www.sciencemag.org/content/350/6258/325/suppl/DC1

Supplementary Materials for

Suboptimization of developmental enhancers

Emma K. Farley,* Katrina M. Olson, Wei Zhang, Alexander J. Brandt, Daniel S. Rokhsar, Michael S. Levine*

*Corresponding author. E-mail: msl2@princeton.edu (M.S.L.); ekfarley@princeton.edu (E.K.F.)

Published 16 October 2015, *Science* **350**, 325 (2015)
DOI: 10.1126/science.aac6948

This PDF file includes

Other Supplementary Material for this manuscript includes the following:
(available at www.sciencemag.org/content/350/6258/325/suppl/DC1)

Materials and Methods

Electroporation:

Adult *Ciona intestinalis* were obtained from M-Rep (San Diego, CA) and maintained in artificial seawater (Instant Ocean) at 18°C under constant illumination. Dechorionation, in vitro fertilization and electroporation was carried out as described in *(34)*. For each electroporation, typically eggs and sperm were collected from 20 adults, 50 µg of DNA was resuspended in 100 µl buffer. Embryos were fixed at the appropriate developmental stage for 15 min in 4% formaldehyde. The tissue was then cleared in a series of washes of 0.01% Triton-X in PBS. Samples were mounted in 50% glycerol in PBS with 2% DABCO compound for microscopy. Differential interference-contrast microscopy was used to obtain transmitted light micrographs with a Zeiss Axio Imager A2 using the ×40 EC Plan Neofluar objective, the same microscope was used to obtain GFP images. All constructs were electroporated at least twice in two completely separate experiments (biological replicates).

Counting embryos:

For each experiment once embryos had been mounted on slides, slide labels were covered with thick tape and randomly numbered by a lab member not involved in this project and randomized. In each experiment all comparative constructs were present along with a slide with WT Otx-a as a reference. The x-cite was turned on for 1hr before analysis to ensure the illumination intensity was constant. 100 embryos were counted for each slide, 50 by EF and 50 by KO independently to ensure no deviation from the set scale. Weak, moderate and strong expression definitions are based on exposure time required for 10% saturated pixels. Strong <500ms, moderate <800ms and weak >800ms. Statistical significance of counting data was determined using chi-square test.

Acquisition of images:

For enhancers that were being compared, images were taken on the same day and from electroporations done on the same day, using identical settings. For images, embryos were chosen that represented the average from counting data. Images are rotated and cropped but have no other manipulations.

Library construction:

A DNA oligo of the Otx-a enhancer shown in Fig. 1B was ordered from Elim biosciences with PCR amplification arms containing BseRI sites (5'CATCATGACGAGGAGAAACCAGCAC, 3' AAACCATTCTCCTCTTCCATCAT). This was cloned into the custom designed SEL-Seq (Synthetic Enhancer Library-Sequencing) vector using type II restriction enzyme BseRI. After cloning the library was transformed into bacteria (MegaX DBH10 electrocompetent cells), the culture was grown up till an OD of 1 was reached. DNA was extracted using standard mini-prep plasmid extraction methods. A 30bp barcode was added to this library by PCR using primers that match the vector backbone and contain 15bp Ns followed by EcoR1 site. DpnI digest was carried out to remove all non-barcoded input plasmid, EcoRI digest was used to create a linear barcoded enhancer library, and this was then ligated and transformed into bacteria and grown overnight. The DNA library was extracted from the bacteria using Machete-Nagel Nucelobond Xtra Midi kit.

Enhancer to barcode assignment:

To determine which enhancer was associated with which barcode tag we removed the intervening sequence containing promoter and GFP by restriction digest with XbaI and SpeI and re-ligated the remaining vector to get enhancer and barcode in close proximity. We PCR amplified the region of interest with 8 cycles of PCR, and size selected the PCR product using AMPure beads. We ran the purified sample on a bioanalyser to check the product was the correct size and then sequenced the library using HiSeq 2500 Rapid Run mode.

Electroporation of library into embryos:

The library was electroporated into ~5000 fertilized eggs. Embryos developed until 5hrs 10 min at 22ºC. The embryos were then put into trizol and RNA was extracted following manufactures instructions (Life technologies). The RNA was DNase treated using Turbo DnaseI from Ambion following standard instructions. Poly-A selection was used to obtain only mRNA using poly-A biotinylated beads as per instructions (Dyna-beads Life technologies). The mRNA was then split in two. Half of the mRNA was used in a RT reaction that specifically selected the synthetic tag mRNA, the other half was put in a –RT reaction to ensure no DNA contamination was present (Transcriptor High Fidelity Roche). The RT product was PCR amplified using 8 PCR cycles. The PCR product from the –RT reaction was analyzed on a bioanalyser no product was seen. The PCR product from the RT reaction was size selected using Agencourt AMPure Beads (Beckman Coulter), checked for quality and size on the bioanlayser and sent for sequencing on the HiSeq2500 Rapid Run mode. Two biological replicates were sent for sequencing.

SEL-Seq data analysis:

In this study, we established a new analysis framework to systematically analyze SEL-Seq data and uncover motifs of active enhancers de novo.

We first analyzed the DNA library and constructed a dictionary of unique barcode tag-enhancer pairs by allowing 2 bp mismatches in the ~69 bp enhancers to buffer the sequencing mistakes. Since we are sampling a tiny fraction of the huge pool of possible random enhancers (1.4e+8 out of 3.2e+29), the chance of having any 2 enhancers that are truly differentiated by 1 or 2 bp is negligible (2.3e-9). Therefore, it is entirely reasonable to believe that any sequences within 2 bp mismatches are caused by sequencing mistakes. In case an enhancer has more than one isoforms with $\leq$ 2 bp mismatches between each other, it would be crucial for the downstream analysis to distinguish the correct sequences of the enhancers from those with sequencing mistakes. Therefore, for each enhancer we defined the isoform with the most reads in the DNA library as the dominant isoform and used it for the following analysis. If more than one barcode tag was associated with a single enhancer, the maximum reads of these tags were used in the following RNA-seq analysis, similar signals from two tags that both matched one enhancer severed as an internal control. We removed the tags that were attached to multiple enhancers. The dictionary contains 2,534,802 enhancers that are uniquely mapped to one or more barcode tags.

We parsed the 2 biological replicates of RNA-seq data and calculated the expression level as reads per million total reads (RPM) for each barcode tag. By referring to the tag-enhancer dictionary we assigned the expression level to the uniquely mapped enhancers (dominant isoforms). Totally 163,708 enhancers were detected by RNA-seq and 21,799 of them were defined as active enhancers by RPM ≥ 4 in either of the 2 replicates. Enriched sequence motifs (6 – 8 bp) in the active enhancers were identified de novo using HOMER software *(35)* by comparing the DNA library as a background. Due to the complexity of the enhancer library our two biological replicates sampled the library, such that some enhancers are found only in one library. However, the motifs identified by each replicate separately are the similar to those identified when both data sets are combined. Our automated analysis framework can easily be applied to similar studies in the future.

Scoring binding sites in *C.i* and *C.s* Otx-a enhancer:

The WT Otx-a enhancers of *C. intestinalis* and *C. savignyi* were scanned for 6 – 8 bp extended GATA and Ets motifs using HOMER software *(35)*. The motif instances were scored as the log odds probabilities: score = $\sum_{i=1}^{n} \log(\frac{P_i}{0.25})$, where p_i is the probability of the nucleotide in position i based on position weight matrices. Each GATA or Ets motif instance in WT Otx-a enhancer was assigned with the maximum score among 6 – 8 bp motifs. Then a relative score was calculated by comparing it with the 8 bp optimal motif.

Scoring relative affinities:

Relative affinities were calculated using median signal intensities of the universal protein binding microarray (PBM) data for mouse ETS-1 *(10)* and GATA-6 *(11)* proteins from UniProbe database (http://thebrain.bwh.harvard.edu/uniprobe/index.php). The percentage of relative affinities represent the fold changes of median signal intensities of the native 8-mer motifs comparing to the optimal 8-mer motifs for optimal Ets and GATA, respectively.

