Table S1. Collected examples of functional "GATA" sites in the promoters of *C. elegans* intestinal genes[a]

Click here to Download Table S1

Table S2											
	Position Frequency Matrix from McGhee et al 2009 Figure 2A										
		Position									
		1	2	3	4	5	6	7	8	9	10
Base	A	0.64	0.12	0.03	0	1	0	1	0.96	0.1	0.66
	C	0.07	0.5	0.01	0	0	0	0	0	0.26	0.08
	G	0.15	0.17	0	1	0	0	0	0.03	0.58	0.17
	T	0.15	0.21	0.96	0	0	1	0	0.01	0.06	0.09

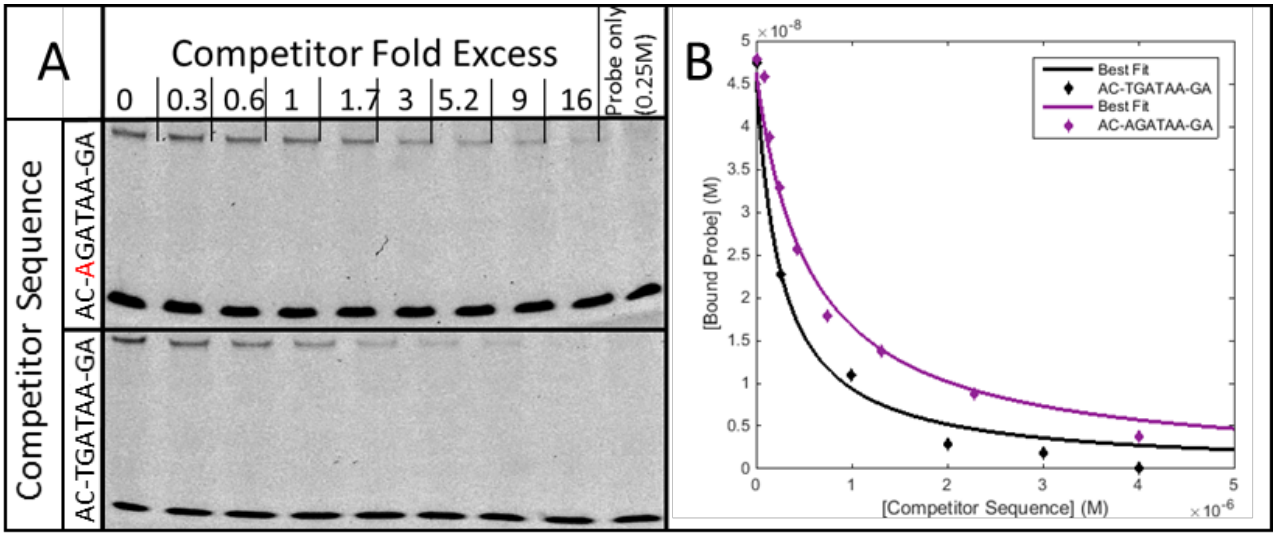


Figure S1. Representative gel images of competitive EMSA assays used to produce estimates of relative binding constants.

A. Free (unshifted) and ELT-2 Bound (shifted) bands are detected by the fluorescence (inverted contrast) of FAM-Labelled double-stranded oligodeoxynucleotide hairpins containing the highest affinity site ACTGATAAGA. Bottom panel shows self-competition of the labelled ACTGATAAGA-containing double stranded oligodeoxynucleotide with an unlabelled double stranded oligodeoxynucleotide containing the same high affinity site. The upper panel shows competition of the highest affinity labelled probe with an unlabelled double-stranded oligodeoxynucleotide containing the same sequence but with AGATAA replacing TGATAA.

B. Analysis of competitive EMSA gels shown in (A) in order to yield estimates of K_{rel}, the relative binding constant to ELT-2. The measured amount of bound (shifted) fluorescent probe is plotted against the total concentration of added competitor. Curves are generated as described in Supplementary Methods and used to yield estimates of K_{rel}.

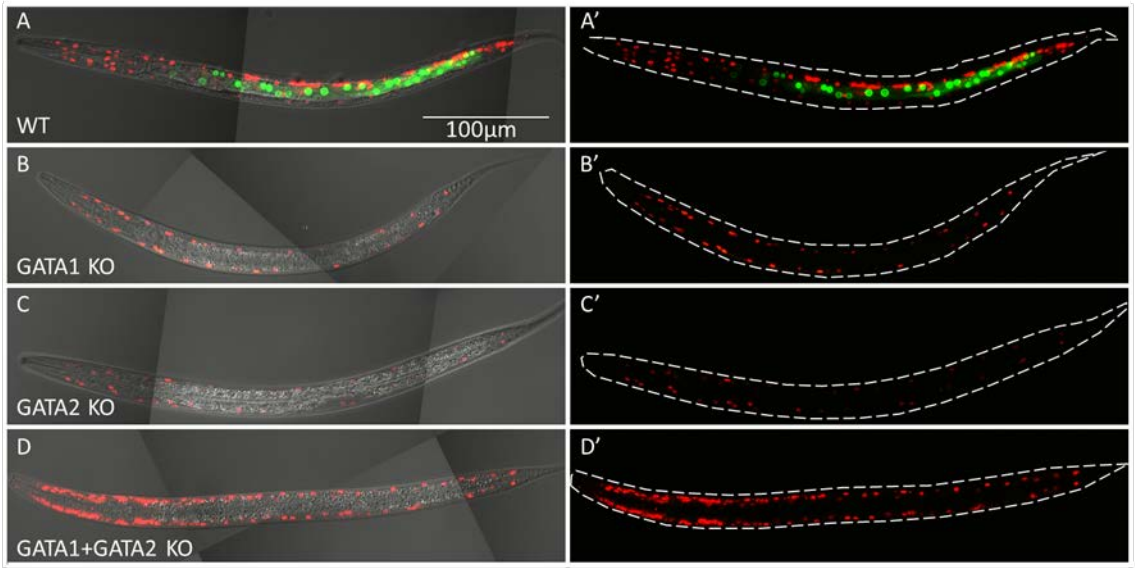


Figure S2: Analysis of WT and TGATAA KO *asp-1* Promoter Activity by Transgenic GFP Reporter Fluorescence

Green fluorescence is observed in the intestinal cells when GFP reporters are placed under transcriptional control of the WT *asp-1* promoter (A and A') but little fluorescence is observed when the *asp-1* promoter TGATAA sites are ablated, either singly (B, B', C, C') or together (D, D'). Red fluorescence reflects body wall muscle expression of a *myo-3promoter::rfp* reporter to indicate successful transgenesis. Left panels (A, B, C, D) = superimposed fluorescence (merged green + red) + differential interference contrast images; Right panels (A', B', C', D') = fluorescence (merged green + red). These images are representative of the majority of worms observed. Images were captured on a Zeiss Axioplan2i microscope with a Hamamatsu Orca digital camera and AxioVision software (version 4.8.1). All images were captured with the same settings and exposure time and displayed with the same brightness and contrast.

