UNIVERSITY OF CALIFORNIA SAN DIEGO

Investigating the Grammar of the Otx-a Enhancer within *Ciona intestinalis*

A Thesis submitted in partial satisfaction of the requirements for the degree
Master of Science

in

Biology

by

Jessica Lauren Grudzien

Committee in charge:

Professor Emma Farley, Chair
Professor Xin Sun
Professor Deborah Yelon

2018

The Thesis of Jessica Lauren Grudzien is approved, and it is acceptable in quality and form for publication on microfilm and electronically:

Chair

University of California San Diego

2018

DEDICATION

I would like to dedicate this to my loving family:

To my mother, Chris, and father, Daniel, for being my foundation and supporting me every single day throughout my endeavors and being there to catch me if I fall.

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

ACKNOWLEDGEMENTS

I would like to thank Dr. Emma Farley for helping and supporting me throughout my career as a researcher. Her knowledge and guidance have proved invaluable.

I would also like to thank Dr. Xin Sun and Dr. Deborah Yelon for being members of my committee.

I would like to thank Dr. Granton Jindal for helping me with my Principal Component Analysis, and for reviewing my thesis.

I would like to thank Zane Rau in his help in coding the Principal Component Analysis.

I would finally like to thank all the members of the Farley Lab who have made this project successful throughout many trials and errors, especially Katrina Olson, Fabian Lim, and Benjamin Song.

ABSTRACT OF THE THESIS

Investigating the Grammar of the Otx-a Enhancer within *Ciona intestinalis*

by

Jessica Lauren Grudzien

Master of Science in Biology

University of California San Diego 2018

Professor Emma Farley, Chair

Enhancers are elements within our genome that control where and when genes are expressed throughout development. However, how the sequence of the enhancer regulates tissue-specific expression is not fully understood. We can investigate sequence by looking at the transcription factor binding motifs within enhancers. We want to better understand how combinations of motif syntax: the order, orientation, and spacing of motifs, interplays with motif affinity to regulate gene expression. We term the connections between syntax and affinity enhancer grammar. We use the neural plate Otx-a enhancer within *Ciona intestinalis* as a model to investigate if enhancer grammar is present, and to look for motif syntax and affinity trends giving different expression patterns. This enhancer activates when bound by GATA and ETS transcription factors. Our data shows that there is a grammar present within the Otx-a enhancer, as different grammatical variants give varying expression patterns. Our data suggests that changing the motif order to having ETS binding sites on both ends of the enhancer abolishes *Otx* gene expression in most developing embryos. Our data also shows a loss of expression caused by ETS motifs directly next to each other in combination with non-optimal spacing between high affinity GATA and ETS. We also found a grammatical variant with notochord expression, which may be due to two ETS sites close to a FoxA binding site. These findings help us better understand the grammar of the Otx-a enhancer and help us understand how enhancer sequence codes tissue-specific gene expression in development.

Introduction

Successful development requires that the right genes are expressed at the right time and place. Enhancers are the elements within the genome that control when and where genes are expressed, and thus are required for an egg to successfully develop into a multicellular adult organism. Enhancers contain DNA binding motifs to which transcription factors bind. The factors in combination with each other control activation of gene expression in time and space (Long et al., 2016; Shlyueva et al., 2014).

It is possible to identify enhancers in the genome using various epigenetic techniques, however these approaches don’t help us understand how changes in enhancer sequence impact function. Understanding how enhancers encode tissue-specific expression is crucial as changes in enhancer sequence can cause disease and evolutionary adaptation. An example of this is polydactyly, a disease in which an individual has more than the normal number of fingers or toes. Polydactyly has been linked to mutations in a *Sonic Hedgehog* enhancer. These mutations limit the enhancer’s spatial control on gene expression, causing aberrant anterior-posterior patterning during development. These uncontrolled locations of gene expression result in the additional digit growth (Lettice et al., 2017). Though we recognize enhancers and their role in tissue-specific gene expression, we do not fully understand how enhancers encode tissue-specific gene expression.

The Farley lab aims to understand how enhancers encode tissue-specific gene expression. Previous work in the lab suggests that there is a type of language for functional enhancers which we define as enhancer grammar. Grammar, in this context, is referring to the interplay of motif affinity and syntax within enhancers (Barolo, 2016; Farley et al., 2016). Syntax is defined as the order, orientation, and spacing of binding sites. Order is the lineup of motifs within the enhancer region, such as a site upstream of another site. Orientation is the directionality of the binding site, as transcription factors will bind in either a forward or reverse orientation, depending on the sequence. Lastly, spacing is the distance in base pairs between the cores of the motifs that comprise the enhancer. Individual changes in these variables affect gene expression, such as losing tissue-specificity and changing expression levels (Farley et al., 2016, 2015). We look at these syntax variables in combination to better understand how enhancers control tissue-specific gene expression.

The 69-bp-long Otx-a enhancer was used to study enhancer grammar in this thesis due to its role in neuronal development and our extensive understanding of how its linear sequence controls gene expression (Farley et al., 2015). This enhancer regulates the *Orthodenticle homeobox gene* (*Otx*) within *Ciona intestinalis* (*Ciona*). Homeobox genes are regulatory genes which use tissue-specific expression to control formations of body structures, the *Otx* gene is important for neural plate specification in vertebrates and neural specification in *Ciona* (Acampora et al., 2005). The Otx-a enhancer contains five transcription factor binding site motifs: three GATA and two ETS sites. The GATA sites restrict gene expression to the ectoderm, while the ETS sites control activation of the enhancer via fibroblast growth factor (FGF) signaling, required for neural induction during development (Guss, 2001; Rothbacher et al., 2007). GATA and ETS are thus required transcription factors for normal development within the Otx-a enhancer, important for neural plate development within *Ciona* (Bertrand et al., 2003).