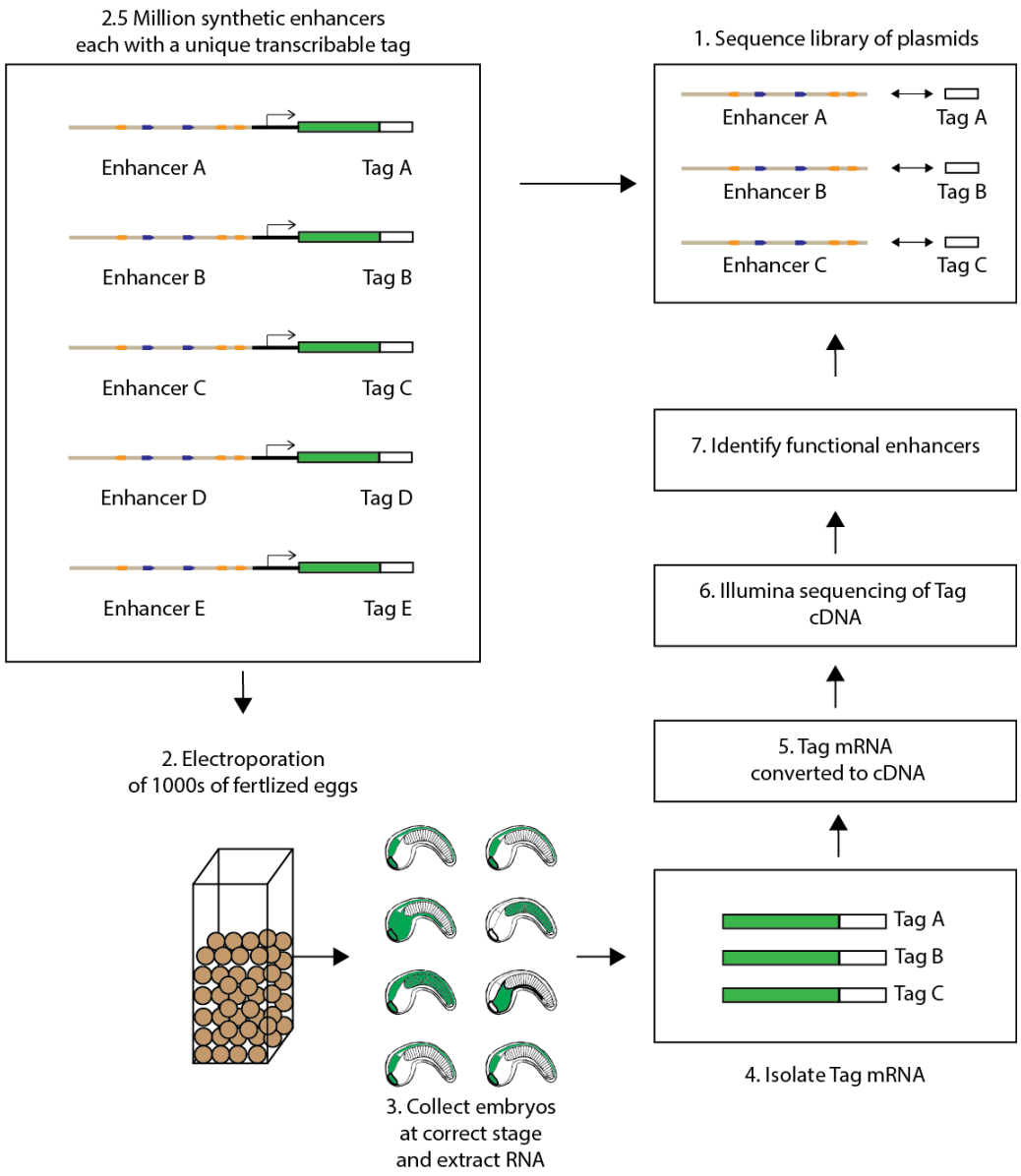


Fig S1. The Synthetic Enhancer Library – Seq (SEL-Seq) experimental design.

(**A**) 2.5 million synthetic enhancers were attached to a minimal promoter, GFP coding sequence, and 30 bp barcode tags. Each enhancer has a unique tag identifier. **Step 1**. The library was sequenced to assign enhancer variants to barcodes (see methods for more information). **Step 2**. The library of enhancer variants was electroporated into thousands of fertilized *Ciona* eggs. **Step 3**. The electroporated embryos developed until a stage at which the endogenous enhancer becomes active. Embryos were then collected and RNA was extracted. **Step 4**. Isolated mRNAs were reverse transcribed using a primer found only in the library to create cDNA from the transcribed barcodes. **Step 5**. cDNAs were amplified using PCR primers that match synthetic barcode mRNAs. **Step6**. cDNAs were subject to deep sequencing. **Step 7**. Each transcribed barcode tag was assigned to its respective enhancer. Active enhancers were used for further analysis (e.g., see below).

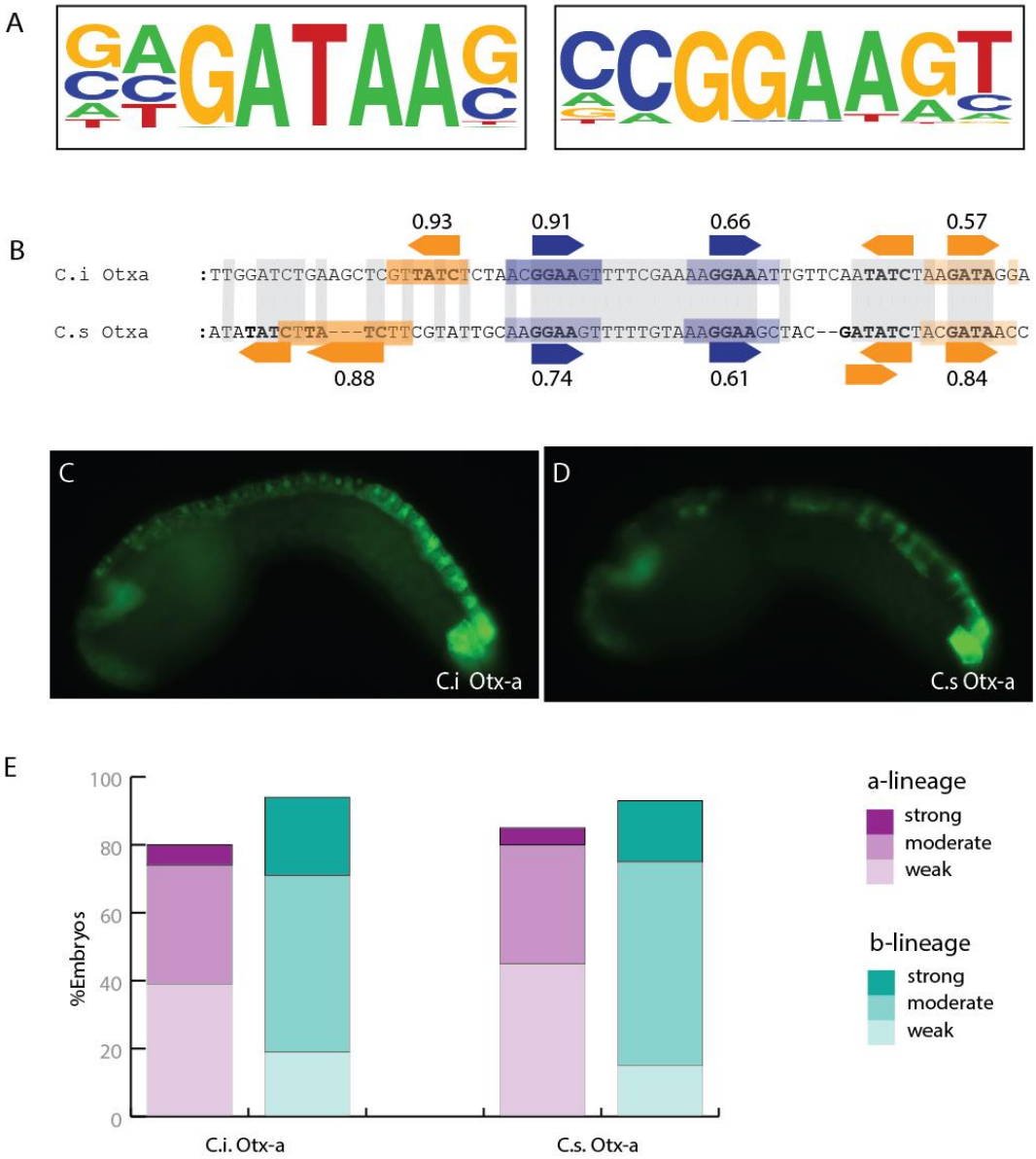
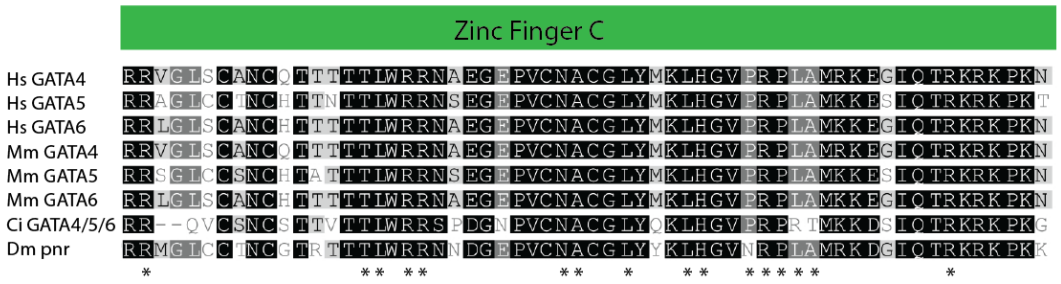


Fig. S2. Motifs identified in functional enhancers match the PWMs for Ets and GATA transcription factors. (**A**) Position Weight Matrixes of the motifs enriched in functional enhancers transcribed ≥ 4 RPM. These PWMs are similar to those identified by high throughput in vitro binding assays (see Fig. S3 and S4). (**B**) Otx-a enhancer sequence in *Ciona intestinalis* (*C.i*) and *Ciona savignyi* (*C.s*) grey boxes highlight conservation between the species, orange boxes highlight sequences matching the GATA PWM; blue highlights sequences matching the ETS PWM. Both *C.i.* and *C.s*. have a mix of good and poor matches to the identified PWMs. Scores of match to PWM are shown above binding sites if the score is above 0.5 (**C**) WT *C.i* Otx-a enhancer drives expression in the palps, anterior brain, dorsal nerve cord and dorsal epidermis along with two tail-tip muscle cells. (**D**) *C.s* Otx-a enhancer drives expression in the same locations as *C.i* Otx-a (**E**) Embryo counts after electroporation with *C.i.* Otx-a or *C.s*. Otx-a. The expression pattern and levels for *C.i.* and *C.s*. are very similar.

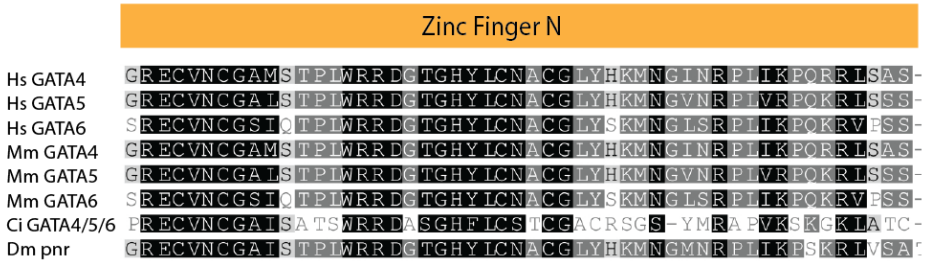
	Helix 1	Helix 2	Helix 3	Strand 4	PWM PCC	PWM
	1 2	3 4	5 6 7 8 9 10 11 12	13 14 15		
Hs-ETS1	LWQF	WGKRK	YEKLSRGLRYYY	RYVY	0.98	
Mm-ETS1	LWQF	WGKRK	YEKLSRGLRYYY	RYVY	0.95	
Ci-ETS1/2	LWQF	WGIRK	YEKLSRGIRYYY	RYVY	1.00	
Dm-PNT1	LWQF	WGIRK	YEKLSRGLRYYY	RYVY	0.98	

Fig. S3. ETS1 DNA binding domain sequence and specificity are conserved from Drosophila to Humans.
ETS-domain showing amino acid residues that contact specific nucleotides, based on the published crystal structures of ETS-domain DNA complexes *(10, 33, 36-39)*. Amino-acid residues contacting DNA are numbered from 1 to 15. PWM Pearson correlation coefficient (PWM PCC) between Ci-ETS1/2 and Human, mouse and Drosophila PWMs are shown. PWMs are compared using Pearson's correlation coefficient (PCC) by converting each matrix into a vector of values. PCC are based on the flanking dinucleotides since the 4 bp core motifs were fixed in our experiments. The PWM graphic is also shown.