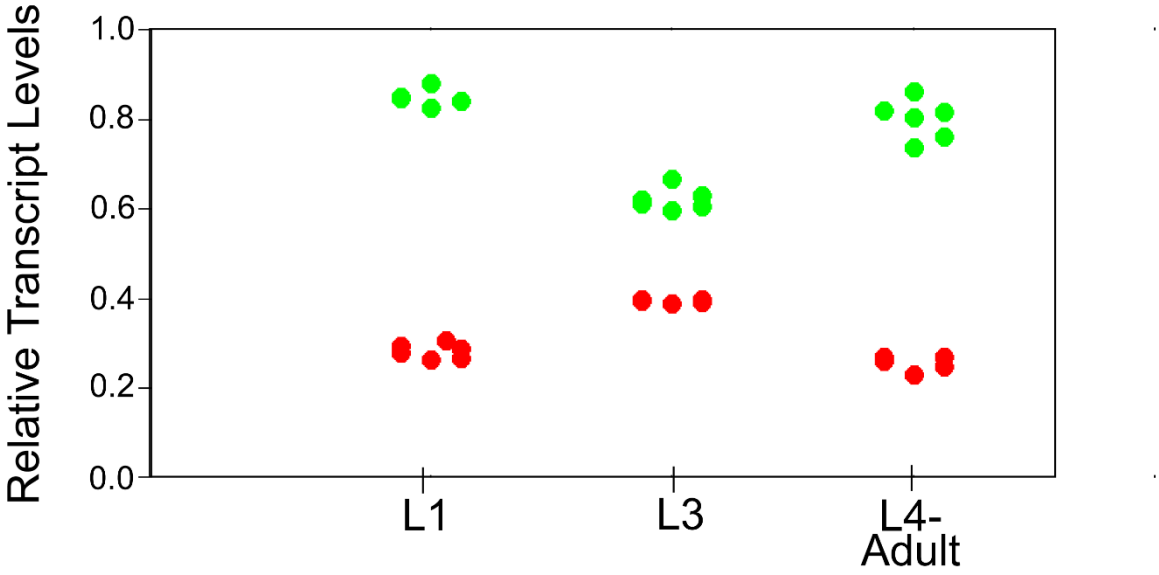


Figure S3. Relative Promoter Activity Measured at Different Developmental Stages Relative transcript levels produced by an *asp-1* variant promoter containing two copies of …AATGATAAGA… (green circles) or two copies of …GCTGATAATG… (red circles) replacing the two wildtype copies of the core TGATAA motifs. The data extends the analysis of Figure 4C of the main text to a weaker promoter and to a slightly stronger promoter.

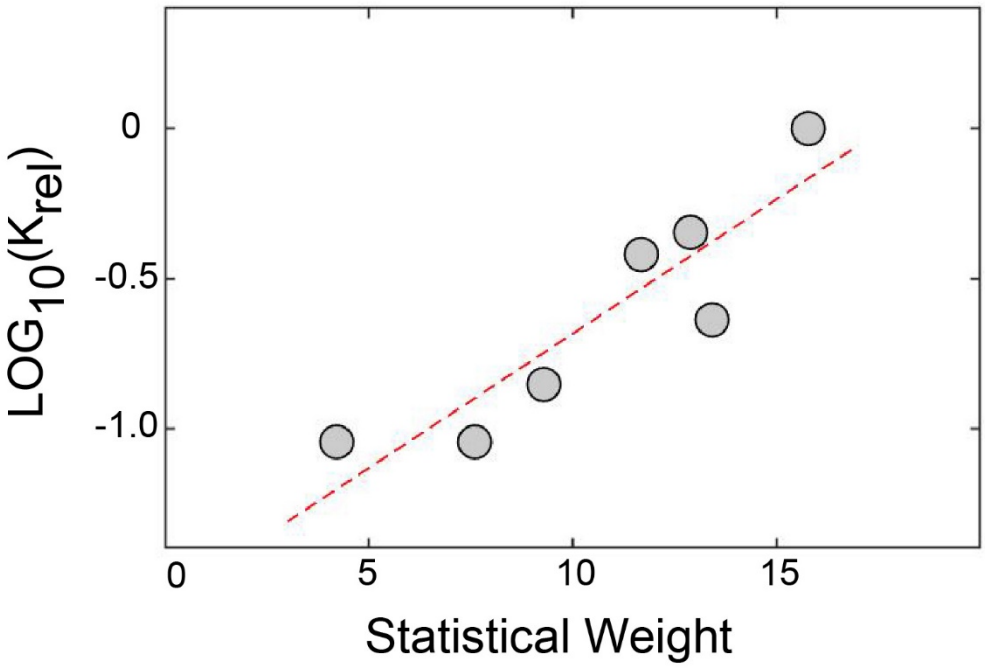


Figure S4

As described in more detail in the text, the position frequency data reproduced in Supplementary Table S2 were used to calculate a Statistical Weight for each of the decameric sequences for which the ELT-2 relative binding affinity had been measured. These statistical weights were then plotted against the logarithm of the relative binding constant. The dashed line represents the least squares regression fit to the data. The fact that this relation is linear is consistent with the conclusion that the frequency with which a *cis*-acting TGATAA regulatory site appears in the promoters of *C. elegans* intestinal genes is determined, in part, by the strength of binding of this sequence to ELT-2.

