How affinity of the ELT-2 GATA factor binding to *cis*-acting regulatory sites controls *C. elegans* intestinal gene transcription

Brett R. Lancaster and James D. McGhee
DOI: 10.1242/dev.190330

Editor: Susan Strome

Review timeline
Original submission: 7 March 2020
Editorial decision: 21 April 2020
First revision received: 2 May 2020
Editorial decision: 27 May 2020
Second revision received: 29 May 2020
Accepted: 6 June 2020

Original submission

First decision letter

MS ID#: DEVELOP/2020/190330

MS TITLE: How Affinity of the ELT-2 GATA Factor Binding to *cis*-acting Regulatory Sites Controls *C. elegans* Intestinal Gene Transcription

AUTHORS: Brett R. Lancaster and James Douglas McGhee

I have now received reviews of your manuscript from 3 experts. Their comments are appended below, or you can access them online: please go to BenchPress and click on the 'Manuscripts with Decisions' queue in the Author Area.

As you will see, all 3 reviewers consider your study to address an important topic and to present interesting results that will be appreciated by others in the field. They recommend numerous text changes to clarify points and improve the readability of your paper, most of which I think you will agree with. Reviewer 3 suggests additional experiments to rule out some possible complications with your assays. Because Reviewers 1 and 2 were not concerned with those possible complications and because the current pandemic is keeping researchers out of their labs, you do not need to tackle those additional experiments. That said, you may want to address in the text possible complications with the assays.

Please submit a revised manuscript, highlighting your changes. You should avoid using 'Tracked Changes' in Word files as those are lost in PDF conversion. Please also provide a point-by-point response detailing how you have dealt with the reviewersâ€™ suggestions.

Reviewer 1

Advance summary and potential significance to field

This paper from the McGhee lab addresses the relationship between TF binding strength and transcription levels in the C. elegans intestine. Previous work from this and other labs showed that GATA transcription factors, especially ELT-2 are the dominant regulators of intestines-specific

expression through binding to TGATAA motifs. Here, they show that ELT-2 affinity in vitro depends substantially on the nucleotides flanking the core TGATAA motif, that these affinity differences predict expression differences in a quantitative reporter assay, and that AGATAA motifs drive lower expression than predicted by their in vitro affinity alone. Overall it is a nice study that will be of interest to people in the gene regulation community, both in C. elegans and in other systems. I do have some suggestions for the authors to address, and as outlined below, the authors should be careful not to overstate the extent to which their reporter results favor or exclude specific models of motif collaboration (synergy, etc).

Comments for the author

Comments
F1)
This seems compelling, but the authors should address the possibility that there is ascertainment bias towards TGATAA motifs relative to AGATAA in the 44 known examples. For example citing some comparative enrichment analysis of TGATAA relative to AGATAA in promoters of intestine specific genes (as defined in earlier studies) to help the reader understand the extend of any enrichment for this site.

Are there cases (beyond the asp-1 example in the current study) where mutating TGATAA to AGATAA has a detectable impact on expression?

F2)
Writing of this section is very methods-rich. Maybe try to move some of the details (e.g. that fluorescein was used as the dye) to methods and try to reword to emphasize the main biological goals and conclusions.

Were these experiments done with full length ELT-2 or just the GATA domain?

124: “identical except for the XXTGATAAXX whose affinity was being tested “ – reword confusing

Unclear from Fig 2B – does the “……” mean that more GATAA variants were tested, or are all tested sequences in the figure as is?

How does the binding specificity as a function of flanking nucleotides compare to that measured previously for other related GATA factors (such as elt-7 by PBM in Narasimhan et al 2015, which also seems to prefer TGATAA to AGATAA)?

F3/4) The assay is nice – not terribly novel conceptually but takes good advantage of this quirk of C elegans transgenesis. Despite authors’ claims that chromatin is not a factor, it is unclear how extrachromosomal array chromatin relates to endogenous locus.

In Fig 4A it would be nice to have p values denoted with symbols on the plot In addition to including ‘controls,’ Fig 4 also as data about the relative role of the TGATAA sites in the asp-1 promoter; I suggest renaming the figure legend.

The authors should clarify how p-values were calculated. It appears there is relatively substantial strain to strain variability so this should be taken into account.