A



B



C

	PWM PCC	PWM
Ci GATA4/5/6	1.00	
Mm GATA4	0.96	
Mm GATA5	0.77	
Mm GATA6	0.93	
Dm pnr	0.97	

Fig. S4. Comparison of GATA4/5/6 DNA binding domains and PWM across species. (**A**) The Zinc finger C has strong conservation between human, mouse and Drosophila. Stars show the amino acid residues that directly contact DNA *(40)*; these are strongly conserved across species. (**B**) The N Zinc finger is poorly understood and shows less sequence conservation than the C finger. (**C**) The PWM for Ci- GATA4/5/6, and the Pearson correlation coefficient for the PMWs of mouse GATA4, 5 and 6 and the Drosophila homolog, pnr (pannier). No Human PWM data are available. Mm GATA 5 shows lower correlation with Ci GATA4/5/6 (PCC=0.77), GATA4/5/6 homolog in *Drosophila melanogaster* Dm pnr (PCC=0.80), Mm GATA4 (PCC=0.87) and GATA6 (PCC=0.79) based on UniProbe PBM data *(11, 41)*.

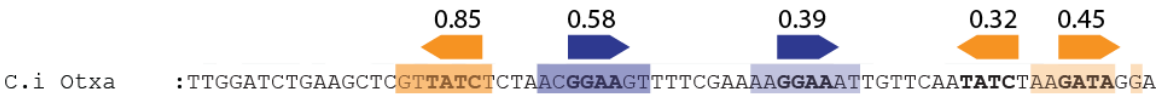


Fig. S5. Relative affinities of native Otx-a binding sites. Sequence of the *C.i.* Otx-a enhancer, indicating the relative affinities of each of the five binding sites
For example, the Ets-2 binding site (0.39) is predicted to contain a binding affinity that is 2.5-fold lower than the optimal motif. Relative binding affinities were calculated using median signal intensities of the universal protein binding microarray (PBM) data for mouse ETS-1 *(10)* and GATA-6 *(11)* proteins from the UniProbe database (http://thebrain.bwh.harvard.edu/uniprobe/index.php). The percentage of relative affinities represent the fold changes of median signal intensities of the native 8-mer motifs comparing to the optimal 8-mer motifs (CCGGAAGT and GAGATAAG for optimal ETS and GATA, respectively). Note that these numbers are smaller than those obtained by a traditional comparison of PWMs. The current method involves a direct determination of relative binding frequencies based on the analysis of the raw DNA binding datasets.

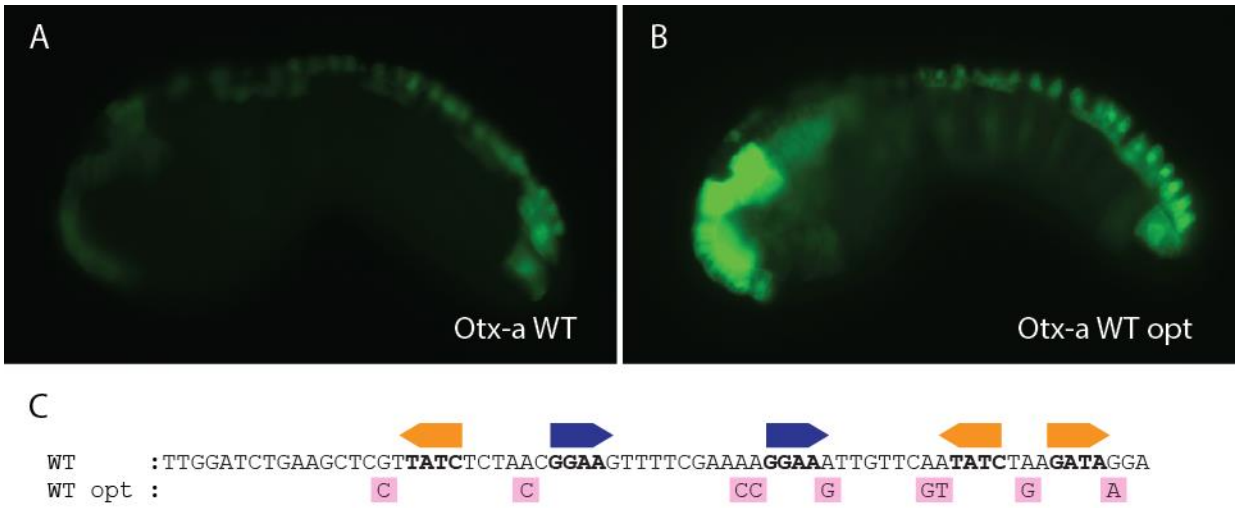


Fig. S6. Optimizing the sites in WT enhancer leads to ectopic expression in FGF signaling regions. (**A**) The native *C.i.* Otx-a enhancer drives expression in the palps, anterior brain, dorsal nerve cord and dorsal epidermis along with two tail-tip muscle cells. (**B**) When this enhancer is mutated to create optimal matches to the identified PWM, we observe ectopic expression in the notochord, endoderm and posterior nerve chord. (**C**) Sequences of WT enhancer and changes to create WT opt (enhancer with optimized flanking) shown in pink.

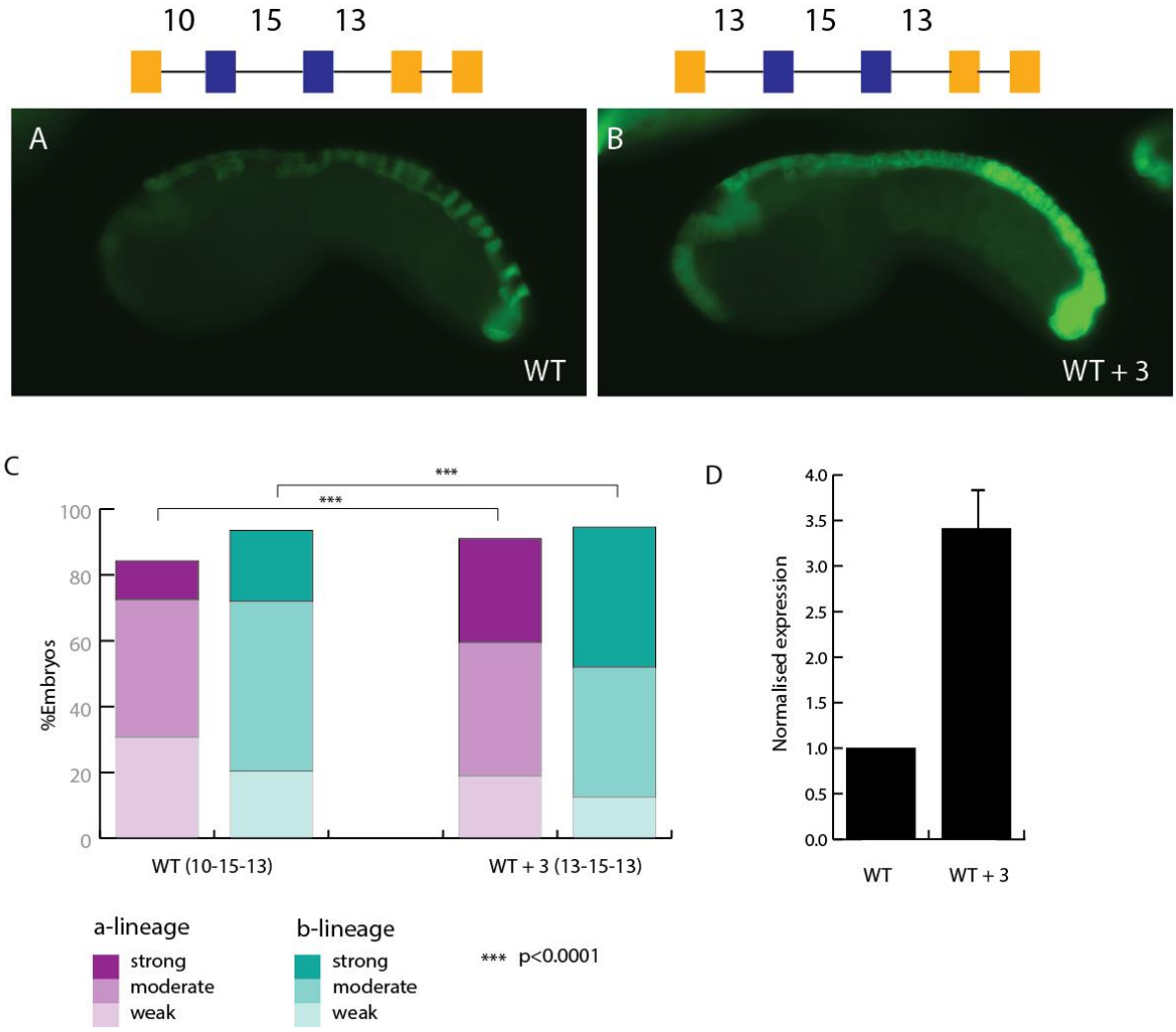


Fig. S7. Spacing of binding sites is important for expression.
(**A**) The WT Otx-a enhancer drives expression in the palps, anterior brain, dorsal nerve cord and dorsal epidermis along with two tail-tip muscle cells. (**B**) WT Otx-a + 3, contains a 3 bp insertion between the GATA-1 and ETS-1 sites. This insertion leads to a significant increase in endogenous neural expression. Both images are taken at the same exposure times with the same setting, 500ms (**C**) Graph showing embryo counts after electroporation with WT and WT+3 Otx-a enhancer variants. Levels of expression were determined by exposure time, two biological replicates were counted double-blind, with 50 embryos per replicate. n=200 for each construct. (**D**) qPCR results showing expression levels of GFP mRNAs with the WT and WT+3 enhancers. Three biological replicate electroporations were carried out at 5hrs 10. Results were normalized to three house-keeping genes and GFP levels produced by the WT enhancer. Error bar shows standard deviation; the WT sample does not have an error bar as each WT replicate was normalized to 1.

