental Figure S2. (A) *gfp* transcript counts for wild-type strains 1879:WT (n=187) and 613:WT (n=459). (B) Significance test of the difference between these strains.

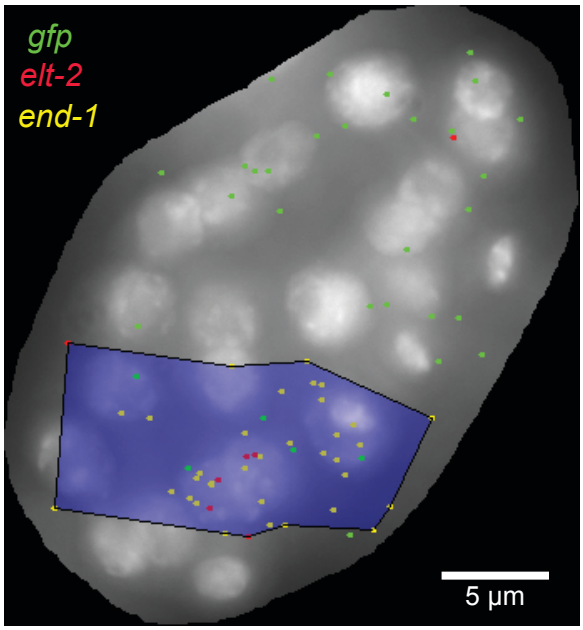


Supplemental Figure S3. Transcriptional start site (TSS) data from Saito *et al.* 2013. Promoter diagram scaled to TSS data. Gray line indicates position of ACTGATAAGA *A-site* motif. Some peaks may represent 5' ends of degraded RNAs or abortive transcripts. Plot generated using IGB (Nicol *et al.* 2009).

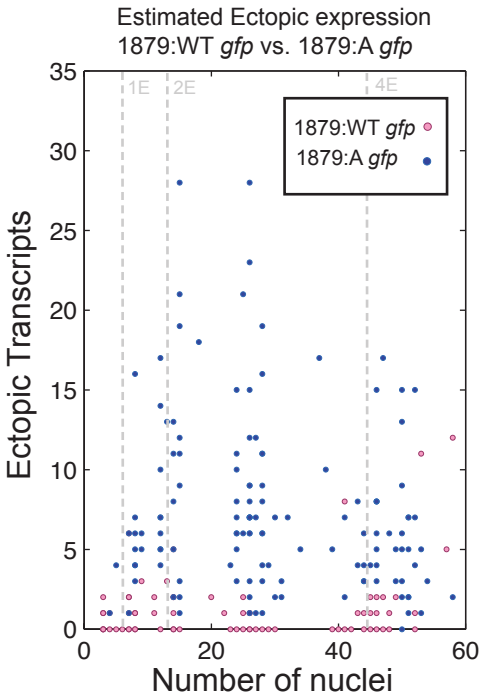


Supplemental Figure S4. (A) gfp transcript counts for A-site mutants 1879:A (n=253) and 613:A (n=217) compared to each other and to wild-type strains 1879:WT (n=187) and 613:WT(n=459). (B-C) Significance tests of the effect of the A-site mutation. D) Significance test of the effect of shortening the promoter on transcription in the A-site mutants.

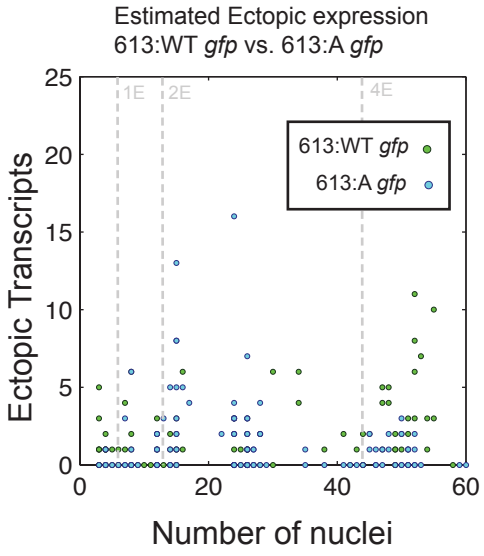
A)



B)



C)