F5)
The modeling is thought provoking although the concept that activity is a joint function of binding site affinity and TF concentration is not a new one. More importantly for this specific analysis, the evidence for a “monotonic decrease” with decreased binding site affinity is weak.
In fact, the second highest affinity site (TGATAA with Krel ~= 0.45) appears to lead to equivalent expression levels as the maximum affinity sites. Similarly, the site at Krel of ~0.2 appears more active than the 0.4 Krel site. This, along with the observations regarding the AGATAA sites seem to me to argue that there is a lot more going on besides just affinity.
Regardless of this point, it seems like the data are equally compatible both with the synergistic and partial synergy models, as well as likely many other models not considered.

Reviewer 2

Advance summary and potential significance to field

This paper introduces a new method for measuring relative transcription initiation rates from two variant promoter sequences simultaneously. That is used to determine the relationship between intrinsic binding affinity and activation for a set of ELT-2 binding site variants. In general there is a direct relationship between affinity and activation, except for an exception of the site AGATAA. This shows much reduced activation compared with its affinity, which is consistent with its low usage in vivo.

Comments for the author

This is a very good paper that addresses the important topic of the relationship between transcription factor (TF) binding affinity to different sequences measured in vitro, and the functional activity in vivo of those sequences. It tests several different variants in the flanking region of binding sites for the ELT-2 TF, for each of which the relative affinity has been determined, on their ability to activate transcription from a reporter gene. They introduce a novel technique, SQRIPT, to measure the intrinsic transcription initiation rate for each variant which compensates for other confounding factors besides binding affinity in vivo. The method provides very good quantitative activity values and their results make several interesting points about the product of the TF concentration and the affinity of the binding sites. In addition they study one particular variant, between the common TGATAA and the uncommon AGATAA, in detail. They confirm (consistent with previous work) that the binding affinities to ELT-2 measured in vitro for those two sequences are very similar, differing only by about a factor of 2. But they have widely different usage in vivo, with the TGATAA sites used about 30 times more frequently. That fact presented a conundrum about whether the binding was also nearly equivalent in vivo, and some other factor led to their disparate usage, or if they binding affinity in vivo was different. They present convincing data that the binding in vivo is actually distinct, with the AGATAA site behaving similar to other sites with much reduced affinity. They provide some speculation about possible mechanisms for the distinction.
I have only a couple of suggested revisions that I think will help clarify some points.

1. I didn't ever see a complete list of the flanking variants that they used. In Figure 2A they show binding graphs for the highest affinity site and four variants plus a control sequence. Figure 2B is a table containing six flanking variants plus the AGATAA variant, but I'm not sure if the "..." row indicates there are others which are not included, of that is the complete set. Assuming those six are the complete set, they could just add (in the text or the legend) that "part A shows the binding curves for a subset of the varints, and the complete list of relative affinities are shown in part B". or if there are more that are not included, that should also be stated and the data presented somewhere.

2. I found Figure 3B confusing. In the top row why are R1 and R2 shown in both orientations? is that important? and in the next row they show two bands for each of R1 and R2, and I was trying to figure out how those were related. From supplemental figure S6 I realized that the two bands are the two ends of the PCR fragments after cleavage (because the sums of the lengths are the same), and it would be useful to have that stated in the legend to figure 3. But is that related to the upper row with R1 and R2 in two orientations? I don't think so.

Reviewer 3

Advance summary and potential significance to field

The authors advance a quantitative system for the in vivo study of the importance of binding sites to gene activation in C. elegans. The results presented in the paper are straightforward: binding site affinity tracks with in vivo expression levels, but that is not the whole story. The results should

be of interest to those in C. elegans but also in other model organisms, particularly those interested in mechanisms of gene expression.