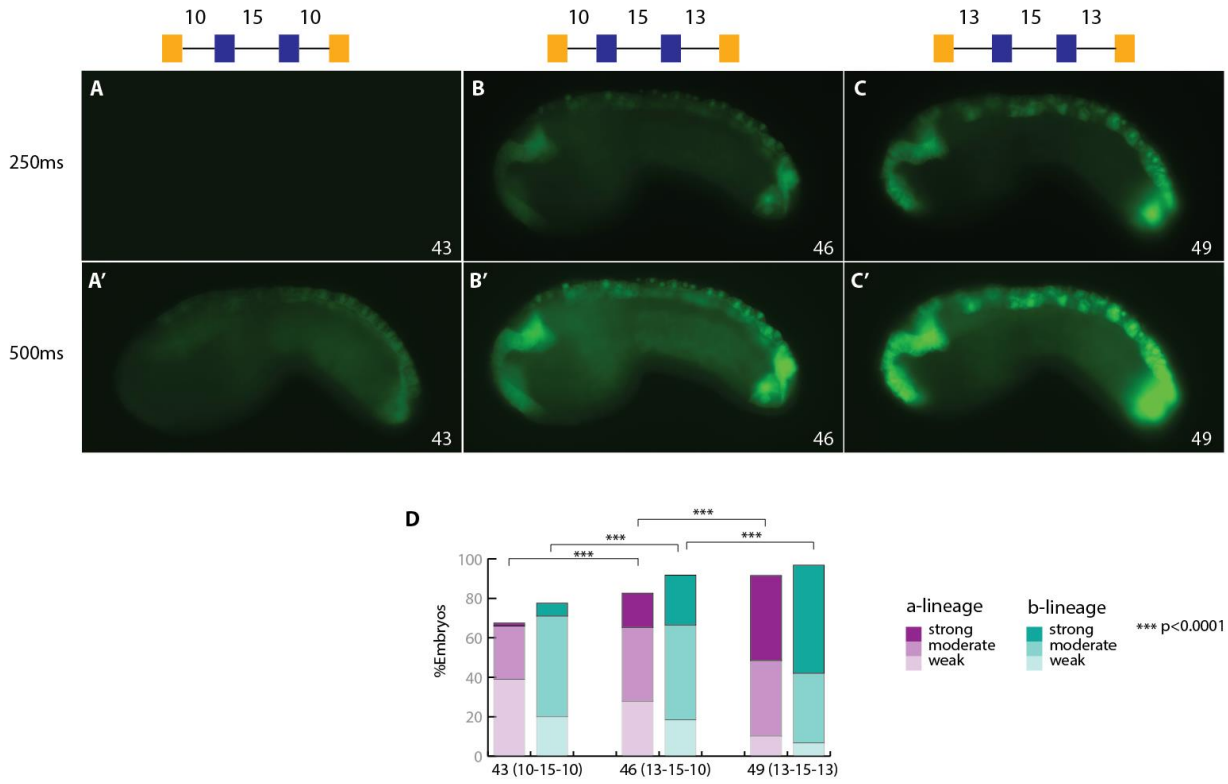


Fig. S8. Spacing of binding sites is important for the levels of expression generated by a synthetic Otx-a enhancer variant.
Synthetic variant Otx-a 46 contains 2 GATA sites and 2 ETS sites, with successive spacing of 10 bp, 15 bp, and 13 bp (10-15-13). This synthetic enhancer produces levels of expression that are similar to those seen for the native enhancer (panels **B** and **B'**). This variant was created by replacing GATA2 with the GATA3 sequence. The same embryo was photographed for 250 ms and 500 ms (**B** and **B'**), respectively. Deletion of 3 bp from the interval between ETS-2 and GATA-3 causes a significant reduction in the levels of expression (**A**, **A'**). In contrast, insertion of 3 bp between GATA-1 and ETS-1 causes a significant increase in the levels of expression (**C**, **C'**). This increase was quantified by qPCR assays (see Fig. S7). (**D**) Graph showing embryo counts following electroporation with the different spacing variants. The changes in expression between variants is significant to $p<0.0001$. Levels of expression were determined by exposure time, for each replicate two biological replicated were counted double-blind, with 50 embryos per count. n=200 for each construct.

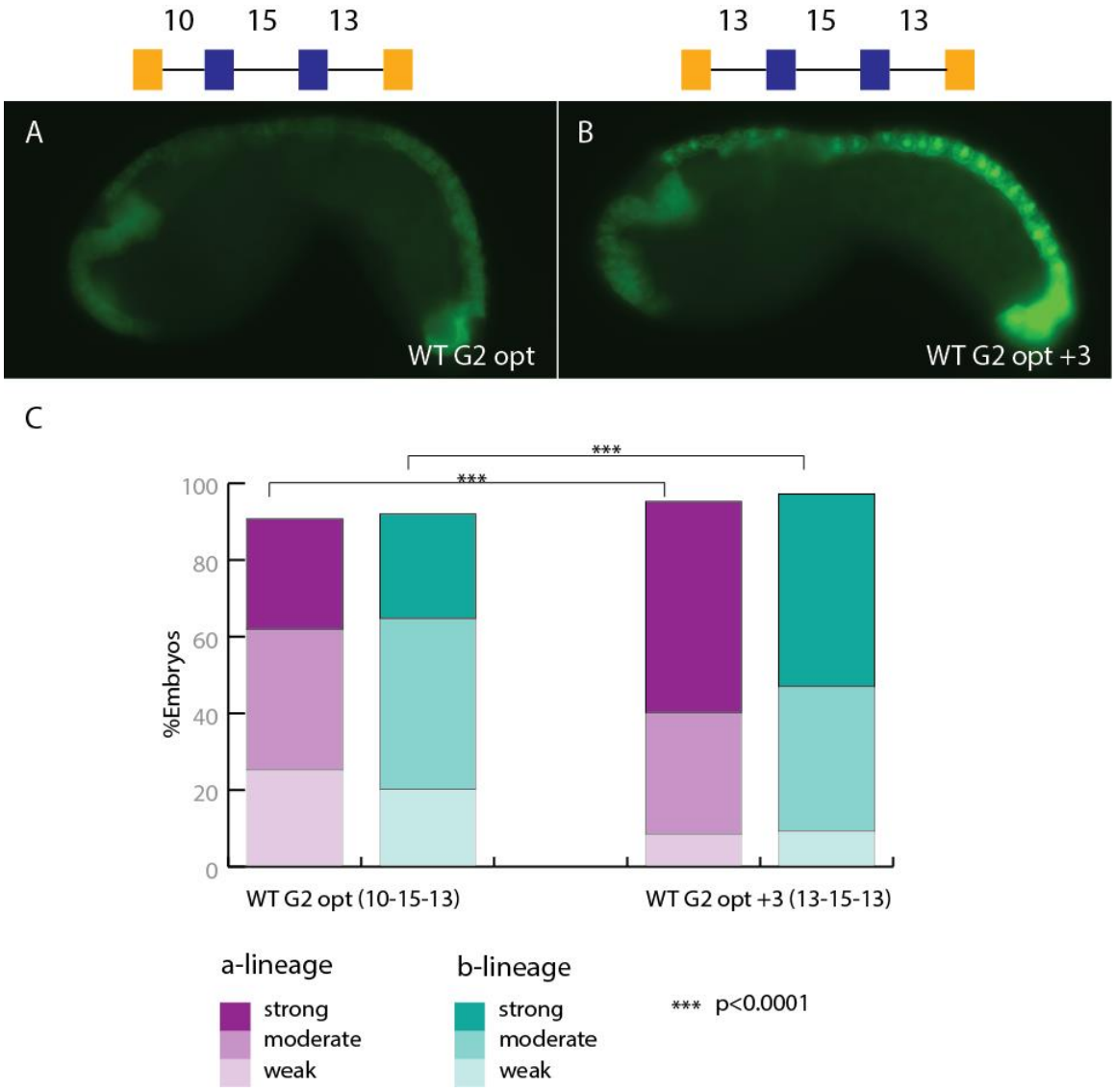
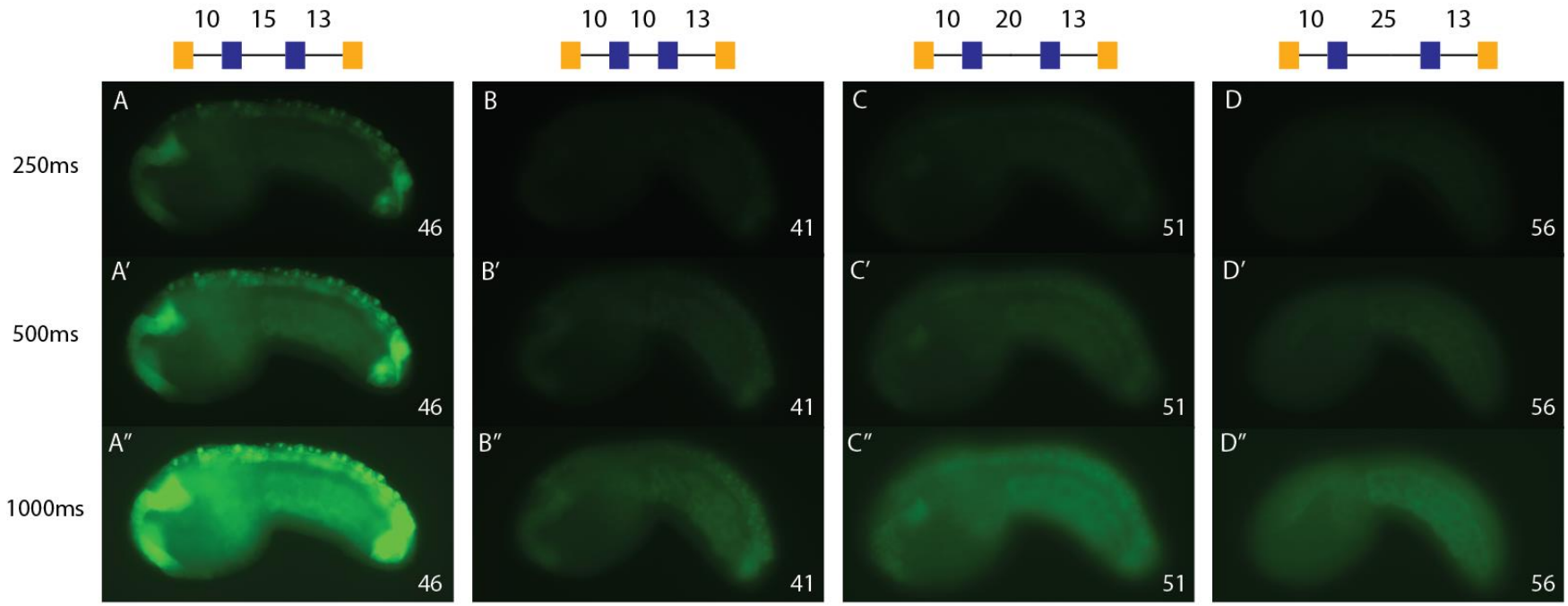


Fig. S9. Spacing of optimal binding sites is important for the levels of gene expression. Addition of 3 bp between GATA1 and ETS1 increases levels of neural expression in WT G2 opt enhancer.