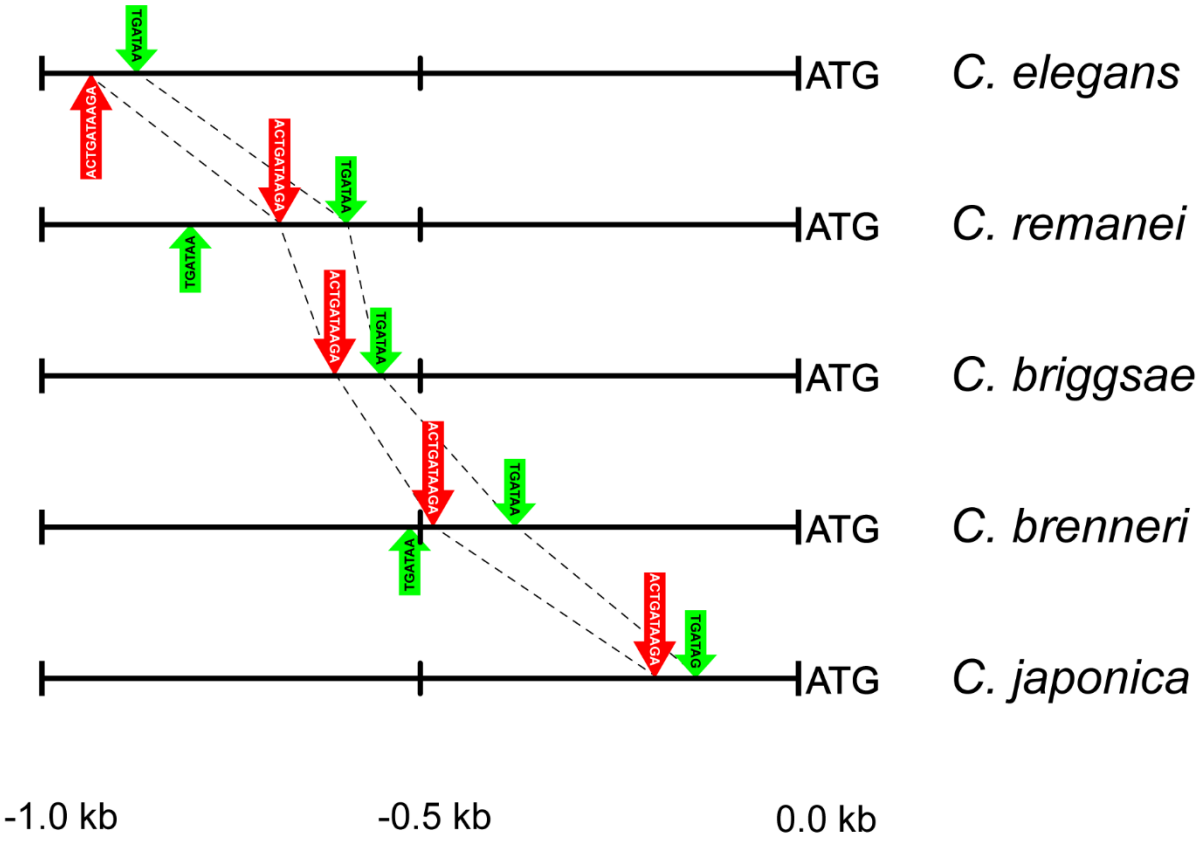


Figure S5. Conservation of Paired TGATAA sites in *Caenorhabditis asp-1* Promoter Homologs.
All TGATAA sequences are shown for the 1 kb regions upstream of *asp-1* homologs in related *Caenorhabditis* species (counting from the ATG initiation codon). The decameric sequence, ACTGATAAGA, with the highest binding affinity to ELT-2 is shown as the red arrow. All other sequences containing TGATAA are shown as the green arrows. In the Discussion section of the main text, we note that each homologous promoter contains two TGATAA sites (connected by the dashed lines) spaced <145 bps apart; one of these sites is the ELT-2 binding sequence with the highest affinity.

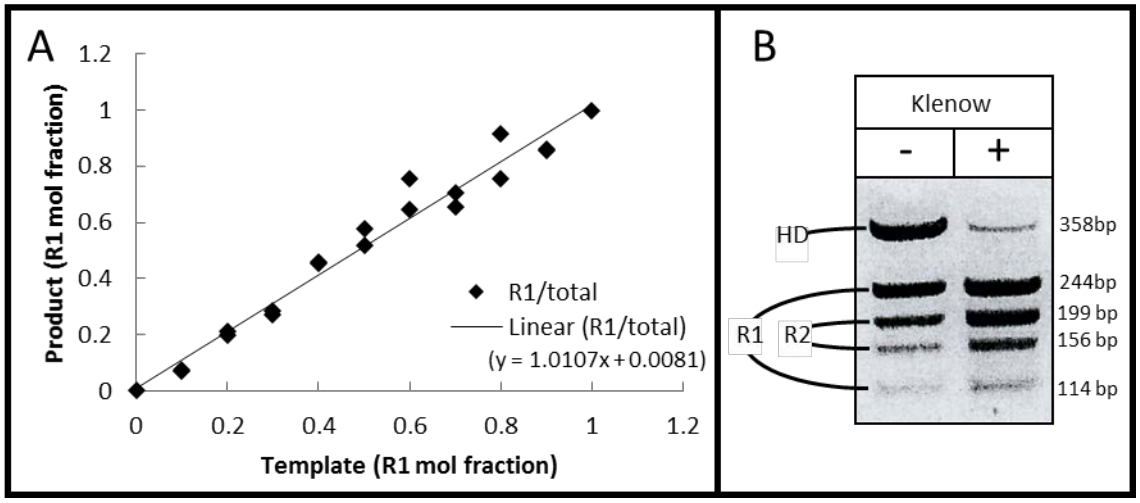


Figure S6: Validation of SQRIPT Procedures

(A) Plasmids containing reporter sequences R1 or R2 were mixed in defined mol fractions as PCR templates. Following amplification, *KpnI* digestion and electrophoresis, pixel intensities of the *KpnI* restriction fragments were quantitated. The slope of the best fit line is close to 1, indicating minimal amplification bias.

(B) In some R1 + R2 amplifications, a fraction of the amplification product was resistant to exhaustive *KpnI* digestion. We deduced that this product was a heteroduplex formed between the two highly similar reporter fragments. First, re-amplification of this resistant product regenerated fragments that were *KpnI* sensitive. Secondly, incorporation of a final extension period using Klenow polymerase with additional added primers removed the majority of the resistant product, as shown in the "+" lane. In practice, the intensity measured for any *KpnI* resistant product was apportioned equally between intensities assigned to reporter R1 and R2.

Supplementary Methods

Analysis of Competitive Band Shifts (See Figure 2 of main text)

Let the labelled oligodeoxynucleotide (present as a double stranded self-complementary hairpin "oligo") be represented as A.