Comments for the author

The ELT-2 is a tissue-specific specification/differentiation master gene that directs the intestinal fate in C. elegans. This lab has published dozens of papers validating the role of ELT-2 in this capacity.
In this paper, the authors attempt to associate binding site affinity of ELT-2 with expression of a target gene, asp-1, an enzyme encoding a protease involved in digestion, whose dominant activator is likely to be ELT-2.
The authors first present a 'consensus' binding site for ELT-2 that reflects multiple lines of study from prior work, including site enrichment present in promoters of genes from intestinally expressed mRNAs, and other in vivo and in vitro work. They use competitive EMSAs to measure binding affinity of recombinant ELT-2 to a series of sequences to find the optimized ACTGATAAGA binding site. Interestingly the modification of the flanking 2 bp on either side of the TGATAA core can affect binding affinity by some 10-fold.
The authors describe a system which they call 'SQRIPT' to compare altered transgene expression of modified asp-1 transgenes from the same extrachromosomal arrays. This system should theoretically normalize for effects of multicopy arrays and chromatin accessibility as whatever effects occur in the assembly and propagation of these, they would be expected to affect both reports the same way. This way the authors can compare the in vitro EMSA and in vivo reporter assays. In passing, the only other GATA factors in the intestine at these late stages are ELT-7 and the tiny (and likely non-functional) ELT-4, and the authors used strains in which both of these have been deleted, leaving endogenous chromosomal ELT-2 the only functional GATA factor remaining in the intestine.
In one result, it is found that in vitro binding of ELT-2 to sites containing AGATA vs. TGATA are relatively close, with AGATA affinity approximately 78%-94% (with a substantial range around these) of that of TGATA. In the in vivo, however, the AGATA motif is quite a bit weaker in its apparent response to ELT-2-dependent expression.
In another, the two adjacent GATA sites in the asp-1 promoter are found in the SQRIPT assay to function largely but not completely synergistically. The authors do a further nuclear run-on assay to measure nascent mRNA levels and conclude that the SQRIPT assay reflects differences in rates of transcription initiation.
Next, the authors mutate the pairs of ELT-2 sites in the asp-1 promoter to measure the effect of mutation to the sites that were tested in the EMSA competition assay. They find that the stronger the binding, the higher the expression level. The authors present a relationship of relative transcript levels measured with the in vivo assay as compared with binding affinity (Krel) in vitro.
In a last experiment, the authors choose two binding sites, one with a core AGATA site, and one with TGATA, but with a context that makes the relative binding affinity in vitro approximately equal. A surprising result is that the in vivo expression data do not come out the same, that is there is another aspect to ELT-2 activity on the arrays in vivo that goes beyond simple affinity as measured in vitro.
This is a nice study that presents a useful assay for in vivo TF binding site assays in C. elegans, and the finding that simple binding site affinity in vitro does not determine in vivo activity is a useful finding. The results should be of interest to those in C. elegans but also in other model organisms. The paper is well-written and is suitable for Development but I have some comments for revisions.

Major comments:
1. The work does rely on comparison of multicopy arrays. Although this should 'cancel out' differences between control and altered *transgenes*, it would be extremely helpful to have a chromosome-level validation. Such an experiment could be performed by introducing reporter plasmids into the genome by the well-established method of MosSCI into defined chromosomal sites. Then the comparison of expression would be done by the same downstream analysis by quantitative RT-PCR. If one or two test cases reproduced essentially the same results as with the SQRIPT system, then the validation would be complete. It would be very useful for other labs to use a variant of this system, if they knew that in at least one study that single-copy transgenes behaved indistinguishably.

2. The authors could have tested SQRIPT with changing phasing/spacing of two wild-type TGATAA sites and by adding a third site at equal spacing to the original. Is there an increased likelihood of binding and expression?
3. The presence of what must be hundreds of asp-1 transgenes in the arrays does not of course imply 100% occupancy of all available binding sites with endogenous ELT-2 protein. It would be nice if there could be an estimate of this occupancy somehow.
4. In a similar vein it is known that in the initial assembly of transgene arrays, the plasmids undergo homologous recombination (Mello, 1991; from the Abstract: "homologous recombination drives array assembly"). The 1.5 kbp of distance between the ELT-2 sites and the KpnI sites provides a large target for homologous recombination, in fact. This means that there will be in any array an assemblage of "hybrid" reporters in which the promoters are exchanged relative to the downstream KpnI sites. This would be expected to be in the minority of overall reporters, but even if it was 5%, it might be significant. Perhaps an experimental estimation of this effect could be made by some means, such as comparing relative expression when two arrays were derived independently. Although the use of the two transgenes in one array is designed to minimize differences in the arrays, there is probably enough consistency in the assemblage of arrays from independent injections using the same concentration of plasmid DNA to be able to estimate possible origin of hybrid reporters. Alternatively perhaps a PCR-based strategy could be found to amplify undesired hybrid transgenes.
Minor Comments:
line 113 – 'among' (not 'between') different co-regulated promoters line 360 – the authors should cite a recent study that proposed that the upstream endodermal GATAs END-1 and END-3, and also ELT-7, might have evolved by a gene duplication mechanism from an ancestral simpler network involving only ELT-2 (Maduro, 2020, PMID 31740453). This is a simple explanation for their interchangeability.
Supplementary figure S5. – The species is C. japonica, not C. japonicum.