(**A**) Synthetic variant of WT enhancer containing the GATA-3 sequence at the position of the native GATA-2 site. This GATA site was mutagenized to create optimal dinucleotide sequences in the flanking regions. (**B**) Same as (A) except that 3 bp was inserted between the GATA-1 and ETS-1 sites. This change in spacing leads to a significant increase in expression. All images are taken at the same exposure times 250ms with the same settings on the same day. (**C**) Graph showing embryo counts following electroporation with the preceding enhancer variants. The augmented levels of expression obtained upon insertion of 3 bp is significant to $p<0.0001$. Levels of expression were determined by exposure time, for each replicate two biological replicated were counted double-blind with 50 embryos each. n=200 for each construct.


10 15 13
10 10 13
10 20 13
10 25 13
250ms
500ms
1000ms
A
A′
A″
B
B′
B″
C
C′
C″
D
D′
D″
46
41
51
56

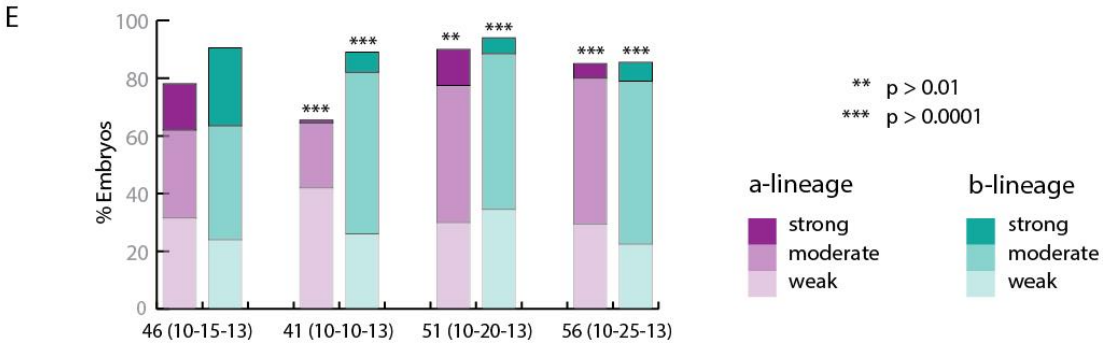
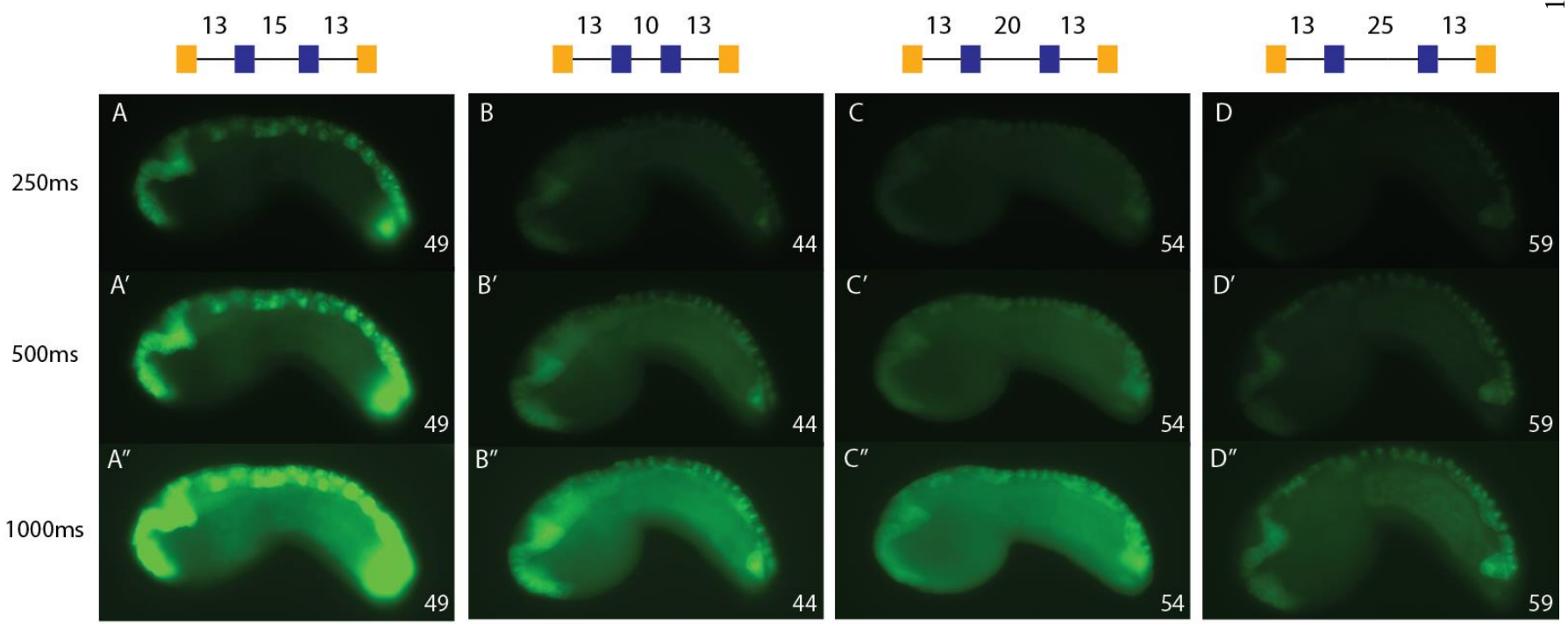

E
% Embryos
100
80
60
40
20
0
***
**
46 (10-15-13)
41 (10-10-13)
51 (10-20-13)
56 (10-25-13)
** p > 0.01
*** p > 0.0001
a-lineage
strong
moderate
weak
b-lineage
strong
moderate
weak

Fig. S10. Spacing of binding sites impacts levels of expression. Altering spacing between the ETS sites from 15 to 10, 20 or 25bp reduces expression in the context of Otx-a 46.

(**A**) Expression of enhancer Otx-a 46 (10-15-13 spacing) at 250ms, (**A'**) 500ms and (**A''**). 1000ms (**B**) Expression of enhancer Otx-a 41 (10-10-13 spacing) at 250ms, 500ms (**B'**) and 1000ms (**B''**). Expression is much weaker than 46. (**C**) Expression of enhancer Otx-a 51 (10-20-13 spacing) at 250ms, 500ms (**C'**) and 1000ms (**C''**). Expression is much weaker than 46. (**D**) Expression of construct Otx-a 56 (10-25-13 spacing) at 250ms, 500 (**D'**) and 1000ms (**D''**). Expression is much weaker than 46. (**E**) Counting of embryos electroporated with 46, 41, 51 and 56. A spacing of 15bp between the two ETS give the highest expression with any changes from this WT spacing significantly reducing expression $p<0.0001$. Levels of expression were determined by exposure time, for each replicate two biological replicated were counted blind. n=200 for each construct


13 15 13
13 10 13
13 20 13
13 25 13
250ms
500ms
1000ms
A
A′
A″
B
B′
B″
C
C′
C″
D
D′
D″
49
44
54
59

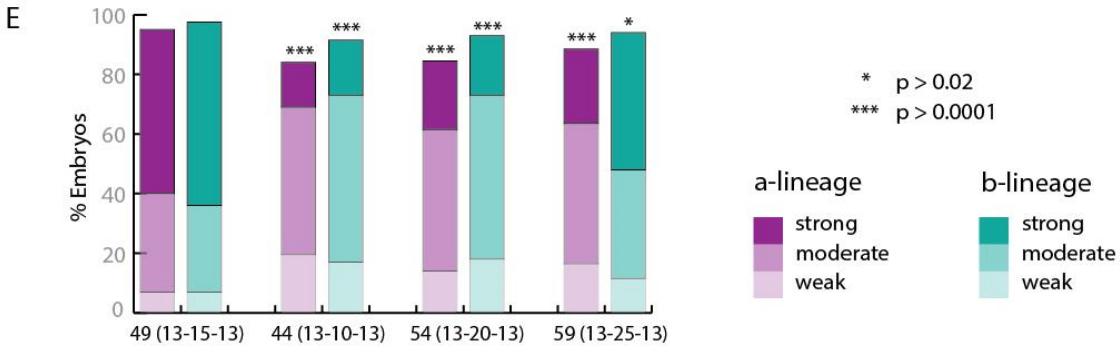

E
% Embryos
100
80
60
40
20
0
***
*
49 (13-15-13)
44 (13-10-13)
54 (13-20-13)
59 (13-25-13)
* p > 0.02
*** p > 0.0001
a-lineage
strong
moderate
weak
b-lineage
strong
moderate
weak

Fig. S11. Spacing of binding sites impacts levels of expression. Altering spacing between the ETS sites from 15 to 10, 20 or 25bp reduces expression in the context of Otx-a 49.