Let the unlabelled competitor oligo be represented as C. C is also a self-complementary oligodeoxyonucleotide hairpin, identical to A except in the residues within 3 base pairs of the core GATA sequence.

Let the protein ligand (in the present case ELT-2) be represented as L.

Let the total concentration (activity) of A in the binding experiment be represented as A_{tot} (M).
Then $A_{tot} = A_b + A_f$ (M) where the subscripts b and f represent bound and free oligo, respectively.
Likewise for the competitor oligo: $C_{tot} = C_b + C_f$ (M).
Similarly for the protein ligand: $L_{tot} = L_f + A_b + C_b$ (M) where ligand can be either free in solution (L_f) or bound in a one-to-one complex with oligo A or oligo C ---(1)

Let the association constant of the ligand L binding to oligo A be K_A (1/M) and to competitor oligo C be K_C (1/M), where $K_A = A_b/(A_f \cdot L_f)$ ---(2) and $K_C = C_b/(C_f \cdot L_f)$ (1/M). ---(3)

And (rearranging the ratio of (3) to (2)): $C_b = A_b \cdot (K_C/K_A) \cdot (C_f/A_f)$ ---(4)

A typical competition experiment begins by mixing oligo A with ligand L under conditions such that most of oligo A remains free, aiming for $A_b \sim 0.15\ A_{tot}$. The precise fraction is measured on sample lane 1 (no added competitor C). Competitor oligo C is then mixed in increasing concentrations maintaining the same L_{tot} and A_{tot} (and with constant sample volumes) and the amount of A_b is measured (on sample lanes 2,3,4,....) for each total concentration of C. The objective is to measure A_b as a function of added C (i.e. C_{tot}) and then to estimate the ratio of the affinity constants, $K_{rel} = K_C/K_A$. In these experiments, A_{tot} and C_{tot} are known (within the accuracy of the assumption that their extinction coefficients are equal). The total protein concentration L_{tot} is constant in each sample but is known only approximately for reasons explained below. In practice, L_{tot} is estimated from a self-competition experiment in which unlabelled A competes with labelled A.

Equations (1) to (4) above are now combined and rearranged in order to obtain A_b as a function of C_{tot} and in a form that will allow $K_{rel} = K_C/K_A$ to be estimated. Other parameters are constants, either measured independently of the experiment (A_{tot}, C_{tot}) or, as just noted, estimated from a separate self-competition experiment (L_{tot}).

$$A_b = K_A \cdot A_f \cdot L_f$$

$$= K_A \cdot (A_{tot} - A_b) \cdot (L_{tot} - A_b - C_b) \quad \text{---(5)}$$

$$= K_A \cdot (A_{tot} - A_b) \cdot (L_{tot} - A_b - A_b \cdot (K_C/K_A) \cdot (C_{tot}/A_{tot})) \quad \text{---(6)}$$

where the C_b term in equation (5) is replaced by C_b from equation (4) above, at the same time replacing C_f/A_f with C_{tot}/A_{tot}. The basis of this approximation is that, under the experimental conditions, most of either oligo will not be bound, i.e. $A_f \lesssim A_{tot}$, $C_f \lesssim C_{tot}$, and the ratio C_{tot}/A_{tot} will be a better approximation than either of the individual quantities.

Equation (6) is expanded and rearranged as a quadratic in A_b, substituting $K_{rel} = K_C/K_A$.

$$A_b^2 \cdot [1 + (K_{rel}) \cdot (C_{tot}/A_{tot})] - A_b \cdot [1/K_A + A_{tot} + L_{tot} + A_{tot} \cdot (K_{rel}) \cdot (C_{tot}/A_{tot})] + A_{tot} \cdot L_{tot} = 0 \quad ---(7)$$

Solving the quadratic (and choosing the negative root in order that $A_b \leq A_{tot}$)

$$A_b = \{(1/K_A + A_{tot} + L_{tot} + K_{rel} \cdot C_{tot}) - [(1/K_A + A_{tot} + L_{tot} + K_{rel} \cdot C_{tot})^2 - 4 \cdot (1 + K_{rel} \cdot (C_{tot}/A_{tot})) \cdot A_{tot} \cdot L_{tot}]^{1/2}\}/\{2 \cdot (1 + (K_{rel}) \cdot (C_{tot}/A_{tot}))\} \quad ---(8)$$

To obtain numerical estimates of K_{rel} for any particular competitor C, the first step is to perform a self-competition of labelled A with unlabelled A in order to obtain estimates of K_A and L_{tot}. An initial estimate of the total ELT-2 concentration present in each binding reaction (obtained from conventional protein assays) is likely to be an overestimate because a fraction of the ELT-2 protein could be inactive in the binding assay; in addition, the effective ELT-2 concentration (i.e. its activity) will be lowered because of binding to components of the reaction other than the specific sequence probes, in particular poly(dI-dC) added to suppress non-specific binding. Using the simple vector commands available in MATLAB, trial values of L_{tot} and K_A are used to calculate A_b (for each of the known values of A_{tot} in the titration experiment) using Equation (8) above; subsequently the sum of the squares of the deviations between these trial values of A_b and the observed set of values for A_b are calculated. The following figure (Supplementary Figure S7) uses the data obtained from a typical self-competition experiment to plot this sum of the squares of the deviations as a function of K_A for a range of values of L_{tot} as shown on each curve. It can be seen that minimizing the sum of the squares of the deviations provides well defined estimates for L_{tot} of ~20 nanomolar and for K_A of ~2x10^7 (1/M).