First revision

Author response to reviewers' comments

Reviewer # 1.
F1)

- *(With respect to Figure 1) "This seems compelling, but the authors should address the possibility that there is ascertainment bias towards TGATAA motifs relative to AGATAA in the 44 known examples."*

We suggest that the possibility of "ascertainment bias", as suggested by the reviewer, does not arise with respect to Figure 1 for two main reasons. First of all, the data that were used to produce Figure 1 do not derive from a sampling process, random or biased. Instead, the data represent all examples that I could find in the literature where *cis*-acting motifs (invariably GATA sites) were actually experimentally manipulated and shown to have a significant influence on intestinal gene transcription (usually reporter expression). In most (not all because in some cases the authors assumed the site was a GATA site) of these studies, the sequences were defined objectively by deletion/mutation experiments and hence are not prone to ascertainment bias. The second point to make is that, in *C. elegans* and other organisms with compact promoters (e.g. yeast), numerous approaches looking for statistical enrichment of *cis*-acting motifs have shown that intestinal promoters are clearly enriched in TGATAA sites and at the same time depleted in (A/C/G/)GATAA sites. This is not ascertainment bias (i.e. a bias in sequences intrinsic to the overall genome) because promoters from muscle or neural genes do not show this bias. I'm including a Figure from our old SAGE paper (McGhee *et al*,; 2009; Figure 2B) to demonstrate this enrichment. We had referenced this study (and others) with respect to Figure 1 and had explicitly pointed out that such sequence patterns (i.e. extended TGATAA sequences like those shown in Figure 1) "have been computationally identified as **over-represented motifs** in the regulatory regions of all genes transcribed in the *C. elegans* intestine, from embryos to adults ….(emphasis added). Presumably, this enrichment has occurred via evolutionary selection acting on the

responding gene control processes.

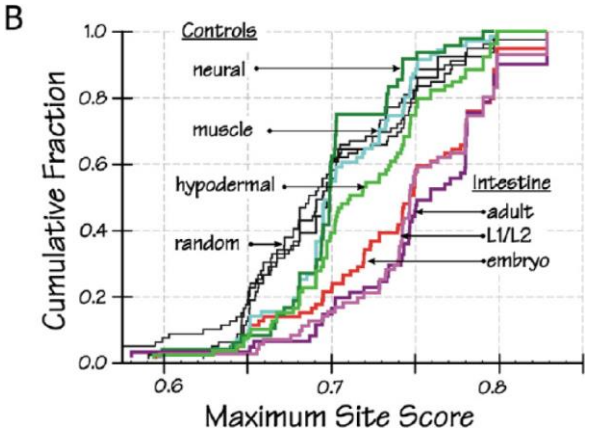

Figure 2B of McGhee, J. D., T. Fukushige, M. W. Krause, S. E. Minnema, B. Goszczynski *et al.*, 2009 ELT-2 is the predominant transcription factor controlling differentiation and function of the C. elegans intestine, from embryo to adult. Dev Biol 327: 551-565.

- *Are there cases (beyond the asp-1 example in the current study) where mutating TGATAA to AGATAA has a detectable impact on expression?*

None of which we are aware. The closest we could find is an experiment in which we had mutated an ATGATAG site in the *elt-2* promoter (Weisenfahrt *et al,* 2016). However, this was done as part of a series of simultaneous knockouts and it was impossible to conclude anything about the potency of any individual site.

F2) I presume the reviewer is referring to the text associated with Figure 2.

- *Writing of this section is very methods-rich. Maybe try to move some of the details (e.g. that fluorescein was used as the dye) to methods and try to reword to emphasize the main biological goals and conclusions.*

As suggested by two reviewers, the section describing the binding competition was both methods-heavy and unclear. We rewrote this section, removing any details that were repeated in the Methods section and shortening it by ~ 40%. We hope it will be clearer to understand. The Figure Legend 2 was also rewritten as suggested by the reviewer. The ellipses "…" in Figure 2B was an unfortunate choice and not meant to suggest there were other sequences that were not included in the Table. The original aim was to set off the TGATAA measurements from the AGATAA measurement; the dots have been removed.