Ciona, commonly known as sea squirts, is an ideal model organism in order to test many different enhancers in all cells of a developing embryo. *Ciona* is a urochordate, which is the closest ancestor to vertebrates (Delsuc et al., 2006). The developmental programs and tissues of the *Ciona* embryo have many similarities with vertebrate embryos. We can follow every single cell through its developmental stages as *Ciona* has defined cell lineages (Passamaneck and Di Gregorio, 2005). This makes tracing gene expression throughout cell stages easier to quantify. The key advantage of Ciona is the power of electroporation in this organism, as we can do high-throughput studies of enhancer variants using thousands of *Ciona* embryos in a short period of time (Christiaen et al., 2014, Farley et al., 2015).

To better understand how enhancers control tissue-specific gene expression, we have previously manipulated different aspects of syntax and affinity within the Otx-a enhancer. Farley et al. had previously explored the spacing and affinity of motifs within the Otx-a enhancer (Farley et al., 2015). The GATA and ETS binding sites within this enhancer do not have the spacing optimal for highest levels of transcription. When Farley et al. introduced a more optimal spacing in combination with high affinity GATA and ETS motifs within the enhancer; the *Ciona* embryos lost tissue-specificity and had robust ectopic expression within different tissues. Normal, tissue-restricted expression for the Otx-a enhancer is located in the anterior sensory vesicle and palps (a 6.5 lineage) and the dorsal nerve cord, epidermis, and two tail tip muscle cells (b 6.5 lineage) of the developing embryo (Fig. 1). In their studies, ectopic expression was found in places such as the notochord and anterior endoderm, which are tissues that have FGF signaling as well (Shi and Levine, 2008; Yasuo and Hudson, 2007). Thus, Farley et. al discovered the importance for sub-optimized spacing between high affinity motifs in Otx-a for tissue-specific expression in development (Farley et al., 2015).

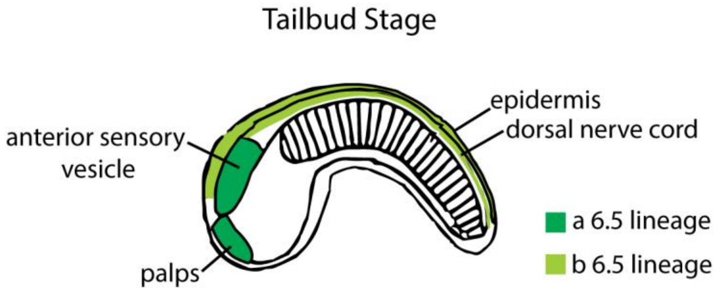


Figure 1: Tailbud-stage *C. intestinalis* embryo with wild-type Otx-a expression. Not pictured is two tail muscle b 6.5 cells with GFP expression. Adapted from "Suboptimization of developmental enhancers", by Farley et al., 2015.

Another study by Farley et. al explored enhancer grammar within their notochord expressing variant, RS 6. They discovered that this notochord variant had a ZicL binding site, which is known to regulate notochord-specific activity by acting on the *Brachyury* gene (Imai et al., 2002; Yagi, 2004). They found that reversing the orientation of an ETS site closest to ZicL abolished normal notochord expression, which emphasized an additional aspect of motif syntax important for gene expression. They also found that when changing the spacing from optimal to suboptimal between a low affinity ETS site closest to ZicL, expression levels decreased. However, by making this low affinity ETS have high affinity, they were able to rescue strong expression levels. This study illustrated the interplay between both syntax and affinity, and this interplay is what we are defining as grammar. (Farley et al., 2016).

There is scientific need to further explore how enhancer grammar defines tissue-specific expression throughout development. Farley et al. discovered how spacing and affinity between motifs changed gene expression within the Otx-a enhancer. We can deepen our understanding of the grammar of the Otx-a enhancer by looking at motif orientation and order in combination with spacing. A library of about 500,000 Otx-a enhancer grammatical variants, called Otx-a Library Scrambled (OLS), has different combinations of motif syntax changed from wild-type. I have tested 30 of these grammatical variants' gene expression profiles to validate future high-throughput experiments on the set of 500,000 variants. Five of these variants, numbered OLS 31-35, only have spacing changed from wild-type, while OLS 36-60 have differences in order, orientation, and spacing. These selected enhancer grammatical variants are fused to a bpFog minimal promoter and green fluorescent protein (GFP) and electroporated into *Ciona* embryos to map tissue expression. We can use the results of these experiments to find out if there is an Otx-a enhancer grammar, and better understand how this grammar encodes tissue-specific expression.

Materials and Methods

Electroporation

Adult *Ciona intestinalis* were obtained from M-Rep (San Diego, CA) and maintained in real seawater from the Pacific Ocean at 18°C with constant illumination. Dechorination, in vitro fertilization and electroporation were carried out as detailed in reference (Christiaen et al., 2014). For each electroporation, eggs and sperm were collected from 4-8 adults, and 70 µg DNA was resuspended in 500 µL buffer. Embryos were fixed for 15 minutes in 3.7% formaldehyde at the early tailbud stage, around 8 hours 10 minutes after fertilization. The embryos were then put through washes of PBTT overnight. The following day the embryos were washed in 0.01% Triton-X in PBS and mounted with Prolong Gold Mounting Media. Differential interference-contrast microscopy was used to obtain transmitted light micrographs with an Olympus Fluoview 3000 Confocal Laser Scanning Microscope using the x40 objective, the same microscope was used to obtain GFP images. All constructs were electroporated at least twice in two separate experiments (biological replicates).