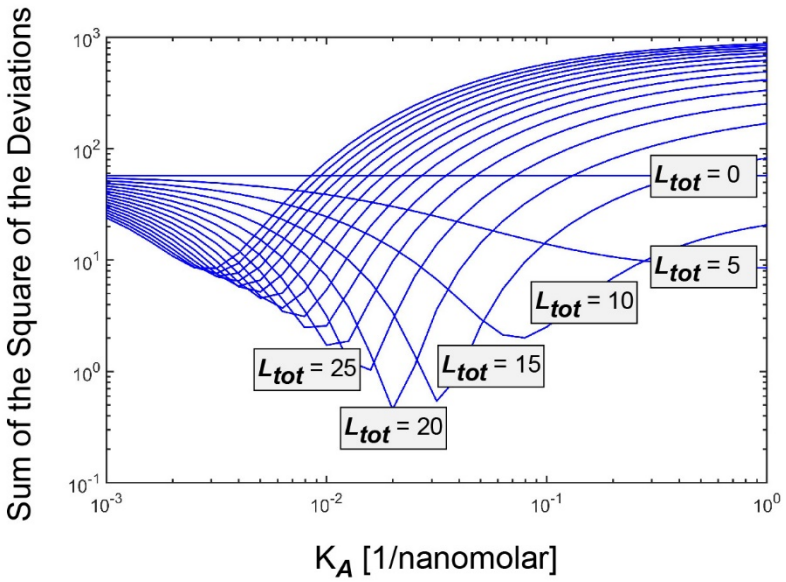


Figure S7

Using the values of L_{tot} and K_A determined by the self-competition experiment, the known values of A_{tot} and C_{tot} and the measured values of A_b, a similar strategy was used to obtain an estimate of $K_{rel} = K_C/K_A$. The sum of the squares of the deviations between calculated A_b (from Equation (8)) and measured A_b is then plotted as a function of trial values of K_{rel} , as shown in Supplementary Figure S8. For this particular set of binding data, the best estimate of K_{rel} is 0.12.

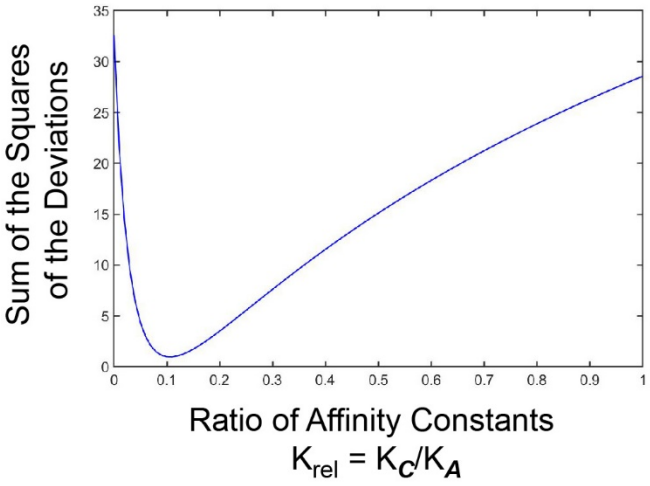


Figure S8

Relative Binding Affinity of ELT-2 to …AGATA… compared to …TGATA… motifs.

We measured the relative binding affinity of ELT-2 to AGATA… vs. TGATA… sequences using the method of "SpecSeq", as developed by Stormo and coworkers [1, 2]. We began with a degenerate "library" of oligodeoxynucleotides, synthesized according to the formula:

`tcctactcttctgtatgtcgNNNNGATANNNNcctaaccgactccgttaatt`

where the lower case sequences bind to appropriate primers, first to render the entire library double-stranded and labelled with fluorescein at the 5'-end of one strand and secondly to amplify ELT-2 bound and unbound fractions by PCR prior to sequencing.

We first determined the quantity of poly(dI-dC). poly(dI-dC) to be added to each EMSA reaction such that non-specific binding of ELT-2 is suppressed but specific binding remains. Increasing amounts of poly(dI-dC). poly(dI-dC) were added to a mixture of ELT-2 protein and a double stranded oligodeoxynucleotide (5'-labelled with FAM) that contains a tightly binding ACTGATAAGA motif (left series of lanes) or a mutated non-binding GATA motif (ACGTCGCCGA; right series of lanes), followed by electrophoresis. Images of typical EMSA gels are shown in Supplementary Figure S9, where it can be seen that 10-20 nanograms of poly (dI-dC). poly(dI-dC) per reaction is able to suppress non-specific binding, while at the same time leaving specific binding apparently unchanged.

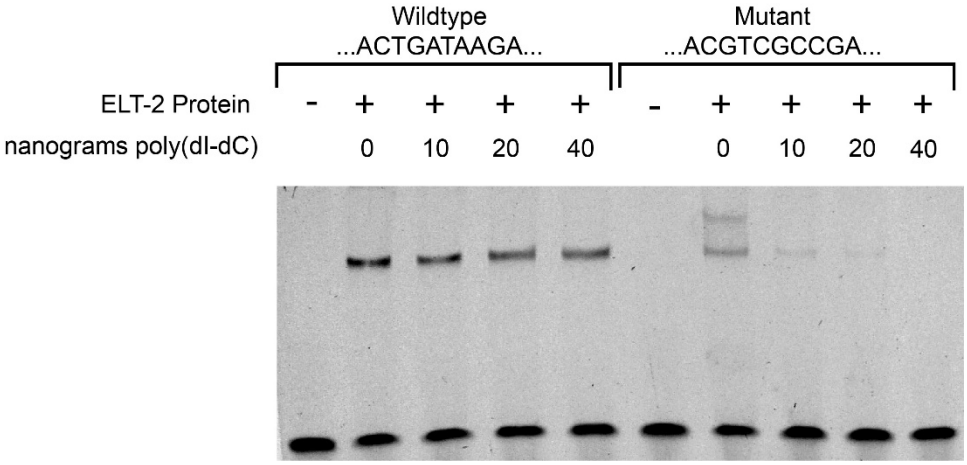
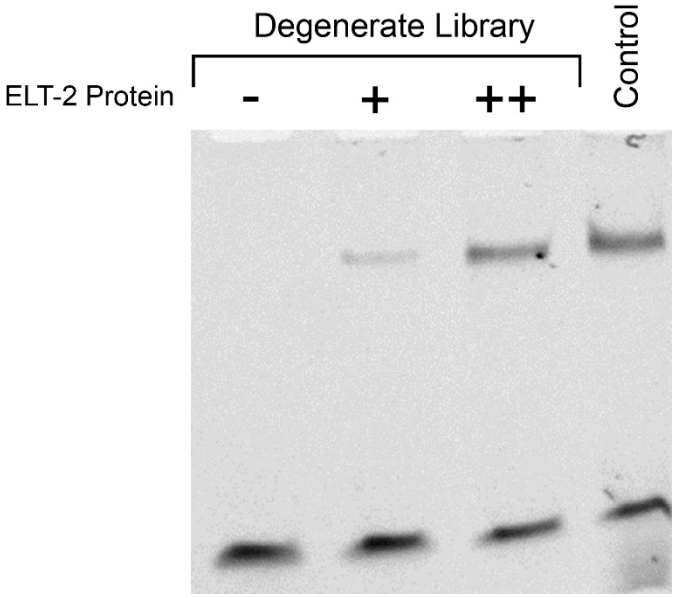