- *Were these experiments done with full length ELT-2 or just the GATA domain?*

All band shift experiments were performed with full length ELT-2 protein and this has now been noted in the text as well as in the Methods section.

- *124: "identical except for the XXTGATAAXX whose affinity was being tested " - reword, confusing*

Unclear from Fig 2B - does the "……" mean that more GATAA variants were tested, or are all tested sequences in the figure as is?

Both of the above points were taken care of in the rewriting of this entire section.

- *F3/4) The assay is nice - not terribly novel conceptually but takes good advantage of this quirk of C elegans transgenesis. Despite authors' claims that chromatin is not a factor, it is unclear how extrachromosomal array chromatin relates to endogenous locus.*

I take this as a comment more than a criticism. We certainly acknowledge that it is unknown how similar or how different the "chromatin" of extrachromosomal arrays is to the chromatin of

endogenous genes. All we are saying is that these genes are properly regulated (to a good approximation) and that the control reporter is in the same environment as the test reporter, allowing us to isolate the effect of TGATAA mutations and ELT-2 affinity on reporter expression. Our aim was not to study array chromatin but to develop an experimental system in which any "chromatin" effect, whatever it is, can be normalized away.

- *How does the binding specificity as a function of flanking nucleotides compare to that measured previously for other related GATA factors (such as elt-7 by PBM in Narasimhan et al 2015, which also seems to prefer TGATAA to AGATAA)?*

We are reluctant to expand on detailed comparisons of the sequence preferences for other GATA factors, if only because the methods applied were vastly different (and not always immune to criticism). However, in response to the reviewer's suggestion, we have added the following sentence to the Discussion. "We note that binding motifs for several additional *C. elegans* GATA factors, both endodermal (ELT-7 but not END-1/3) and hypodermal (ELT-3, ELT-6, EGL-18) also appear to be enriched in TGATAA sequences (Shao *et al.*, 2013; Narasimhan *et al.*, 2015); it will be interesting to determine if these other GATA factors can, like ELT-2, discriminate *in vivo* against AGATAA sequences independently of binding affinity."

- *In Fig 4A it would be nice to have p values denoted with symbols on the plot*
In addition to including 'controls,' Fig 4 also as data about the relative role of the TGATAA sites in the asp-1 promoter; I suggest renaming the figure legend.
The authors should clarify how p-values were calculated. It appears there is relatively substantial strain to strain variability so this should be taken into account.

As suggested by the reviewer, I have added p-values to Figure 4A and 4B, as indicated by the brackets. We had stated in the text that significance of comparisons was estimated by t-tests; we have now also stated this in the Figure Legend to Figure 4. To incorporate variation in different transgenic strains, the test was applied to grouped data, i.e. replicate measurements obtained from the replicate independent transgenic strains produced for the same construct. I have also changed the title to Figure 4, replacing "Controls…" with "SQRIPT assay characterization of *asp-1* transcription".

F5) *The modeling is thought provoking although the concept that activity is a joint function of binding site affinity and TF concentration is not a new one. More importantly for this specific analysis, the evidence for a "monotonic decrease" with decreased binding site affinity is weak. In fact, the second highest affinity site (TGATAA with Krel ~= 0.45) appears to lead to equivalent expression levels as the maximum affinity sites. Similarly, the site at Krel of ~0.2 appears more active than the 0.4 Krel site. This, along with the observations regarding the AGATAA sites seem to me to argue that there is a lot more going on besides just affinity.*
Regardless of this point, it seems like the data are equally compatible both with the synergistic and partial synergy models, as well as likely many other models not considered.