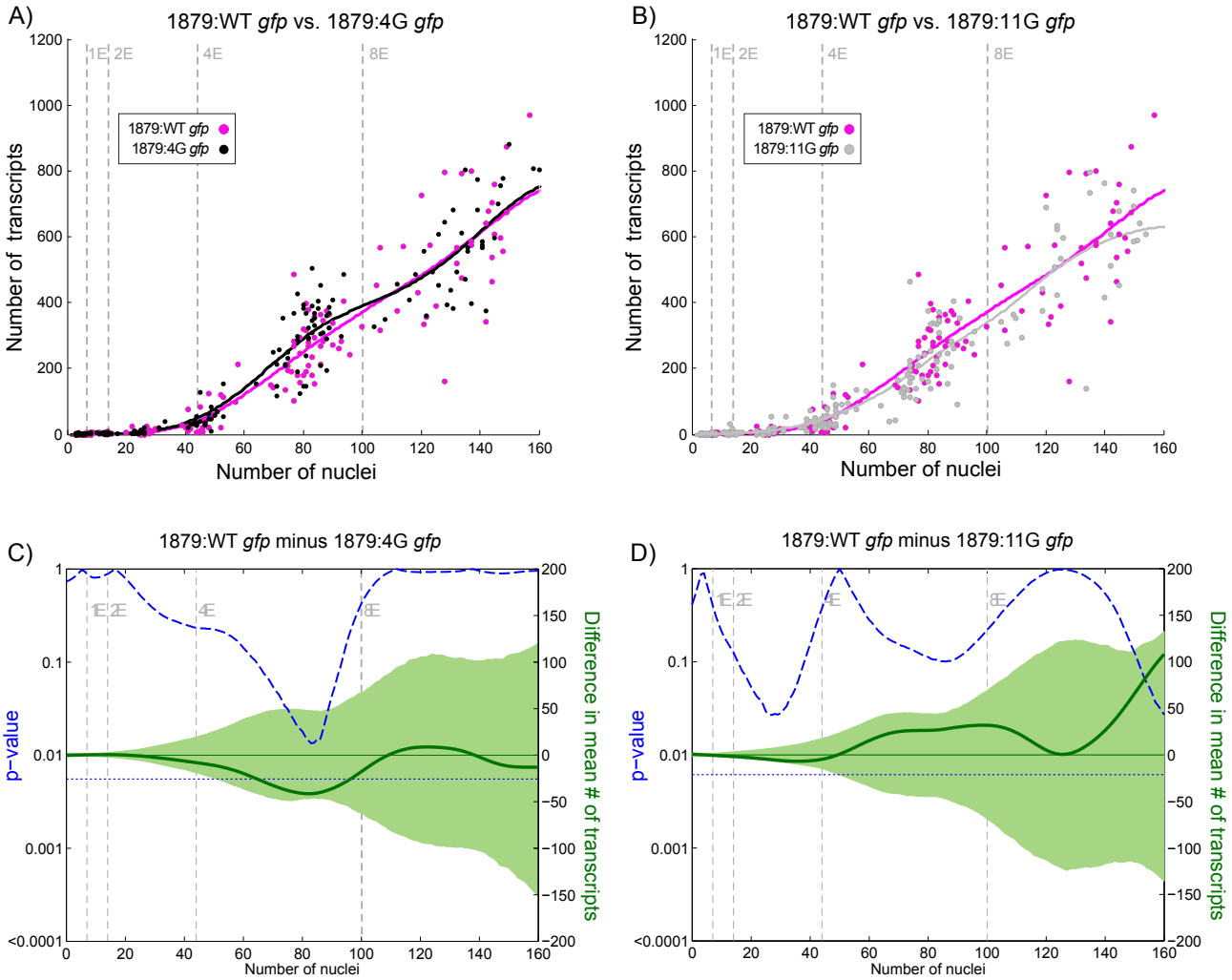


D)

A	B	Binned 2E stage p-value (Reject A=B)
Ectopic 1879:A (n=75)	Ectopic 1879:WT (n=39)	<0.0001
Ectopic 613:A (n=70)	Ectopic 613:WT (n=57)	0.157
Ectopic 613:WT (n=57)	Ectopic 1879:WT (n=39)	0.256
Ectopic 1879:A (n=75)	Ectopic 613:A (n=70)	<0.0001

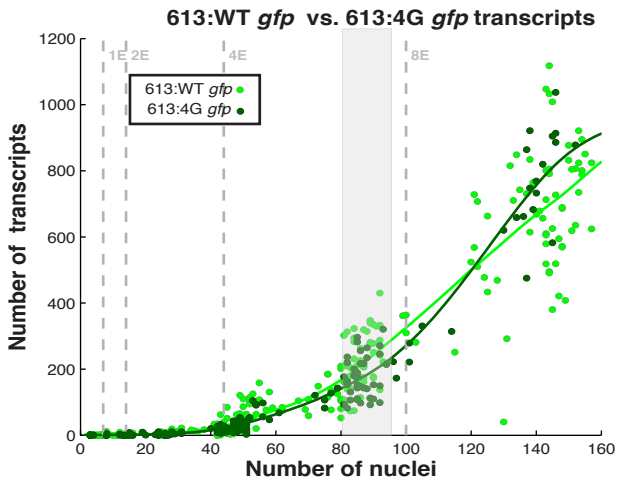
Šidák corrected alpha (α = 0.05) = 0.0127

Supplemental Figure 5. Ectopic expression estimates in *A-site* mutants. A) Example of alpha shape construction. *gfp* spots outside of the shape are counted as ectopic. B) Estimated ectopic 1879:A (n=132) versus estimated ectopic 1879:WT (n=72). C) Estimated ectopic 613:A (n=104) versus estimated ectopic 613:WT (n=105). D) Results of one-sided permutation test performed on binned 2E stage ectopic expression data (2000 repetitions per test).