Counting Embryos

The first biological repeat was not randomized or blinded as nothing was known about the expression patterns of each construct. For the second biological repeat, after embryos were mounted on slides, slide labels were covered with thick tape and randomly numbered and randomized. All comparative constructs were tested in the same experiment with wild-type Otx-a and water as a control. Illumination intensity was constant for all experiments. A minimum of 50 embryos were counted for each slide, depending on the amount of well-developed embryos present. The only exceptions to this was wild-type Otx-a Batch 3, which had a sample size of 31 embryos, and one biological repeat of OLS 33, which had a sample size of 47 embryos. Expression was counted if it was present in any visible amount in the specific tissue type. Statistical error of counting data is shown as standard error of the mean.

Acquisition of Images

All comparative images were taken with identical settings. WT Otx-a, OLS 47, 48, OLS 47 No ETS-2, and OLS 48 No ETS-2 were imaged on the same day as each other. OLS 57 and OLS 57 No ETS-2 were imaged on the same day as each other. Images were rotated and cropped to have the same orientation as each other. All embryo images had brightness adjusted up 100% in Photoshop (CC 2018, Adobe Photoshop) to better see expression in tissues.

Enhancer Grammatical Variants Creation

Synthetic enhancers OLS 31-35 were chosen by randomly selecting variants from a computer-generated library of Otx-a enhancer sequences that had the spacing between all GATA and ETS sites randomized, but order and orientation of the sites held constant. Enhancers OLS 36-40 were chosen randomly from a computer-generated library of Otx-a enhancer sequences that had spacing, order, and orientation of the GATA and ETS sites changed from wild type, called Otx-a Library Scrambled (OLS). For OLS 41-60, 10 more variants were randomly chosen from the OLS library. These 10 were then manipulated in order to change one or two aspects of their syntax, and the pair was created to fill the remaining 10 variants (OLS 41 and 42 are paired, and so on). These sequences were ordered from Integrated DNA Technologies with PCR amplification arms containing BseRI sites (5’CATCATGACGAGGAGAAACCAGCAC, 3’ AAACCATTCTCCTCTTCCATCAT). The PCR product was cloned into custom designed SEL-

Seq (Synthetic Enhancer Library Sequencing) vector using type II restriction enzyme BseRI. After cloning, the library was transformed into bacteria (Top10 chemically competent cells), the culture was grown for 8 hours in 4 mL 2xYT growth media. DNA was extracted via standard mini-prep plasmid extraction, and enhancer region sequenced to ensure plasmid accuracy. The culture was then grown overnight for 15 hours, and DNA extracted from the bacterial liquid culture via Nucleobond Xtra Midi kit from Macherey-Nagel (Davos, Switzerland).

Enhancer Grammatical Variants with No ETS Site Creation

The core of the ETS site sequence was changed from TTCC to TCTA to make the site non-functional, as done in previous work (Beh et al., 2007). Mutagenesis primers were created via Integrated DNA Technologies, and PCR was performed with the template DNA and PfuTurbo enzyme. Dpn1 enzyme was added to remove template DNA afterwards, and after incubation the samples were transformed overnight. Culture was grown up for 8 hr in 4mL 2xYT and DNA was extracted via standard mini-prep plasmid extraction. The plasmid DNA was checked for sequence accuracy to ensure the ETS site was made non-functional. Culture was then grown overnight for 15 hours, and DNA extracted from the bacteria via Macherey-Nagel Nucleobond Xtra Midi kit from Macherey-Nagel (Davos, Switzerland).

Software Analysis

To find additional patterns of enhancer grammar within the dataset of variants, a Principal Component Analysis was performed on the experimental data using the MathWorks Statistics and Machine Learning Toolbox pca function in MATLAB and the biplot function to visualize the results (R2018a, The MathWorks).

Results

Wild-type Otx-a Expression in Different Batches of Prepared DNA

The wild-type Otx-a enhancer was used as a control in each electroporation to determine if the percentage of well-developed embryos with tissue expression stayed relatively constant across experiments. This control was also used to check that the grammatical variants tested in the specific experiment were not giving a false representation of gene expression, which could result if something went wrong in the electroporation step. Each batch of wild-type Otx-a DNA was midi-prepped in different stages in my experimental timeline.

Percentage of well-developed embryos with a 6.5 and b 6.5 expression increased with batches of midi-prepped DNA from batch 1 to batch 4. Batch 4 had 3% more embryos with anterior sensory vesicle expression, and 11% more embryos with dorsal nerve cord expression than batch 1. Batch 1 also had an unusual amount of embryos with anterior sensory vesicle expression, not seen in any other experiment. Batch 3 had the highest expression of all batches, but had a small sample size (n=31) and only one biological repeat (Fig. 2C). There is an overall increasing trend of embryos with expression seen throughout wild-type Otx-a DNA preps.

Testing the Grammatical Variants OLS 31-60

We organized expression results from two biological repeats (n=50 embryos minimum, exception n=47 in one repeat of OLS 33) for OLS 31-60. The data was organized into categories based on the total percentage of well-developed embryos with expression (Fig. 4). We looked at the schematics and sequence of all OLS variants in conjunction to the expression data (Fig. 3B, Table 1B). We found a potential grammatical constraint using this visual analysis. In the lowest bracket (0-10% embryo expression), OLS 38, 40, and 57 had similar enhancer structure of ETS motifs touching (Fig. 3A).