We are certainly not trying to take credit for the application of The Law of Mass Action. Nonetheless, we would like to think that we have made a useful contribution by developing an analysis method that can isolate the product (TF Concentration X Highest Binding Affinity) in a way that can be measured and also for pointing out that this product, not just the affinity, should be recognized as the key factor in understanding transcription *in vivo;* furthermore, this product should be used to assess possible consequences of motif mutations, either in disease or in evolution; usually this is not the case. To address the reviewer's second point above, whether we should be able to state that reporter expression monotonically decreases with decreasing ELT-2 affinity, I suggest that this is almost a matter of choice or even style. We have deliberately formulated the simplest model in order to extract, if possible, general principles. We would agree with the reviewer that there maybe "a lot more going on besides just affinity" but that should not inhibit attempts to find general principles. We would also point out that the average relative transcript levels obtained at Krel ~ 0.45 are indeed lower than at Krel =1 but the reason that they are not **much** lower revealed one of the main conclusions of our analysis, namely that the response curve **plateaus** at higher relative ELT-2 affinities. Yes, the data bounce around a bit but we would like to ascribe this to normal variation resulting from a complicated assay. In fact, we

are pleased that the data can actually be fit so well with such simple thermodynamic models, whether based on complete or on partial synergy, although we favour neither one model nor the other. The purpose of introducing the complete synergy model first is that it is simpler; extension of this model to include partial synergy only leads to a minor adjustment in the predicted response curve. We added an additional sentence to emphasize and to clarify this progression but the most important point is that the same general conclusions emerge from either model: ELT-2 affinity dominates the transcriptional response driven by TGATAA sites and the Kmax*ELT-2free product must be roughly 10.

Reviewer # 2 The first comment concerns presentation of the competitive binding data, and is similar to the comment by reviewer # 1. As noted above, we have rewritten this section completely, modified Figure 2B somewhat, amplified the Figure Legend, and trust that these revisions satisfy the reviewer's comments.

2. I found Figure 3B confusing. In the top row why are R1 and R2 shown in both orientations? is that important? and in the next row they show two bands for each of R1 and R2, and I was trying to figure out how those were related. From supplemental figure S6 I realized that the two bands are the two ends of the PCR fragments after cleavage (because the sums of the lengths are the same), and it would be useful to have that stated in the legend to figure 3. But is that related to the upper row with R1 and R2 in two orientations? I don't think so.

Figure 3B depicted the R1 and R2 reporters in both orientations to emphasize that, within a multicopy array, reporter disposition is not known and could include both orientations. We have tried to explain this more clearly in the Figure Legend. The two green bands and the two red bands shown on the schematic gel image at the bottom of Figure 3B were indeed meant to represent the two *KpnI* fragments generated by restriction of the amplified cDNA; this is the entire basis of the SQRIPT assay and required us to insert *KpnI* sites that distinguish R1 and R2 as shown in Figure 3A. To make clearer that the four fragments derive from bisection of the two reporter sequences, I have added white lines over the amplified fragments in Figure 3B and have added the caption on the figure "Electrophoretic separation of *KpnI* digestion products".

Reviewer # 3

*1) The work does rely on comparison of multicopy arrays. Although this should 'cancel out' differences between control and altered *transgenes*, it would be extremely helpful to have a chromosome-level validation. Such an experiment could be performed by introducing reporter plasmids into the genome by the well-established method of MosSCI into defined chromosomal sites. Then the comparison of expression would be done by the same downstream analysis by quantitative RT-PCR.*

This is exactly the experiment that would have been (and perhaps may be done) next on our list if we are (somehow) able to pursue this project. We agree that the results of such an experiment would be "extremely helpful", assuming that the single copy reporters provide sufficient sensitivity in the RT- PCR assay. We can only plead that the experiment as suggested (but apparently not required by the reviewer) is beyond the scope of the present investigation. In an ideal world, we would have already done it.

2) The authors could have tested SQRIPT with changing phasing/spacing of two wild-type TGATAA sites and by adding a third site at equal spacing to the original. Is there an increased likelihood of binding and expression?

We had performed a limited number of such addition/transposition/switch experiments with the GATA motifs in the reporter promoters. Like many such promoter-bashing experiments, the possible variations are endless and the interpretations are often ambiguous. For example, we had introduced a third TGATAA site just downstream of the natural pair and had found that transcription was enhanced minimally if at all. The straightforward interpretation is that, exactly as would have been predicted from Figure 5, the wildtype *asp-1* promoter is already working near saturation and addition of one more site would not be expected to enhance transcription. Alternatively, perhaps the additional site had been introduced too close to the start site of

transcription to be effective. These and several additional (highly suggestive but not definitive) promoter-rearrangement experiments are recorded in Brett's thesis. In the current study, we have focused on targeted mutations in the TGATAA sites within their natural sequence context, arguing that at least this approach should yield interpretable results.