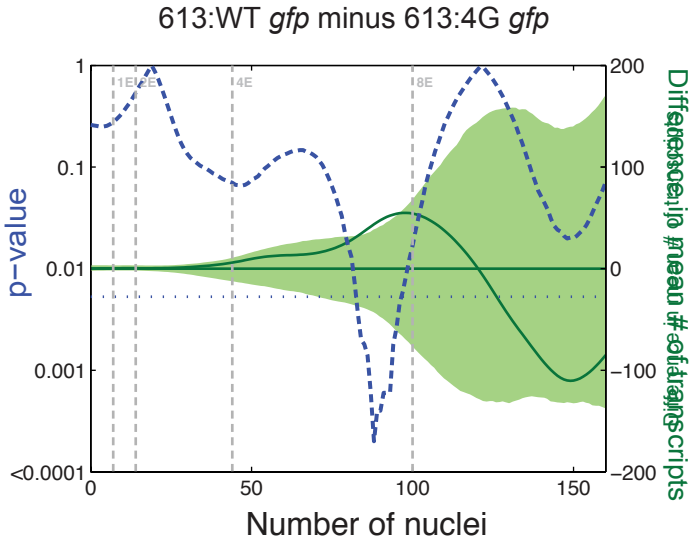


Supplemental Figure 6. Mutation of 4G and 7G HGATAR motifs have minimal impact on gene expression levels or variability. (A-B) Comparison of gfp transcript counts in strains with wild-type and mutant versions of the 1879 bp *elt-2* promoter. Lines represent smoothing splines of the data. (C-D) There are no significant differences between these strains and the wild-type promoter.

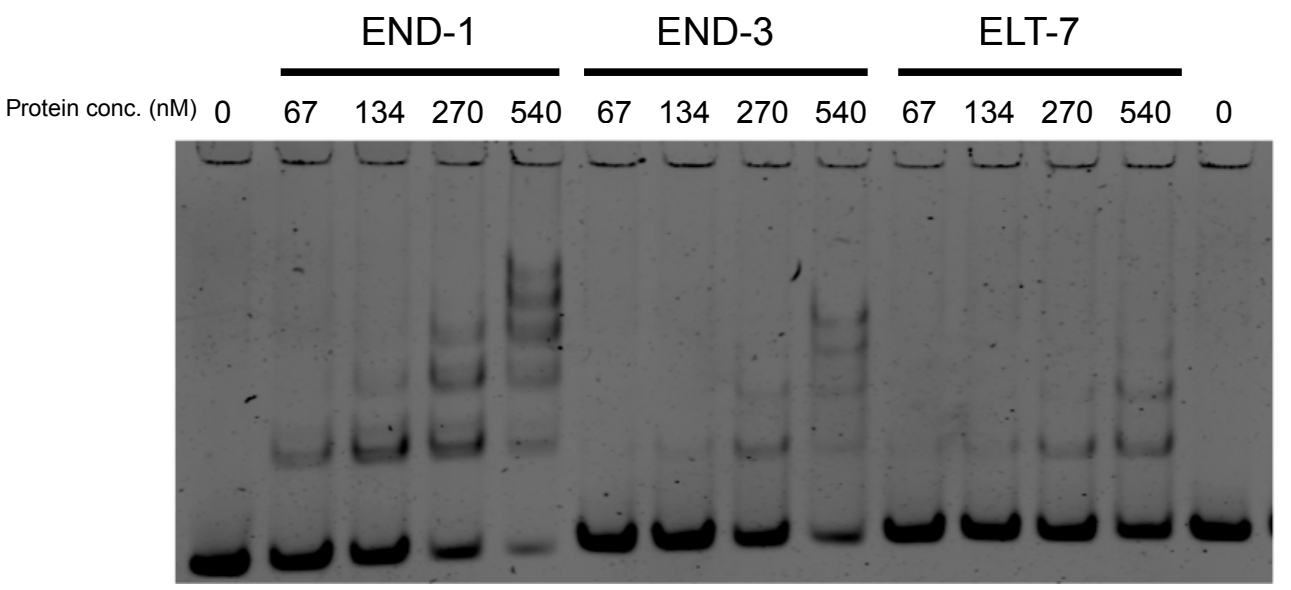
A)



B)



Supplemental Figure S7. (A) *gfp* transcript counts for wild-type strain 613:WT (n=459) versus 613:4G (n=154). Shaded grey area represents timepoints where mean expression trajectory of 613:WT is significantly different from that of 613:4G. (B) Significance test of the effect of the 4G mutations in the shorter promoter context.