(**A**) Expression of enhancer Otx-a 49 (13-15-13 spacing) in the same embryo at 250ms, 500ms (**A'**) and 1000ms (**A''**). (**B**) Expression of enhancer Otx-a 44 (13-10-13 spacing) in the same embryos at 250ms, 500ms (**B'**) and 1000ms (**B''**). Expression is much weaker than 49. (**C**) Expression of construct Otx-a 54 (13-20-13 spacing) in the same embryos at 250ms, 500ms (**C'**) and 1000ms (**C''**). Expression is much weaker than 49. (**D**) Expression of construct Otx-a 59 (13-25-13 spacing) in the same embryo at 250ms ,500ms (**D'**) and 1000ms (**D''**). Expression is much weaker than 49. (**E**) Counting of embryos electroporated with 49, 44, 54 and 59, the WT spacing, 15bp between the two ETS give the highest expression, with any change in spacing between the two ETS significantly reducing expression, $p<0.0001$. Levels of expression were determined by exposure time, for each replicate two biological replicated were counted blind. n=200 for each construct. Note that better spacing between the ETS and GATA sites, namely, 13bp rather than 10bps seen in the 49 vs 46 enhancer, can compensate for suboptimal spacing between the ETS as 44, 54 and 59 all show higher expression than 41, 51, 56 (Compare FigS11 with Fig S10).

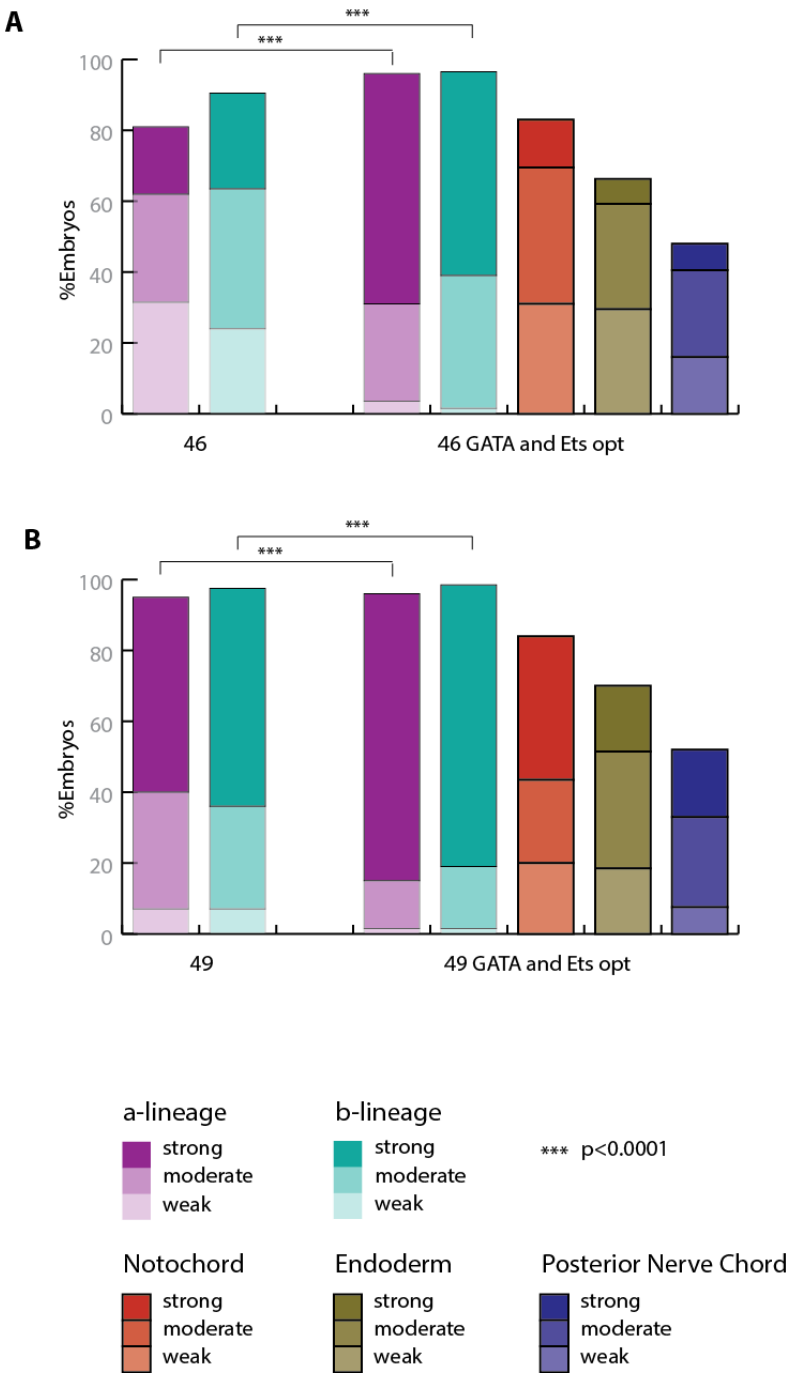


Fig. S12. Counting for 46, 46opt, 49, and 49opt. (**A**) Counting of embryos electroporated with enhancer 46 and 46opt, with optimized GATA and ETS flanking. (**B**) Graph showing counting of embryos electroporated with 49 and 49opt, with optimized GATA and ETS flanking. Optimizing the GATA and ETS gives stronger endogenous expression and high levels of ectopic expression for both 46 and 49. Levels of expression were determined by exposure time, for each replicate two biological replicated were counted blind. n=200 for each construct.

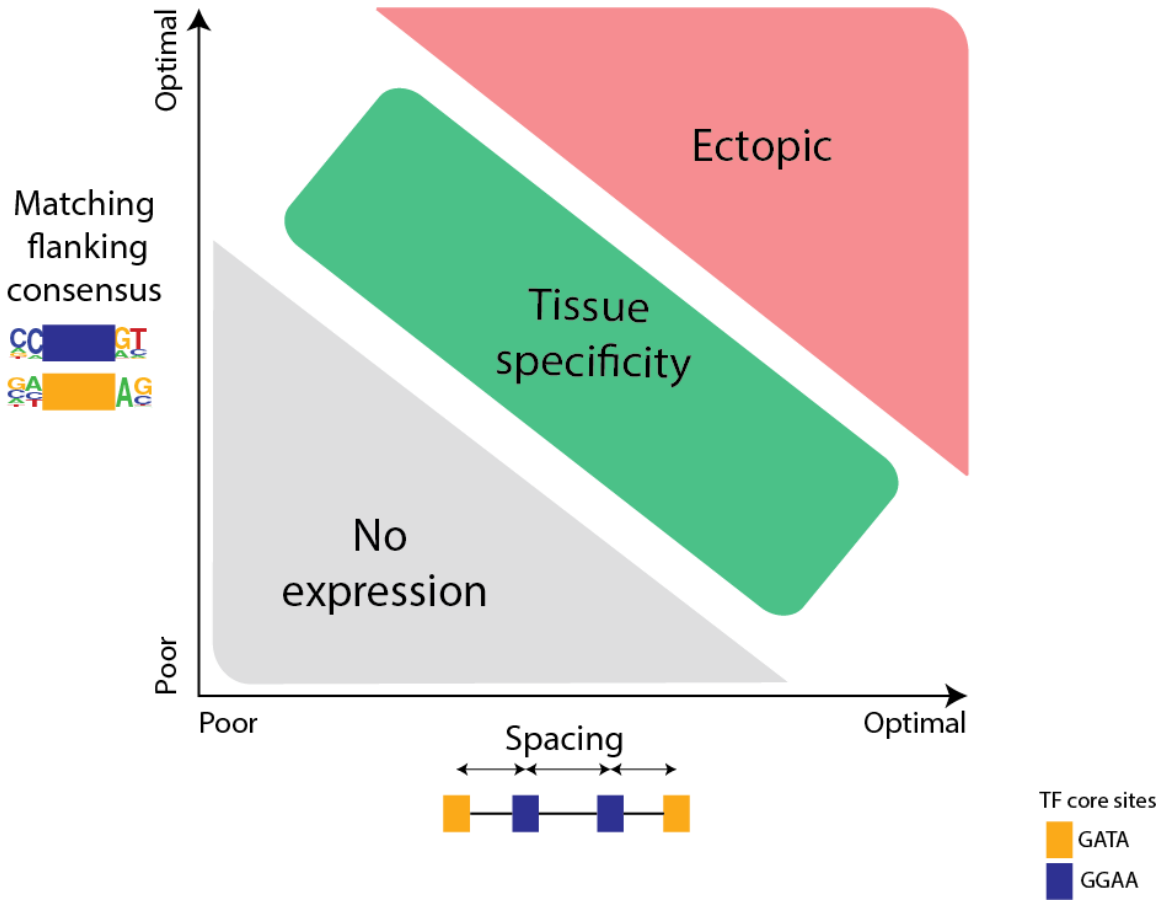


Fig. 13. Tissue specificity requires multiple tiers of enhancer suboptimization. Schematic showing match to identified flanking consensus on Y-axis and spacing on X-axis. Core binding sites are shown as orange and blue boxes. Optimal flanking and spacing leads to ectopic expression, while poor spacing and flanking results in no expression. Suboptimal spacing and flanking provide tissue specific expression.