3) The presence of what must be hundreds of asp-1 transgenes in the arrays does not of course imply 100% occupancy of all available binding sites with endogenous ELT-2 protein. It would be nice if there could be an estimate of this occupancy somehow.

We agree with the reviewer that such an estimate of ELT-2 occupancy would be nice. However, this is an entirely different project and we suggest that it is outside the scope of the present study. Probably the experiment would best be done at the same time as the single copy genomic insertion experiments suggested above by this reviewer, using anti-ELT-2 Chip-seq to ask whether ELT-2 relative affinity does indeed reflect ELT-2 relative occupancy *in vivo* to these single copy promoters; the experiment would have to be able to distinguish the highly similar promoters in the immunoprecipitates but this should emerge from the sequencing.

4) In a similar vein it is known that in the initial assembly of transgene arrays, the plasmids undergo homologous recombination (Mello, 1991; from the Abstract: "homologous recombination drives array assembly"). The 1.5 kbp of distance between the ELT-2 sites and the KpnI sites provides a large target for homologous recombination, in fact. This means that there will be in any array an assemblage of "hybrid" reporters in which the promoters are exchanged relative to the downstream KpnI sites. This would be expected to be in the minority of overall reporters, but even if it was 5%, it might be significant. Perhaps an experimental estimation of this effect could be made by some means, such as comparing relative expression when two arrays were derived independently. Although the use of the two transgenes in one array is designed to minimize differences in the arrays, there is probably enough consistency in the assemblage of arrays from independent injections using the same concentration of plasmid DNA to be able to estimate possible origin of hybrid reporters. Alternatively perhaps a PCR- based strategy could be found to amplify undesired hybrid transgenes.

The reviewer raises an important point, namely that, in the process of establishing the extrachromosomal array, any recombination event between the GATA sites and the reporter-distinguishing *KpnI* sites would have the effect of " randomizing" the resulting transcripts and decrease the observed reporter-specific signals. We were worried about this possibility from the start and, if it could have been shown to be significant (it wasn't), were prepared to delete some of the relatively long *asp-1* leader sequence, as well as shift the inserted *KpnI* sites close to the 5'-end of the coding sequence, thereby lowering the opportunities of inter-reporter recombination. However, we then realized that such putative recombination could **not** be a **major** limitation of the SQRIPT assay. Our strongest argument comes from the observation (e.g. Figure 4A) that ablation of both of the *asp-1* TGATAA sites lowers reporter expression to close to background (5-10% of the wildtype control). This result could not be obtained if the linkage of the TGATAA sites to the reporter-distinguishing *KpnI* sites were randomized by recombination during array assembly. This residual level of reporter expression could be due to recombination, as the reviewer surmised, or it could be due to transcriptional run-on in the array, or it could be due to low level of *bona fide* regulated transcription driven by other sites in the mutated *asp-1* promoter, e.g. the GATA sites that are not TGATA. The importance of transcriptional read-through or transcriptional promiscuity could be assessed by the non- promoter control (5-10% of wildtype levels) and, as described in the text, this basal level was subtracted from the measured relative rate for each promoter. So overall, maybe as much as 5-10% of the signal distinguishing test promoter from control promoter could be due to recombination. This is a **possible** limitation of the SQRIPT assay but not a serious one. The overall uncertainty in the assay (from intrinsic variability of the RT-PCR reactions, from variation between independent transgenic strains, etc.) is of the same order of magnitude. Overall, we are focusing on first order effects in order to extract the most general conclusions and none of the above considerations weaken these general conclusion. I would hesitate to compare test and control constructs in separate arrays, as suggested by the reviewer. I expect there would be just too much variability, if only because of different copy numbers. Besides, comparing test and control promoters in "identical" environments is a strength of SQRIPT.

5) line 113 - 'among' (not 'between') different co-regulated promoters

The reviewer did not explain why “among” should be preferred over “between” in this context. I would disagree with this suggestion. According to a number of web sources, the use of “between” need not be confined to discussions of ONLY two objects but, more broadly, should be used to discuss multiple objects as long as these objects have well defined identities. “Among” is better used when the multiple objects under discussion do not have individual identities (e.g. trees in a forest). We have used “between” because the co-regulated promoters have discrete individual identities (and references), and I suggest that “between” remains the correct usage…..(much ado about not much….but still fun).