Figure S9

Figure S10 shows a fluorescent image of a typical band shift gel using the degenerate libraries, with no ELT-2 protein (left) and then two different levels of ELT-2 protein (synthesized in baculovirus infected insect cells) mixed in with the double stranded degenerate fluorescently-labelled library. In the lane on the right, ELT-2 is binding to a single-sequence (containing the tight binding motif ACGATAAGA) double stranded oligonucleotide that has the same length and that serves as a positive control. Bound and free sequences were excised from the gel, PCR-amplified, ligated to sequencing primers and processed for next generation sequencing (Illumina MiSeq). For each sample, sequencing returned between 600,000 and 700,000 reads, of which >95% contained an expected core sequence that allowed it to be identified (i.e. ctgtatgtcgNNNNGATANNNNcctaaccgac) and then extracted using a Perl script. Individual 12-mer motifs within the extracted sequences were counted using a MatLab script.

Supplementary Figure S10



Following Zuo and Stormo [2], the relative affinity of AGATA… motifs relative to TGATA… motifs, assuming independent contributions of individual base pairs to the binding affinity, can be assessed by the following ratio of ratios:

(Number of AGATA…Reads in Bound Fraction)/(Number of AGATA…Reads in Unbound Fraction)
(Number of TGATA…Reads in Bound Fraction)/(Number of TGATA…Reads in Unbound Fraction)

The resulting estimate of relative affinity was calculated to be 0.67 and 0.90 for the two different loadings of ELT-2 protein.

The next more complicated model assumes nearest neighbour interactions between the A (or the T) residue and the residue lying immediately upstream (i.e., both A and T have the constant G on their 3'-side). In this case the above formula was re-applied but first to compare the relative affinities of AAGATA… and ATGATA… sequences, then of CAGATA… and CTGATAA… sequences and so on. The resulting estimate of relative affinity, averaged over the four possible nearest neighbour residues, was 0.91 +/- 0.43 and 0.97 +/- 0.27 for the two different loadings of ELT-2 protein.

In principle, next nearest neighbour effects could also be investigated but we judge that the results are sufficient to demonstrate that ELT-2 does not greatly favour binding of TGATA… over AGATA…

Thermodynamic Modelling of the Gene-Response Curve Relating Relative Levels of Reporter Transcripts to Relative Affinity Constants of the Promoter xxTGATAAxx Sites

Let K_{max} (1/M) be the affinity constant for ELT-2 binding to the TGATAA site with the highest naturally occurring affinity (ACTGATAAGA).

Let K (1/M) be the affinity constant for ELT-2 binding to any other individual TGATAA site.

Let L (M) be the concentration (activity) of free unbound ELT-2 in the nucleus *in vivo*.

Thus the (dimensionless) product K·L is an important descriptor of a simple hyperbolic binding curve. That is, if K·L = 1, the site is half-occupied; if K·L <<1, the site is largely free, and; if K·L >> 1, the site is approaching saturation.

Based on the results of Figure 4 (main text), we initially assume that, in our current SQRIPT reporter system, **both** of the two TGATAA sites in each test promoter must be occupied in order for reporter transcription to occur. We further assume that the two sites are occupied independently and that the probability of a site being occupied can be calculated from simple equilibrium considerations.

Thus, the probability that any particular TGATAA site is occupied is given by:

$$\Theta = K{\cdot}L/(1+K{\cdot}L)$$

Since, in our experiments, both TGATAA sites in the same reporter promoter are identical, the transcript levels from this reporter are proportional to Θ^2, i.e. transcript levels from an individual promoter $= \propto \cdot \Theta^2$ where $\propto$ is a constant (ignoring units because $\propto$ will cancel out).

Thus, the transcript levels produced from any reporter if both TGATAA sites were to be completely saturated $= \propto$ (i.e. $\Theta = 1$ as would occur at "infinite" *in vivo* ELT-2 levels).

And now express all transcript levels as y = Relative Transcript Levels, i.e. the transcript levels produced by an individual reporter relative to the transcript levels produced by the reporter driven by the wildtype promoter and incorporated into the same transforming array. y is the parameter measured in our *in vivo* SQRIPT experiments.

Therefore, the maximum possible relative transcript levels (produced when both TGATAA sites are fully saturated) is given by

$$y_{\Theta=1} = \propto/(\text{ transcript levels produced by wildtype control promoter})$$

and y = the Relative Transcript Level produced by any other reporter is given by

$$y = \propto \cdot \Theta^2/(\text{ transcript levels produced by wildtype control promoter})$$

Thus: $y = y_{\Theta=1} \cdot \Theta^2$

And now replace $K = K_{rel} \cdot K_{max}$ in the expression for Θ;

$$y = y_{\Theta=1} \cdot \Theta^2$$

$$= y_{\Theta=1} \cdot \{K{\cdot}L/(1+K{\cdot}L)\}^2$$

$$= y_{\Theta=1} \cdot \{K_{rel} \cdot K_{max} \cdot L/(1+K_{rel} \cdot K_{max} \cdot L)\}^2$$

This rearrangement allows us to plot y (measured in the *in vivo* SQRIPT experiments) as a function of K_{rel}, which was measured by the competitive band shift experiments. We treat $y_{\Theta=1}$ and the product $K_{max} \cdot L$ as adjustable parameters that can be determined by fitting the data. Optimal values for the two parameters can easily be determined to appropriate precision by numerical trials using the vector commands provided by Matlab. The figure below (Supplementary Figure S11) shows the sums of the squares of the differences between the observed y and a trial value of y calculated assuming particular numerical values of $y_{\Theta=1}$ and $K_{max} \cdot L$. As can be seen, the values

that minimize the sums of the squares of the deviations are: $y_{\Theta=1}$ is roughly 1.4 +/- 0.3; $K_{max}\cdot L$ lies in the range from 5 to 10. Calculated Relative Transcript Levels (y) as a function of K_{rel} (ranging from 0 to 1) are superimposed on the actual SQRIPT data in Figure 5 of the main text, using $y_{\Theta=1} = 1.3$ and $K_{max}\cdot L = 10$. The data are fit somewhat less well using $K_{max}\cdot L = 5$ or 15.