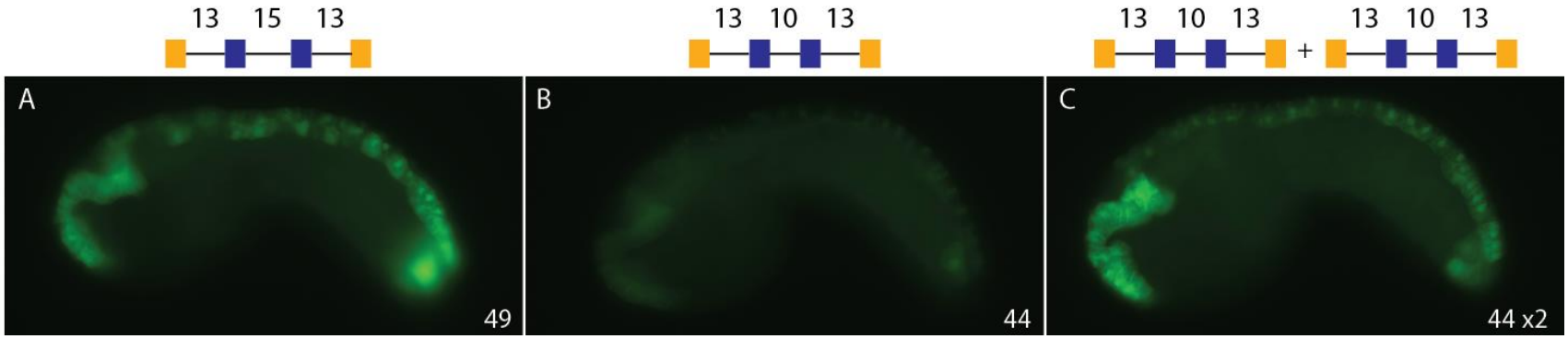

13 15 13
13 10 13
13 10 13 + 13 10 13
A
49
B
44
C
44 x2

Fig. S14. Clusters of weak suboptimal enhancers show augmented expression relative to single copy, and can circumvent weaker expression from suboptimal enhancers.
(**A**) Construct with WT flanking and more optimal spacing shows strong expression in neural plate. (**B**) Same as A, but spacing between two Ets sites is now 10bp rather than 15. This change in spacing leads to reduced expression. (**C**) Multimerizing the suboptimal construct leads to expression equivalent to that seen for the more optimal enhancer Otx-a 49 (compare C with A). Images all taken at 250ms with no adjustment to the images, all setting for each image the same, all images taken on same day.

Table S1. List of enhancer variants with transcribed barcodes in Otx-a high-throughput enhancer screen. (Separate file)

Sequences are in reverse complement orientation. Expression is shown in Reads Per Million (RPM).

Table S2. Description of all individually tested constructs tested to validate the library sequencing data. (Separate file)

Table shows scores of expression for each tested construct in the endogenous Otx-a location and ectopic expression, 0 = no expression, 5 = strong expression. For each construct two biological repeats were carried out, with 100 embryos analyzed in each repeat.

Table S3. Summary of findings for enhancer spacing experiments. Scores for expression were determined based on counting for each construct, 0 = no expression, 25 = highest expression. The scale in this table is not comparable to Table S2 scores. A new scale was required for Table S3 as expression levels were much higher than those seen in Table 2.

Construct name	Spacing	Types of sites	Endogenous expression level	Ectopic expression level
		On WT background		
WT	10--15--7--13	g1, e1, e2, g2, g3	4	0
WT + 3	13--15--7--13	g1, e1, e2, g2, g3	12	0
		Minimal with GATA 2 optimized, GATA 3 removed		
WT G2 opt	10--15--13	g1, e1, e2, opt g	8	0
WT G2 opt + 3	13--15--13	g1, e1, e2, opt g	16	0
		Minimal with GATA 3, GATA 2 removed		
32	8--8--8	g1, e1, e2, g3	1	0
37	10--9--10	g1, e1, e2, g3	3	0
42	10--11--13	g1, e1, e2, g3	4	0
43	10--15--10	g1, e1, e2, g3	4	0
46	10--15--13	g1, e1, e2, g3	8	0
49	13--15--13	g1, e1, e2, g3	16	0
41 (46 E-5)	10--10--13	g1, e1, e2, g3	2	0
51 (46 E+5)	10--20--13	g1, e1, e2, g3	3	0
56 (46 E+10)	10--25--13	g1, e1, e2, g3	3	0
44 (49 E-5)	13--10--13	g1, e1, e2, g3	5	0
54 (49 E+5)	13--20--13	g1, e1, e2, g3	7	0
59 (49 E+10)	13--25--13	g1, e1, e2, g3	11	0
46 opt GATA	10--15--13	opt g, e1, e2, opt g	18	2
46 opt Ets	10--15--13	g1, opt e, opt e, g3	16	4
46 opt Ets and GATA	10-15--13	opt g, opt e, opt e, opt g	19	6
49 opt GATA	13--15--13	opt g, e1, e2, opt g	24	3
49 opt Ets	13--15--13	g1, opt e, opt e, g3	14	3
49 opt Ets and GATA	13--15--13	opt g, opt e, opt e, opt g	25	12

Key to types of sites:
g1 = WT GATA1 site (GAGATAAC)
e1 = WT Ets1 site (ACGGAAGT)
e2 = WT Ets2 site (AAGGAAAT)
g2 = WT GATA2 site (TAGATATT)
g3 = WT GATA3 site (AAGATAGG)
opt g = optimal GATA from PWM (GAGATAAG)
opt e = optimal Ets from PWM (CCGGAAGT)

Table S4. Sequence of all spacing constructs tested.

Name	Enhancer Sequence
WT	TTGGATCTGAAGCTCGTTATCTCTAACGGAAGTTTTCGAAAAGGAAATTGTTCAATATCTAAGATAGGA
WT + 3	TTGGATCTGAAGCTCGTTATCTCTACTAACGGAAGTTTTCGAAAAGGAAATTGTTCAATATCTAAGATAGGA
WT G2 opt	TTGGATCTGAAGCTCGTTATCTCTAACGGAAGTTTTCGAAAAGGAAATTGTTGCTTATCTCAGGTAGGA
WT G2 opt + 3	TTGGATCTGAAGCTCGTTATCTCTAGTAACGGAAGTTTTCGAAAAGGAAATTGTTGCTTATCTCAGGTAGGA
32	GTTATCTCACGGAAGTAAGGAAATAAGATAGG
37	GTTATCTCTAACGGAAGTTAAGGAAATTCAAGATAGG
42	GTTATCTCTAACGGAAGTTTAAAGGAAATTGTTCAAGATAGG
43	GTTATCTCTAACGGAAGTTTTCGAAAAGGAAATTCAAGATAGG
46	GTTATCTCTAACGGAAGTTTTCGAAAAGGAAATTGTTCAAGATAGG
49	GTTATCTCTACTAACGGAAGTTTTCGAAAAGGAAATTGTTCAAGATAGG
41 (46 E-5)	GTTATCTCTAACGGAAGTTAAAGGAAATTGTTCAAGATAGG
51 (46 E+5)	GTTATCTCTAACGGAAGTTTTCGATCTGAAAAGGAAATTGTTCAAGATAGG
56 (46 E+10)	GTTATCTCTAACGGAAGTTTTCGATCTGAAGCGAAAAGGAAATTGTTCAAGATAGG
44 (49 E-5)	GTTATCTCTACTAACGGAAGTTAAAGGAAATTGTTCAAGATAGG
54 (49 E+5)	GTTATCTCTACTAACGGAAGTTTTCGATCTGAAAAGGAAATTGTTCAAGATAGG
59 (49 E+10)	GTTATCTCTACTAACGGAAGTTTTCGATCTGAAGCGAAAAGGAAATTGTTCAAGATAGG
46 opt GATA	CTTATCTCTAACGGAAGTTTTCGAAAAGGAAATTGTTCGAGATAAG
46 opt Ets	GTTATCTCTACCGGAAGTTTTCGAACCGGAAGTTGTTCAAGATAGG
46 opt Ets and GATA	CTTATCTCTACCGGAAGTTTTCGAACCGGAAGTTGTTCGAGATAAG
49 opt GATA	CTTATCTCTACTAACGGAAGTTTTCGAAAAGGAAATTGTTCGAGATAAG
49 opt Ets	GTTATCTCTACTACCGGAAGTTTTCGAACCGGAAGTTGTTCAAGATAGG
49 opt Ets and GATA	CTTATCTCTACTACCGGAAGTTTTCGAACCGGAAGTTGTTCGAGATAAG

REFERENCES AND NOTES

1. D. Acampora, A. Annino, F. Tuorto, E. Puelles, W. Lucchesi, A. Papalia, A. Simeone, Otx genes in the evolution of the vertebrate brain. *Brain Res. Bull.* **66**, 410–420 (2005). Medline doi:10.1016/j.brainresbull.2005.02.005

2. F. Beby, T. Lamonerie, The homeobox gene *Otx2* in development and disease. *Exp. Eye Res.* **111**, 9–16 (2013). Medline doi:10.1016/j.exer.2013.03.007

3. U. Rothbächer, V. Bertrand, C. Lamy, P. Lemaire, A combinatorial code of maternal GATA, Ets and beta-catenin-TCF transcription factors specifies and patterns the early ascidian ectoderm. *Development* **134**, 4023–4032 (2007). Medline doi:10.1242/dev.010850

4. P. Khoueiry, U. Rothbächer, Y. Ohtsuka, F. Daian, E. Frangulian, A. Roure, I. Dubchak, P. Lemaire, A *cis*-regulatory signature in ascidians and flies, independent of transcription factor binding sites. *Curr. Biol.* **20**, 792–802 (2010). Medline doi:10.1016/j.cub.2010.03.063

5. V. Bertrand, C. Hudson, D. Caillol, C. Popovici, P. Lemaire, Neural tissue in ascidian embryos is induced by FGF9/16/20, acting via a combination of maternal GATA and Ets transcription factors. *Cell* **115**, 615–627 (2003). Medline doi:10.1016/S0092-8674(03)00928-0

6. C. Hudson, P. Lemaire, Induction of anterior neural fates in the ascidian *Ciona intestinalis*. *Mech. Dev.* **100**, 189–203 (2001). Medline doi:10.1016/S0925-4773(00)00528-1

7. S. Barolo, J. W. Posakony, Three habits of highly effective signaling pathways: Principles of transcriptional control by developmental cell signaling. *Genes Dev.* **16**, 1167–1181 (2002). Medline doi:10.1101/gad.976502

8. T. Oosterveen, S. Kurdija, M. Ensterö, C. W. Uhde, M. Bergsland, M. Sandberg, R. Sandberg, J. Muhr, J. Ericson, SoxB1-driven transcriptional network underlies neural-specific interpretation of morphogen signals. *Proc. Natl. Acad. Sci. U.S.A.* **110**, 7330–7335 (2013). Medline doi:10.1073/pnas.1220010110

9. K. A. Guss, C. E. Nelson, A. Hudson, M. E. Kraus, S. B. Carroll, Control of a genetic regulatory network by a selector gene. *Science* **292**, 1164–1167 (2001). Medline doi:10.1126/science.1058312