Principal Component Analysis to Find Syntax Trends on Expression

We performed a Principal Component Analysis (PCA) on the expression data collected from OLS 31-60 and wild-type to find potential syntax trends that could explain enhancer activity. PCA is a method of summarizing data influenced by multiple variables. It reduces the number of variables by combining them into principal components that show as much variation as possible. This way, we can analyze 30+ combinations of enhancer grammatical variants at the same time and visualize if any attributes about these variants are influencing expression in the same way. Two attribute vectors pointing in the same direction likely correlate, and the magnitude (length) of the vector represents the attribute's contribution to variance on the principal components.

We added expression attributes to the PCA to look for syntax that might have similar effects on expression levels. These expression attributes included no and low a 6.5 and b 6.5 tissue expression to look at syntax that prevented functional Otx-a enhancer activity. If any grammatical variants pointed in the direction of these attribute vectors, these grammatical variants had no/low overall expression. We also plotted expression attributes with a 6.5 and b 6.5 tissue expression above 50% presence in well-developed embryos to determine variants with high Otx-a enhancer activity. We also plotted ectopic tissue expression in the PCA: endoderm, mesenchyme, and notochord tissues, to see if any grammatical variants correlated with ectopic expression. We separated the PCA plots for order and orientation to easily look at the syntax variable's effect on the data. We did not look at spacing in the PCA as the number of spacing combinations was too complex at the time. With all these expression attributes set, we then inputted the expression data from OLS 31-60, as well as wild-type, into the PCA.

We saw certain attributes about grammatical variants correlating with expression present in the PCA plot. For example, the order ETS, GATA, GATA, GATA, ETS correlated with low (<15% embryos with expression) a 6.5 and b 6.5 tissue-specific expression. In contrast, the wild-type motif order GATA, ETS, ETS, GATA, GATA correlated with strong (>50%) a 6.5 and b 6.5 tissue-specific expression. The order GATA, ETS, GATA, ETS, GATA had a strong contribution to the PCA's variance shown by the large vector length, but weakly correlated towards endoderm (>5%) expression. Lastly, the order GATA, GATA, ETS, ETS, GATA weakly correlated towards low/no expression based on its distance on the x-axis from the attribute vector. We define this as weak as it is farther separated from the attribute vector on the y-axis, which represents the second principal component (Fig. 5A).

In the motif orientation PCA plot, certain orientations correlated with different expression patterns. The orientation of sites is based on the order GATA-1, ETS-1, ETS-2, GATA-2, and GATA-3, regardless of the order of sites in the variant. Forward, reverse, reverse, forward, forward correlated strongly with no a 6.5 and b 6.5 lineage expression. Reverse, forward, forward, reverse, forward, or wild-type orientation of binding sites, weakly correlated with high (>50%) a 6.5 and b 6.5 tissue expression. Forward, forward, reverse, reverse, forward correlated with mesenchyme (>10%) and endoderm (>5%) expression (Fig. 5B).

Grammatical Variants with No Anterior Sensory Vesicle or Dorsal Nerve Cord Expression

We took a closer look at the variants with ETS sites touching, discovered through the initial look at the expression data. Of OLS 31-60, only OLS 38, 40, and 57 had absolutely no embryos expressing a 6.5 lineage or b 6.5 lineage expression. The three also all had ETS motifs directly next to each other, 8-bp apart core-to-core (Fig. 6E). However, previous data from OLS grammatical variants 1-30 contradicts this finding. Three OLS grammatical variants, OLS 1, 6, and 15, all had touching ETS motifs, but had embryos with a 6.5 and b 6.5 expression (Table 2, Fig. 6F).

Grammatical Variants with Low Anterior Sensory Vesicle or Dorsal Nerve Cord Expression

We looked at one of the trends suggested by the PCA, which suggested the motif order ETS, GATA, GATA, GATA, ETS resulted in very low a 6.5 and/or b 6.5 expression. Out of OLS 31-60, OLS 36, 37, and 59 showed this order. They also all had less than 10% of embryos with expression (Fig. 7E). This order also showed low expression in our previous data OLS grammatical variants 1-30. Out of the set, one grammar OLS 9 had this motif order, and had no expression measured (Table 2, Fig. 7F).

Notochord Expression

We found one grammatical variant, OLS 48, out of OLS 31-60 that had notochord expression in 22% of the well-developed embryos. We compared this to enhancer OLS 47 that has a similar motif organization yet had no notochord expression (Fig. 8C, 8I). OLS 48 had a different order of binding sites than OLS 47, GATA, ETS, ETS, GATA, GATA, which mirrors wild-type order.

Testing the Importance of ETS Site for Notochord Expression

We made the low affinity ETS site within OLS 47 and 48 non-functional to simplify the syntax within the grammatical variants. By doing this, we could determine if the relationship in spacing between high affinity GATA and ETS influenced notochord expression. We also did not know if the ETS site was being occupied in the enhancer, and by making it non-functional we could determine if the motif was being utilized. The result from both non-functional ETS grammatical variants was the complete loss of gene expression (Fig 9M).

Removing an ETS site from the Non-functional Grammatical Variants

We made the low affinity ETS site within OLS 57 non-functional to explore the idea of spatial constraint preventing proper gene expression when both ETS sites were directly next to each other. We found that gene expression was not rescued within the mutated OLS 57, as it had no expression like its original motif organization (Fig. 10G).