6) line 360 – the authors should cite a recent study that proposed that the upstream endodermal GATAs END-1 and END-3, and also ELT-7, might have evolved by a gene duplication mechanism from an ancestral simpler network involving only ELT-2 (Maduro, 2020, PMID 31740453). This is a simple explanation for their interchangeability.

Lines 357 to 360 describe the interchangeability of the *C. elegans* endodermal GATA factors and do not discuss how this could have arisen during evolution. I have thus added the sentence: “Plausible scenarios have been proposed to explain how this interchangeability could have arisen in evolution (Wiesenfahrt *et al.*, 2016; Maduro, 2020).”

Finally, we have changed *C. japonicum* to *C. japonica* in Supplementary Figure S5.

Second decision letter

MS ID#: DEVELOP/2020/190330

MS TITLE: How Affinity of the ELT-2 GATA Factor Binding to cis-acting Regulatory Sites Controls C. elegans Intestinal Gene Transcription

AUTHORS: Brett R Lancaster and James Douglas McGhee

I have now received reviews of your manuscript from the original 3 reviewers. Their comments are appended below, or you can access them online: please go to BenchPress and click on the 'Manuscripts with Decisions' queue in the Author Area.

As you will see, all 3 reviewers are satisfied with your revised manuscript. Before I formally accept it for publication, please consider the final suggestion from Reviewer 1 for clarifying in your paper the point about the known target bias for TGATAA vs AGATAA.

Please submit a revised manuscript, highlighting any changes you made.

Reviewer 1

Advance summary and potential significance to field

I remain positive about this paper and the authors have in general done a good job responding to my suggestions. The advances/significance remain the same as in my earlier review.

Comments for the author

Re: my point about Figure 1, I apologize for being unclear –the authors nicely listed in the response evidence that the known target site bias for TGATAA vs AGATAA isn’t just because of ascertainment bias. My suggestion was more that a more explicit (albeit briefer) statement of this would be useful

in the text. The current wording about “computationally identified over-represented motifs” I found a bit vague, and it seems like this point is worth making strongly given the later results in the paper. Still, this is really more of a stylistic point and I’d leave it up to the authors.

Reviewer 2

Advance summary and potential significance to field

This paper introduces a new method for measuring relative transcription initiation rates from two variant promoter sequences simultaneously. That is used to determine the relationship between intrinsic binding affinity and activation for a set of ELT-2 binding site variants. In general there is a direct relationship between affinity and activation, except for an exception of the site AGATAA. This shows much reduced activation compared with its affinity, which is consistent with its low usage in vivo.

Comments for the author

I think the revisions have adequately addressed the concerns raised by myself and the other reviewers and I have not further suggested revisions.

Reviewer 3

Advance summary and potential significance to field

Although revisiting an old question in a new way that makes use of the quirk by which transgenes assemble into arrays in C. elegans, the paper advances a straightforward way of obtaining 'first-order' information about transcriptional control of gene expression that should be of interest to those studying gene regulation in a developmental context.

Comments for the author

The authors have done a satisfactory job of addressing reviewer comments.

Second revision

Author response to reviewers' comments

I have replaced the sentence that reviewer # 1 found unclear (i.e. the sentence quoted in the reply to reviewers and containing the statements about "statistically over-represented motifs" with the following:

"It is well established that TGATAA-like sites are enriched (and (A/C/G)GATAA-like sites are correspondingly depleted) in the regulatory regions of all genes transcribed in the C. elegans intestine, from embryos to adults (Pauli et al., 2006; McGhee et al., 2007; McGhee et al., 2009; Dineen et al., 2018); (for reference, Supplementary Table S2 reproduces the position frequency matrix from (McGhee et al., 2009))."

I have also taken this opportunity to incorporate a brief statement of acknowledgements within the main body of the revised manuscript, whereas previously I had included acknowledgements as a separate file.

Third decision letter

MS ID#: DEVELOP/2020/190330

MS TITLE: How Affinity of the ELT-2 GATA Factor Binding to *cis*-acting Regulatory Sites Controls *C. elegans* Intestinal Gene Transcription

AUTHORS: Brett R Lancaster and James Douglas McGhee
ARTICLE TYPE: Research Article

I am happy to tell you that your manuscript has been accepted for publication in Development, pending our standard ethics checks.