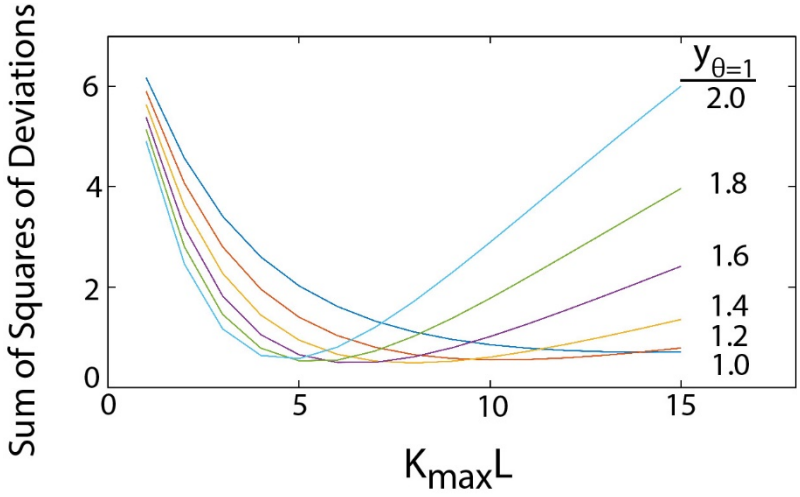


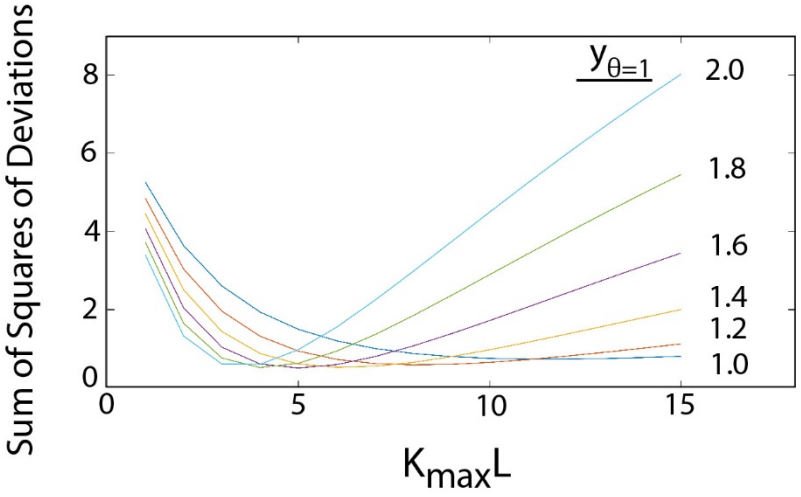
Figure S11

The above model used to predict the y vs. K_{rel} data has assumed that single occupancy of either of the promoter's two TGATAA sites does not activate reporter transcription; only double occupancy is productive. How much do promoter states with single TGATAA occupancy contribute to the observed transcript levels? Consider the strain in which GATA2 has been knocked out (Figure 4 of main text), leaving the reporter promoter under control of the single ACTGATAAGA site, i.e. the optimal site. The mean relative transcript levels produced by this reporter is 0.27 +/- 0.13 (SD), compared to the mean relative transcript levels for the double TGATAA knockout of 0.10 +/- 0.06 (SD), i.e. roughly 15% higher. Thus a revised version of the double-occupancy-only model would stipulate that transcript levels produced if a single TGATAA site were completely occupied $\cong 0.15 \cdot \propto$, where $\propto$ represents the transcript levels produced when both TGATAA sites are completely occupied. Thus the transcript levels produced by a reporter would now be calculated as: $\propto \cdot \Theta^2 + 2 \cdot 0.15 \cdot \propto \cdot \Theta \cdot (1 - \Theta)$, i.e. for single occupancy, either of the two TGATAA sites can be occupied but the other site must be unoccupied. The maximum transcript levels would remain the same, since both sites are then fully occupied. Hence, the expression for relative transcript levels (normalized to the relative transcript levels produced when $\Theta = 1$) is now given by :

$$y = y_{\Theta=1} \cdot \{0.7 \cdot \Theta^2 + 0.3 \cdot \Theta\}$$

$$= y_{\Theta=1} \cdot \{0.7 \cdot [K_{rel}\cdot K_{max}\cdot L/(1+K_{rel}\cdot K_{max}\cdot L)]^2 + 0.3 \cdot K_{rel} \cdot K_{max}\cdot L/(1+K_{rel}\cdot K_{max}\cdot L)\}$$

The same numerical trials are then conducted with Matlab, varying $y_{\Theta=1}$ and $K_{max}\cdot L$ and , as shown in the figure (Supplementary Figure S12), plotting the sums of the squares of the differences between the observed and calculated y values. Compared to the previous figure for the double occupancy only model, the goodness of fit (i.e. the minimum achieved value of the sums of the squares of the deviations) remains largely unchanged but the best values of the two parameters change slightly; $y_{\Theta=1}$ is estimated to be slightly higher at 1.6 and $K_{max}\cdot L$ is now estimated to be slightly lower, closer to 5. The dashed lines in Figure 5 of the main text show how well this revised model fits the experimental data.