Discussion

Wild-type Otx-a Expression

We can attribute the increase of wild-type Otx-a expression across batches of DNA, and the abnormal amount of higher a 6.5 expression in batch 1 of DNA, to two main reasons (Fig. 2C). I was still learning how to properly midi-prep DNA when I electroporated batch 1 early in my experimental timeline. With batch 4 being completed nearly half a year after the first batch was made, my technique in DNA extractions and electroporations got better, giving higher-quality electroporation results from the experiments. The percentage of embryos with expression in this experiment is not at the same level as previous studies (Farley et al., 2015), but is consistent and increasing over the experimental timeline of this experiment (Fig. 2C). This consistency allows the data from grammatical variants tested to be directly compared to my wild-type expression results.

No Expression in Variants with ETS Sites Close to Each Other

Having ETS sites directly next to each other, 8 base pairs apart core-to-core, resulted in no a 6.5 or b 6.5 lineage expression in all embryos within the 3 variants in my tested variant set, OLS 31-60 (Fig. 6). This observation builds upon Farley et al.'s study in where bringing ETS sites closer together reduced gene expression (Farley et al., 2015), but in this experiment instead of losing the level of expression we have no embryos having expression at all. From this initial data, we formed two hypotheses to explain the lack of gene expression. The first hypothesis was that the ETS transcription factor needs a minimal space between the two ETS sites to bind functionally to both sites. It may be that the ETS factor has a flexible range of binding, but that the closeness of the two sites surpassed the ability of the protein to bind correctly (Arnosti and Kulkarni, 2005).

The second possible explanation is that having one ETS binding site is sufficient for expression in some syntaxes and having the two ETS sites so close together is causing ETS to undergo competitive binding with itself, preventing enhancer activity. As two molecules cannot overlap in space, spacing between binding sites may have an important role in allowing both ETS factors to bind. If sites are too close together, one bound ETS factor may be preventing another ETS from binding due to steric hindrance. This additional ETS factor trying to bind may be causing problems in the proper activity of the enhancer. Some transcription factors are known to bind cooperatively, which may alleviate this hindrance and explain the contradicting results proposed by the ETS touching grammatical variants in OLS 1-30 (Deplancke et al., 2016).

To test the second hypothesis, we created a mutant of OLS 57 with a non-functional low affinity ETS site. It was found to have no gene expression in any embryo just like OLS 57 originally had (Fig. 10G). This result does not clear up our hypothesis, as it is very difficult to rescue expression from mutating a site in a non-functional enhancer than the same in a functional enhancer to lose expression.

The first hypothesis, which suggested a spatial constraint between ETS sites preventing the binding of the transcription factor in both sites, did not explain the results from the OLS 1-30 dataset. In the other tested variant set, three variants with this spacing pattern (Fig. 3A, 6F) between ETS had embryos with moderate gene expression (Table 2, Fig. 6F). This suggests that there are other combinatorial factors in play that are preventing Otx-a enhancer activity within the variants tested.

The main cause that may be contributing to this expression difference is that in the ETS touching grammatical variants with expression in set OLS 1-30, high affinity GATA and ETS are close to each other in order (Fig. 6F). This is drastically different from the ETS touching variants in OLS 31-60, which do not have high affinity GATA and ETS near each other. These high affinity sites in vicinity of each other may be the factor in having functional Otx-a activity.

Enhancer Organization that does not Activate Gene Expression

The low percentage of embryos with gene expression in three grammatical variants in my tested variant set and one variant in OLS 1-30 suggested a shared syntax trend of the separation of ETS motifs to opposite sides of the enhancer (Fig. 7). This is similar to previous studies in the Otx-a enhancer, in which Farley et al. changed the spacing between ETS sites from 15 bp to 10, 20, and 25 bp and found a reduction in the level of gene expression in all three cases (Farley et al., 2015). While the previous study looked at levels and percent of embryos with expression, in this study we only analyzed percentage embryos with expression. In this previous study the percentage of embryos with gene expression in these ETS spacer grammatical variants did not change drastically from their wild-type enhancer counterparts (Farley et al., 2015).

ETS are separated much farther than 25bp in the OLS grammatical variants with order ETS, GATA, GATA, GATA, ETS (EGGGE) tested in this experiment. OLS 36, 37, 59, and 9 have ETS sites separated by 45, 52, 55, and 57 bp, respectively (Fig. 7). An initial hypothesis is the possibility that there is a maximum distance in which having ETS sites separated results in a complete loss of Otx-a enhancer-mediated gene expression. However, OLS 39 contradicts this hypothesis, as it has a 47 bp separation of ETS sites (Fig. 3B). This grammar has a moderate percentage of embryos with gene expression like the wild-type enhancer (Fig. 4C). This variant does not follow the order pattern shared by the previous four as it has the order of EGGEG (Fig. 3B), which instead suggests that having all three GATA sites in the middle of the enhancer is possibly preventing this gene expression instead.

To test this hypothesis in future directions, a low affinity GATA site can be knocked out from the middle of one of these enhancers to emulate the ETS grammar structure of OLS 39 (EGGEG) to see if gene expression is rescued (Fig. 11). From these results, we can determine if having three GATA sites next to each other is the cause of abolished gene expression in the Otx-a enhancer, or if a different combination of enhancer syntax such as orientation is having an effect instead.