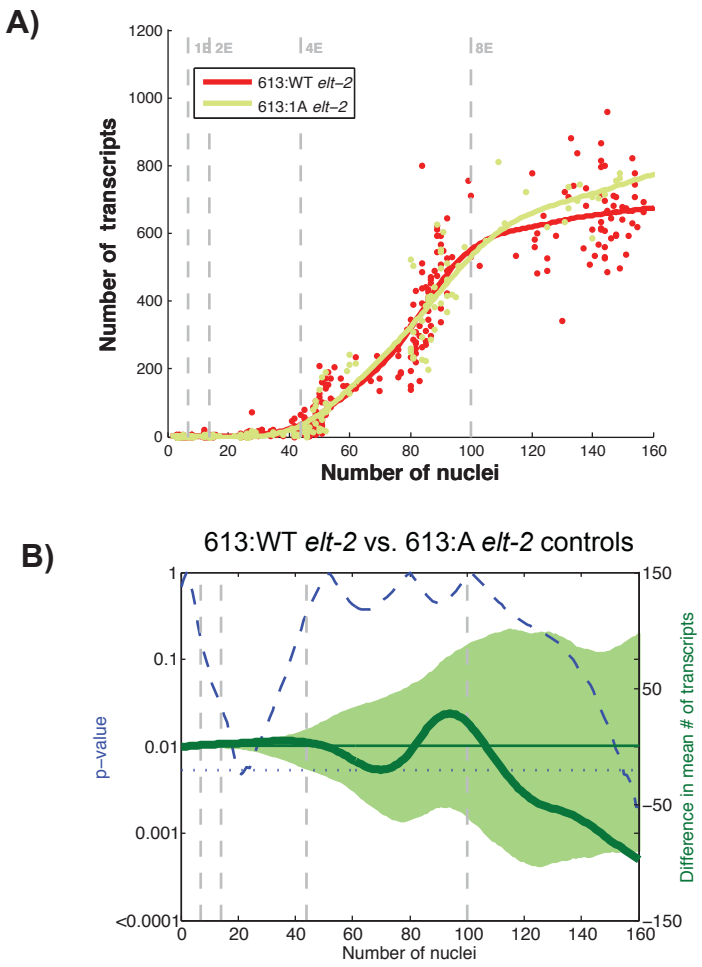
Supplemental Figure S8. EMSA of 80 bp DNA probe corresponding to 4G region. DNA probe concentration in all lanes is 70 nM.

Two-sample t-test for mean gut fluorescence data

A	B	p-value (Reject A=B)
1879:WT (2.5 mg/ml) (n=108)	1879:WT (10 mg/ml) (n=150)	6.64E-62
1879:4G (2.5 mg/ml) (n=106)	1879:4G (10 mg/ml) (n=150)	1.24E-34
1879:11G (2.5 mg/ml) (n=95)	1879:11G (10 mg/ml) (n=85)	0.0451
613:WT (2.5 mg/ml) (n=118)	613:WT (10 mg/ml) (n=159)	2.85E-36
613:4G (2.5 mg/ml) (n=107)	613:4G (10 mg/ml) (n=148)	2.22E-04
1879:4G (2.5 mg/ml) (n=106)	1879:11G (2.5 mg/ml) (n=95)	0.0129
1879:4G (10 mg/ml) (n=150)	1879:11G (10 mg/ml) (n=95)	1.78E-10
1879:11G (2.5 mg/ml) (n=95)	613:4G (2.5 mg/ml) (n=107)	2.32E-18
1879:11G (10 mg/ml) (n=85)	613:4G (10 mg/ml) (n=148)	1.58E-19
1879:WT (2.5 mg/ml) (n=108)	613:WT (2.5 mg/ml) (n=118)	4.77E-53
1879:WT (10 mg/ml) (n=150)	613:WT (10 mg/ml) (n=159)	2.49E-45

Šidák corrected alpha (α=0.05): 0.00465

Supplemental Figure S9. Two-sample t-test p-values for selected mean gut fluorescence values depicted in Fig.5. Values marked in red are not significant at α=0.05.



Supplemental Figure S10. *elt-2* expression is not significantly different between high and low *gfp* expressing strains. A) *elt-2* controls in the high *gfp* expressing 613:WT and the low *gfp* expressing 613:A. B) Significance test.