Supplementary Figure S12

Sequences of Reporters and Collected Primers

Sequence of the *asp-1* gene used in SQRIPT reporters R1 and R2. Genomic sequence begins 1368 bps upstream of the ATG codon (highlighted in green). The two TGATAA motifs are highlighted in yellow. Protein coding sequence is shown in uppercase. The two single base pair mutations that convert the wildtype *asp-1* sequence into reporters R1 and R2, thereby inserting novel *KpnI* sites, are indicated in cyan.

```
    1 gagacatccc gcccccattt taagtgcata ataagtgtat ttagacaaaa
   51 tccccactgg cgctactcca ccaatcattg agaagaaatt cagccttctt
  101 gtatgaaaaa tgctgaaaaa actgcaaaac ttggccaaaa aactctaaat
  151 cagaacgaaa attcaagaaa ccaacgttaa aatctcccac acaataccca
  201 aaattttcaa aaatcttaca ctaaaataat aataataata ccttctgtac
  251 ttttctacag agttcgtcct aagtcatgtg ctaaactggt cacaataaac
  301 tattcttatt ctctaaaaat ctctaaaaat agctcgtact gtactttccc
  351 ccacctaacc acaattagta caatagtaca attacctttc ggctcctttg
  401 ctatactctt tctcttctaa aaccctcttc TTATCAgtgc aaaagcagta
  451 aaaagtgaaa ctaagaagaa atcggaaaac gagaccaaga accTGATAAg
  501 atttctgaaa ccaattgctg catgaggcta attaaacacg aatgacgtaa
  551 aaggaggagg ggggttggag ccggagtttg ggggtactat aaaagatgag
  601 cggagaggtg gagaagcata ttatcttttt ttgttaggtt tctggttttg
  651 cagaactttc tagaaggttt ttcttttagt aataaaaaaa taattactat
  701 gtttgactat gaagttttaa gggtttttat aaaaacttat gtattaatgt
  751 tcatattaaa atgcttttgg gctgtttttt aggcctaacc ctacttttta
  801 cagtgtattt catgtattac accaggaatt agctttcaag atctctaagg
  851 tttaacaact tttaaagttc aaaactaggc tctagctcgg ttttccggtt
  901 ttttaaagcc tagctgaacc tgagactgtc tgcgaccggt tcaagctgac
  951 ctatactgag ttagactatt ggaaactgcc tcaaatcgtg ttaagcttca
 1001 gacccacata aacgggctaa aatctaccca gaatgtttta gagcgtctta
 1051 aagtccctca gacagtcttt aaacgactta tattgcctga ggatagcagc
 1101 ctgaaaaccg ccctagattt ggcatagctt catttcaaac gtttagatcg
 1151 acctggaggg cttaaaagtt ccgtcctata acagcctaga tatcctgaga
 1201 ataggcaaac cctagaccgt actagaatgt ctcagaaggc cttagacccc
 1251 catccaaaag gttcagatct cctccagact gcttcagata gttttaaatt
 1301 gtcctcagac tcgatgtttc ttcttcagaa ccctccaatc cttccctaaa
 1351 ctcttccttc ttccaggtaA TGCAGACCTT CGTTTTGCTC GCCCTTGTGG
 1401 CGGCATGCTC CGCAGCAGTC ATCCAGGTGC CAACGCACAA GACCGAGTCA
 1451 CTCCGTGCCA AGCTCATCAA GGAGGGCAAG TACACCGCCT TCTTGGCTTC
 1501 ACAGCACGCC GCTCGTGCTC AACAGCTCAA CACCGGATTC CAGCCATTCG
 1551 TCGACTACTT CGATGACTTC TACCTCGGAA ACATCACCCT CGGAACTCCA
 1601 CCACAGCCAG CCACCGTCGT TCTTGACACC GGATCATCCA ACCTTTGGGT
 1651 TATCGATGCC GCATGCAAGA CCCAGGCTTG CAACGGATAC CCAGACTCTG
 1701 GATACACCAA GCAGAAGTTC GACACCACCA AGTCGACCAC CTTCGTGAAG
 1751 GAGACCCGCA AGTTCTCGAT CCAATACGGA TCCGGATCCT GCAACGGATA
 1801 CCTCGGAAAG GATGTTCTTA ACTTCGGAGG ACTCACCGTC CAGTCTCAAG
 1851 AGTTCGGAGT TTCCACCCAC CTCGCCGACG TCTTCGGATA CCAACCAGTT
 1901 GACGGAATCC TCGGACTCGG ATGGCCAGCA CTCGCCGTCG ACCAGGTCGT
 1951 CCCACCAATG CAGAACCTCA TCGCCCAAAA GCAATTGGAC GCTCCACTCT
 2001 TCACTGTCTG GCTTGACCGC AACCTCCAGA TCGCCCAAGG AACCCCAGGA
 2051 GGTCTCATCA CCTACGGAGC CATCGACACC GTCAACTGCG CCAAGCAAGT
 2101 CACCTACGTT CCATTGAGCG CCAAGACCTA CTGGCAATTC CCACTCGACG

                 GGTACC = reporter R1
 2151 CGTTCGCAGT CGGAACCTAC TCTGAGACCA AGAAGGATCA AGTCATCTCC

           GGTA CC = reporter R2
 2201 GACACCGGAA CCTCATGGCT CGGAGCACCA AACACCATCG TCTCCGCCAT
 2251 CGTCAAGCAG ACCAAGGCCG TCTTCGACTG GTCCACCGAG CTTTACACCG
 2301 TCGACTGCTC CACCATGAAG ACCCAGCCAG ACCTCATCTT CACCATCGGA
 2351 GGAGCCCAAT TCCCAGTCAA GTCTGTCGAG TACGTCCTTG ACCTTCAACT
 2401 CGGAGGTGGA AAGTGCGCTC TCGCTGTCTT CTCTATGGGA TCCGGAGGAT
 2451 TCGGACCATC ATGGATTCTT GGAGACACCT TCATCCGTCA ATACTGTAAC
 2501 GTCTACGATA TCGGAAACGG CCAAATCGGA TTCGCCACCG CCGTCCACAA
 2551 GGGATTGTAA gaatggtgtt tttcctgtat ggttatgtat tgctttagtg
 2601 tacaatttgg acacaattct ttgcctcaat tctttgtctc gaataaaatc
 2651 tttaatttct ga
```

FAM labelled self-complementary oligodeoxynucleotides used as probes in competitive band shift assays. TGATAA sites (and reverse complements) are highlighted in yellow and cyan, respectively.

(FAM-labelled AC-TGATAA-GA probe)
oBL76 FAM-aaccctcttcttatcagtgcaaaagcccccgctttgcactgataagaagagggtt

(FAM-labelled GC-TGATAA-TG probe
oBL77 FAM-aaccctcttcattatcagcgcaaaagcccccgctttgcgctgataatgagagggtt

Primers used in SQRIPT cDNA synthesis and PCR amplification of reporter fragments

oBL21 ggaggtctcatcacctacgg (forward primer for amplifying reporter fragment)

oBL22 ctccgagttgaaggtcaagg (reverse primer for amplifying reporter fragment AND primer for reverse transcription reaction)

References for Supplementary Methods

1. Stormo, G.D., Z. Zuo, and Y.K. Chang, *Spec-seq: determining protein-DNA-binding specificity by sequencing.* Brief Funct Genomics, 2015. **14**(1): p. 30-8.
2. Zuo, Z. and G.D. Stormo, *High-resolution specificity from DNA sequencing highlights alternative modes of Lac repressor binding.* Genetics, 2014. **198**(3): p. 1329-43.