Another observation contributing to the lack of gene expression may be the lack of high affinity GATA and ETS sites next to each other in all variants tested with order EGGGE. OLS 39, which we compared to for far ETS distances, has expression and has high affinity GATA and ETS motifs next to each other in motif order. A follow up study to this high affinity motif observation would be to make a variant with order EGGGE have high affinity GATA and ETS next to each other in order, without disrupting any other syntax variables. In that context, we would be able to determine if the lack of expression is actually from the three GATA motifs separating the ETS motifs, or if it was from the lack of high affinity GATA and ETS motif vicinity.

Orientation of Sites is Important for Enhancer Function

The PCA of OLS variants 31-60 showed various combinations of motif orientations that suggested trends with enhancer activity. One such trend was the motif orientation forward, reverse, reverse, forward, forward (FRRFF). This orientation strongly correlated with no a and b

6.5 lineage expression in the PCA, based on its overlap with the attribute vectors (Fig. 5B). Looking through the data, the only two variants with this specific motif orientation were OLS 57 and 58. These two variants both had no b 6.5 lineage, and very low/no a 6.5 lineage expression (Fig 4A). We further looked through all tested grammatical variants to find trends in the motif orientations, such as ETS both reversed and GATA-1 and ETS-1 pointed at each other, but we found no pattern that was true across grammatical variants. This suggests that the specific orientation of these motifs may influence the lack of gene expression, but only through additional combinations of syntax.

Our PCA also showed a motif orientation that gave ectopic expression based on the directionality of the vector close to both mesenchyme and endoderm expression. This orientation grammar was FFRRF (Fig. 5B). Interestingly, the only two grammatical variants with this motif orientation was the notochord expression OLS 48 and its pair, OLS 47. Both grammatical variants had a high amount of ectopic expression compared to the rest of the tested grammatical variants. Once again, we looked through the grammatical variants and found no apparent orientation pattern between individual motifs that would explain this ectopic expression. This suggests that the orientation of these motifs may influence ectopic expression, but once again only through additional combinations of syntax.

The last motif orientation trend brought up by the PCA was wild-type orientation (Fig. 5B). This orientation correlated with high expression, but this may be because of its presence in most grammatical variants with wild-type order. The data suggests that this orientation of motifs may influence high a 6.5 and b 6.5 expression within developing embryos in combination with proper syntax.

Grammatical Variants with Notochord Expression

We found notochord expression in OLS 48 (Fig. 8E, 8I). This expression was not present in any other tested grammatical variant in OLS 31-60. In addition, the grammatical variant had an overall high percentage of embryos with gene expression, similar to wild-type but with ectopic expression in the mesenchyme and endoderm (Fig. 8I). When comparing OLS 48 to its close grammatical variant pair, OLS 47, the major difference was the switch in order of binding sites, low affinity ETS and high affinity GATA. We investigated the sequence of OLS 48 to check for the presence of notochord motifs that may be absent in OLS 47, which would account for this notochord expression. We found that there was a motif for notochord specific transcription factor FoxA present in both of the variants 47 and 48 (José-Edwards et al., 2015). A hypothesis is that the two ETS binding sites present in OLS 48 may be allowing a dimer of ETS to interact with the FoxA factor. This interaction would not be possible in OLS 47, which lacks two ETS sites close to the FoxA site. In combination, this ETS dimer and FoxA may be the cause of notochord expression in grammatical variant OLS 48.

Future Directions

To follow up this study, we will continue to explore syntax trends suggested in the Principal Component Analysis. We did not fully explore the combinations of orientations proposed by the PCA that led to a loss of gene expression. We can also find a way to reduce the complexity of spacing and incorporate it into the analysis. We can also make our PCA more informative by adding more data into it. This would include the addition of the tested OLS 1-30 set. We could then make new Otx-a grammatical variants that follow our predictions and see if their gene expression patterns match what we expect.

Having seen that order, orientation and spacing of sites within the Otx-a enhancer is important for activity, the next step will be to conduct high-throughput studies of the synthetic Otx-a enhancer library, via the Synthetic Enhancer Library – Seq (SEL-Seq) method (Farley et al., 2015). We will be electroporating 500,000 grammatical variants into embryos, and then extracting the mRNA and performing an RNA-seq. We could then use our predicted syntax trends to compare and validate these future high-throughput studies.

Another follow up experiment would be to perform a ChIP-exo in order to determine what transcription factor binding sites are occupied within each enhancer variant. ChIP-exo is a new technique that uses chromatin immunoprecipitation to cross-link the DNA and its bound proteins. It ligates in sequencing adaptors, and these are then digested by lambda exonuclease to identify the sequence bound by the factor. High-throughput sequencing is used to generate the sequence of the fragments and allow us to determine where the transcription factors are binding (Rhee and Pugh, 2011). This ChIP-exo experiment will allow us to confirm what binding sites within the Otx-a enhancer are occupied and help us better understand the grammar of the Otx-a enhancer. We would use ChIP-exo instead of the traditional ChIP-chip and ChIP-seq assays as the latter two would not give the resolution required to identify the sites that are bound. (Rhee and Pugh, 2011).