10. G. H. Wei, G. Badis, M. F. Berger, T. Kivioja, K. Palin, M. Enge, M. Bonke, A. Jolma, M. Varjosalo, A. R. Gehrke, J. Yan, S. Talukder, M. Turunen, M. Taipale, H. G. Stunnenberg, E. Ukkonen, T. R. Hughes, M. L. Bulyk, J. Taipale, Genome-wide analysis of ETS-family DNA-binding in vitro and in vivo. *EMBO J.* **29**, 2147–2160 (2010). Medline doi:10.1038/emboj.2010.106

11. G. Badis, M. F. Berger, A. A. Philippakis, S. Talukder, A. R. Gehrke, S. A. Jaeger, E. T. Chan, G. Metzler, A. Vedenko, X. Chen, H. Kuznetsov, C. F. Wang, D. Coburn, D. E. Newburger, Q. Morris, T. R. Hughes, M. L. Bulyk, Diversity and complexity in DNA recognition by transcription factors. *Science* **324**, 1720–1723 (2009). Medline doi:10.1126/science.1162327

12. A. Jolma, J. Yan, T. Whitington, J. Toivonen, K. R. Nitta, P. Rastas, E. Morgunova, M. Enge, M. Taipale, G. Wei, K. Palin, J. M. Vaquerizas, R. Vincentelli, N. M. Luscombe,

T. R. Hughes, P. Lemaire, E. Ukkonen, T. Kivioja, J. Taipale, DNA-binding specificities of human transcription factors. *Cell* **152**, 327–339 (2013). Medline doi:10.1016/j.cell.2012.12.009

13. K. R. Nitta, A. Jolma, Y. Yin, E. Morgunova, T. Kivioja, J. Akhtar, K. Hens, J. Toivonen, B. Deplancke, E. E. Furlong, J. Taipale, Conservation of transcription factor binding specificities across 600 million years of bilateria evolution. *eLife* **4**, (2015). Medline doi:10.7554/eLife.04837

14. M. A. Hume, L. A. Barrera, S. S. Gisselbrecht, M. L. Bulyk, UniPROBE, update 2015: New tools and content for the online database of protein-binding microarray data on protein-DNA interactions. *Nucleic Acids Res.* **43** (D1), D117–D122 (2015). Medline doi:10.1093/nar/gku1045

15. H. Zhang, M. Levine, H. L. Ashe, Brinker is a sequence-specific transcriptional repressor in the *Drosophila* embryo. *Genes Dev.* **15**, 261–266 (2001). Medline doi:10.1101/gad.861201

16. J. Crocker, N. Abe, L. Rinaldi, A. P. McGregor, N. Frankel, S. Wang, A. Alsawadi, P. Valenti, S. Plaza, F. Payre, R. S. Mann, D. L. Stern, Low affinity binding site clusters confer hox specificity and regulatory robustness. *Cell* **160**, 191–203 (2015). Medline doi:10.1016/j.cell.2014.11.041

17. S. Small, R. Kraut, T. Hoey, R. Warrior, M. Levine, Transcriptional regulation of a pair-rule stripe in *Drosophila*. *Genes Dev.* **5**, 827–839 (1991). Medline doi:10.1101/gad.5.5.827

18. C. I. Swanson, D. B. Schwimmer, S. Barolo, Rapid evolutionary rewiring of a structurally constrained eye enhancer. *Curr. Biol.* **21**, 1186–1196 (2011). Medline doi:10.1016/j.cub.2011.05.056

19. J. Jiang, M. Levine, Binding affinities and cooperative interactions with bHLH activators delimit threshold responses to the dorsal gradient morphogen. *Cell* **72**, 741–752 (1993). Medline doi:10.1016/0092-8674(93)90402-C

20. R. W. Lusk, M. B. Eisen, Evolutionary mirages: Selection on binding site composition creates the illusion of conserved grammars in *Drosophila* enhancers. *PLOS Genet.* **6**, e1000829 (2010). Medline doi:10.1371/journal.pgen.1000829

21. M. M. Kulkarni, D. N. Arnosti, cis-regulatory logic of short-range transcriptional repression in *Drosophila melanogaster*. *Mol. Cell. Biol.* **25**, 3411–3420 (2005). Medline doi:10.1128/MCB.25.9.3411-3420.2005

22. C. I. Swanson, N. C. Evans, S. Barolo, Structural rules and complex regulatory circuitry constrain expression of a Notch- and EGFR-regulated eye enhancer. *Dev. Cell* **18**, 359–370 (2010). Medline doi:10.1016/j.devcel.2009.12.026

23. A. Muhlethaler-Mottet, M. Krawczyk, K. Masternak, C. Spilianakis, A. Kretsovali, J. Papamatheakis, W. Reith, The S box of major histocompatibility complex class II promoters is a key determinant for recruitment of the transcriptional co-activator CIITA. *J. Biol. Chem.* **279**, 40529–40535 (2004). Medline doi:10.1074/jbc.M406585200

24. D. Panne, T. Maniatis, S. C. Harrison, An atomic model of the interferon-beta enhanceosome. *Cell* **129**, 1111–1123 (2007). Medline doi:10.1016/j.cell.2007.05.019

25. W. Shi, M. Levine, Ephrin signaling establishes asymmetric cell fates in an endomesoderm lineage of the *Ciona* embryo. *Development* **135**, 931–940 (2008). Medline doi:10.1242/dev.011940

26. H. Yasuo, C. Hudson, FGF8/17/18 functions together with FGF9/16/20 during formation of the notochord in *Ciona* embryos. *Dev. Biol.* **302**, 92–103 (2007). Medline doi:10.1016/j.ydbio.2006.08.075

27. A. Stolfi, E. Wagner, J. M. Taliaferro, S. Chou, M. Levine, Neural tube patterning by Ephrin, FGF and Notch signaling relays. *Development* **138**, 5429–5439 (2011). Medline doi:10.1242/dev.072108

28. E. Wagner, M. Levine, FGF signaling establishes the anterior border of the *Ciona* neural tube. *Development* **139**, 2351–2359 (2012). Medline doi:10.1242/dev.078485

29. W. A. Whyte, D. A. Orlando, D. Hnisz, B. J. Abraham, C. Y. Lin, M. H. Kagey, P. B. Rahl, T. I. Lee, R. A. Young, Master transcription factors and mediator establish super-enhancers at key cell identity genes. *Cell* **153**, 307–319 (2013). Medline doi:10.1016/j.cell.2013.03.035

30. D. Hnisz, B. J. Abraham, T. I. Lee, A. Lau, V. Saint-André, A. A. Sigova, H. A. Hoke, R. A. Young, Super-enhancers in the control of cell identity and disease. *Cell* **155**, 934–947 (2013). Medline doi:10.1016/j.cell.2013.09.053

31. Q. Li, K. R. Peterson, X. Fang, G. Stamatoyannopoulos, Locus control regions. *Blood* **100**, 3077–3086 (2002). Medline doi:10.1182/blood-2002-04-1104

32. G. M. Clore, A. Bax, J. G. Omichinski, A. M. Gronenborn, Localization of bound water in the solution structure of a complex of the erythroid transcription factor GATA-1 with DNA. *Structure* **2**, 89–94 (1994). Medline doi:10.1016/S0969-2126(00)00011-3

33. R. Kodandapani, F. Pio, C. Z. Ni, G. Piccialli, M. Klemsz, S. McKercher, R. A. Maki, K. R. Ely, A new pattern for helix-turn-helix recognition revealed by the PU.1 ETS-domain-DNA complex. *Nature* **380**, 456–460 (1996). Medline doi:10.1038/380456a0

34. L. Christiaen, E. Wagner, W. Shi, M. Levine, Electroporation of transgenic DNAs in the sea squirt *Ciona*. *Cold Spring Harb. Protoc.* **2009**, pdb.prot5345 (2009). Medline doi:10.1101/pdb.prot5345

35. S. Heinz, C. Benner, N. Spann, E. Bertolino, Y. C. Lin, P. Laslo, J. X. Cheng, C. Murre, H. Singh, C. K. Glass, Simple combinations of lineage-determining transcription factors prime *cis*-regulatory elements required for macrophage and B cell identities. *Mol. Cell* **38**, 576–589 (2010). Medline doi:10.1016/j.molcel.2010.05.004

36. A. H. Batchelor, D. E. Piper, F. C. de la Brousse, S. L. McKnight, C. Wolberger, The structure of GABPalpha/beta: An ETS domain- ankyrin repeat heterodimer bound to DNA. *Science* **279**, 1037–1041 (1998). Medline doi:10.1126/science.279.5353.1037

37. M. A. Pufall, G. M. Lee, M. L. Nelson, H. S. Kang, A. Velyvis, L. E. Kay, L. P. McIntosh, B. J. Graves, Variable control of Ets-1 DNA binding by multiple phosphates in an unstructured region. *Science* **309**, 142–145 (2005). Medline doi:10.1126/science.1111915

38. Y. Wang, L. Feng, M. Said, S. Balderman, Z. Fayazi, Y. Liu, D. Ghosh, A. M. Gulick, Analysis of the 2.0 A crystal structure of the protein-DNA complex of the human PDEF

Ets domain bound to the prostate specific antigen regulatory site. *Biochemistry* **44**, 7095–7106 (2005). Medline doi:10.1021/bi047352t

39. E. P. Lamber, L. Vanhille, L. C. Textor, G. S. Kachalova, M. H. Sieweke, M. Wilmanns, Regulation of the transcription factor Ets-1 by DNA-mediated homo-dimerization. *EMBO J.* **27**, 2006–2017 (2008). Medline doi:10.1038/emboj.2008.117

40. D. L. Bates, Y. Chen, G. Kim, L. Guo, L. Chen, Crystal structures of multiple GATA zinc fingers bound to DNA reveal new insights into DNA recognition and self-association by GATA. *J. Mol. Biol.* **381**, 1292–1306 (2008). Medline doi:10.1016/j.jmb.2008.06.072

41. M. van den Boogaard, L. Y. Wong, F. Tessadori, M. L. Bakker, L. K. Dreizehnter, V. Wakker, C. R. Bezzina, P. A. 't Hoen, J. Bakkers, P. Barnett, V. M. Christoffels, Genetic variation in T-box binding element functionally affects SCN5A/SCN10A enhancer. *J. Clin. Invest.* **122**, 2519–2530 (2012). Medline doi:10.1172/JCI62613