Conclusion

This overall thesis explores the question if the Otx-a enhancer has a grammar. We found that there was a grammar, as different grammatical variants tested gave different expression patterns. We further explored this grammar by looking at potential trends stemming from the interplay of affinity and syntax between motifs within the enhancer. Most of these expression patterns were most influenced by the order of binding motifs within the enhancer. We explored why variants with ETS sites touching in OLS 31-60 gave no expression, while in OLS 1-30 they had expression (Fig. 6, Fig. 10). We found that the close spacing was abolishing gene expression in variants without high affinity GATA and ETS in close vicinity. We also found that having ETS sites separated by three GATA motifs eliminated gene expression in every tested grammatical variant, possibly stemming from the fact that there were no high affinity GATA and ETS in vicinity of each other in the tested variants (Fig. 7). We finally looked at an interesting grammatical variant and proposed that its optimal spacing between ETS sites may be allowing for interactions between ETS and FoxA, causing notochord expression (Fig. 8). Follow up experiments will help flesh out these predicted trends and help in our understanding of functional and non-functional motif syntax within the Otx-a enhancer. Due to the conserved nature of transcription factor binding specificity, many of the trends we identify will also possibly help us understand how enhancers control gene expression in different organisms, helping us better understand the role of proper enhancer grammar in successful development.

Figures

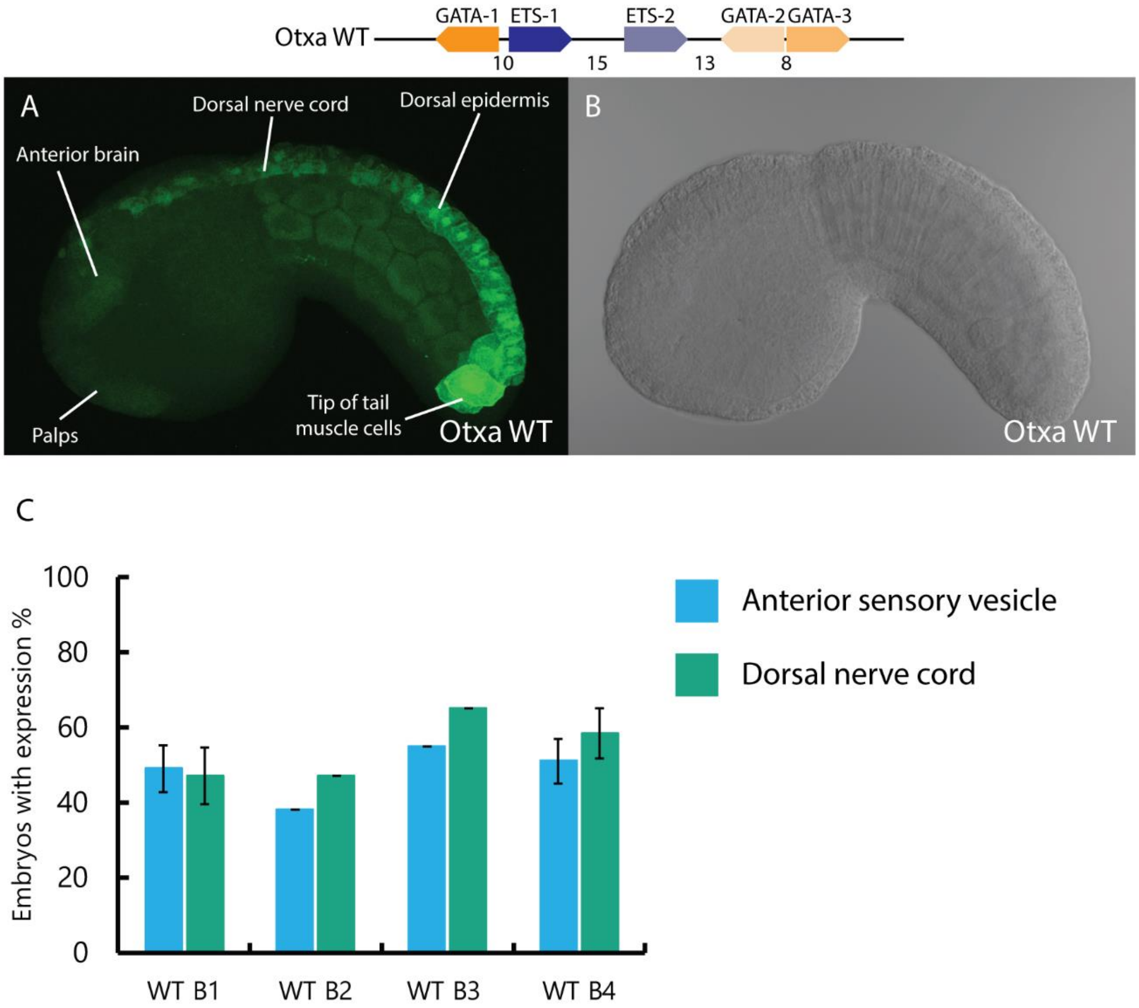


Figure 2: Wild-type Otx-a enhancer expression. Enhancer schematic showing the order, orientation, and spacing (core to core) of ETS and GATA binding sites within the endogenous enhancer. (A) Max intensity projection (MIP) of embryo electroporated with WT Otx-a enhancer, GFP is visible in the anterior brain, palps, dorsal nerve cord, dorsal midline epidermis, and two tail muscle cells. (B) Differential-interference contrast image of the embryo, focused on the neural plane to show notochord and anterior brain structure. (C) Well-developed embryos were counted as having expression in different tissues across different midi-prepped wild-type Otx-a batches. WT batch 1 was midi-prepped first, WT batch 4 was midi-prepped last. Blue columns represent anterior sensory vesicle expression; green columns represent dorsal nerve chord expression. Batch 1 had four biological repeats with a minimum of n=50 in each. Batch 2 and 3 had one biological repeat each, n=72 and n=31 respectively. Batch 4 had 6 biological repeats, with a minimum of n=50 in each repeat.

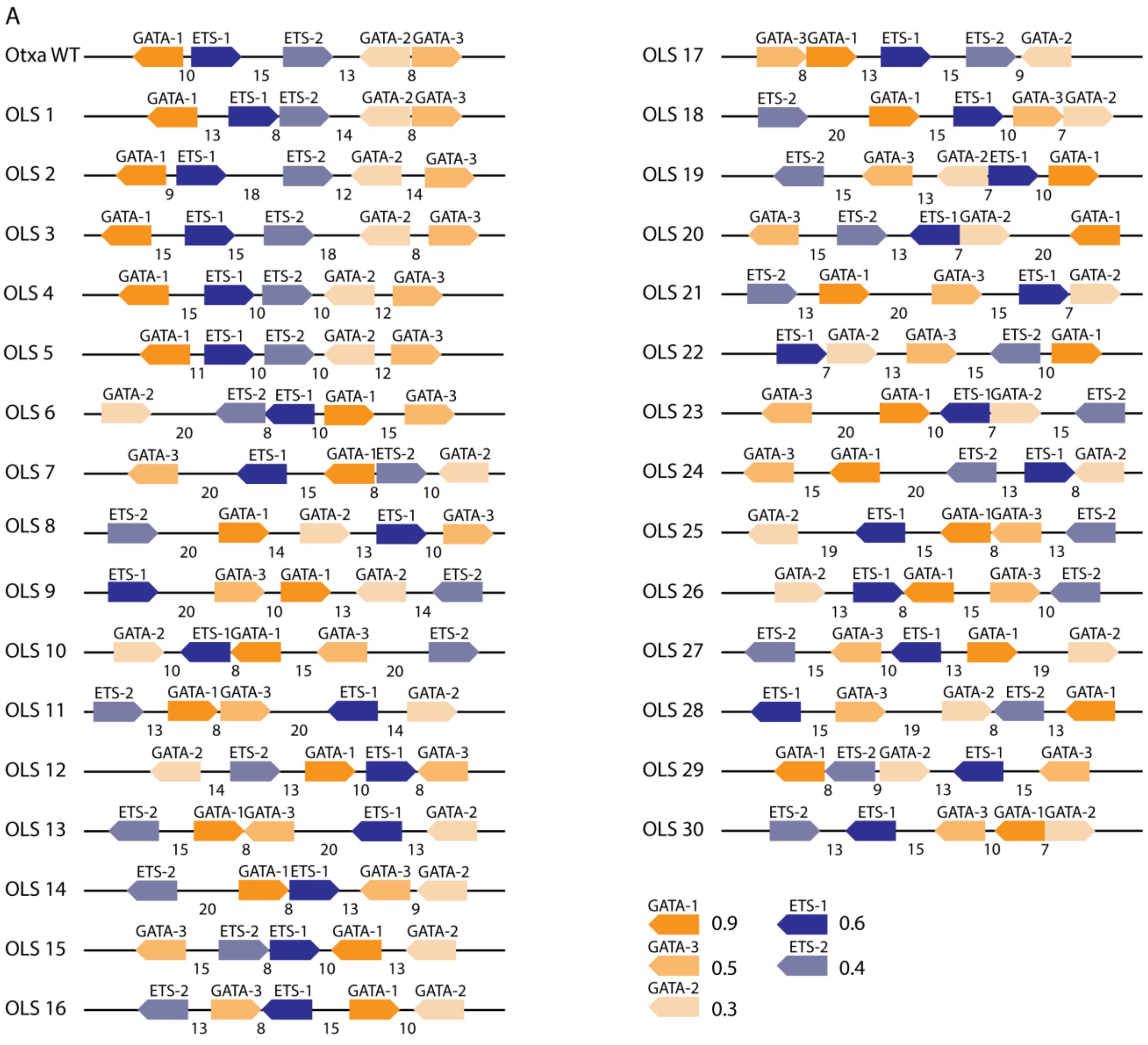

A
Otxa WT GATA-1 ETS-1 ETS-2 GATA-2 GATA-3
10 15 13 8
OLS 1 GATA-1 ETS-1 ETS-2 GATA-2 GATA-3
13 8 14 8
OLS 2 GATA-1 ETS-1 ETS-2 GATA-2 GATA-3
9 18 12 14
OLS 3 GATA-1 ETS-1 ETS-2 GATA-2 GATA-3
15 15 18 8
OLS 4 GATA-1 ETS-1 ETS-2 GATA-2 GATA-3
15 10 10 12
OLS 5 GATA-1 ETS-1 ETS-2 GATA-2 GATA-3
11 10 10 12
OLS 6 GATA-2 ETS-2 ETS-1 GATA-1 GATA-3
20 8 10 15
OLS 7 GATA-3 ETS-1 GATA-1 ETS-2 GATA-2
20 15 8 10
OLS 8 ETS-2 GATA-1 GATA-2 ETS-1 GATA-3
20 14 13 10
OLS 9 ETS-1 GATA-3 GATA-1 GATA-2 ETS-2
20 10 13 14
OLS 10 GATA-2 ETS-1 GATA-1 GATA-3 ETS-2
10 8 15 20
OLS 11 ETS-2 GATA-1 GATA-3 ETS-1 GATA-2
13 8 20 14
OLS 12 GATA-2 ETS-2 GATA-1 ETS-1 GATA-3
14 13 10 8
OLS 13 ETS-2 GATA-1 GATA-3 ETS-1 GATA-2
15 8 20 13
OLS 14 ETS-2 GATA-1 ETS-1 GATA-3 GATA-2
20 8 13 9
OLS 15 GATA-3 ETS-2 ETS-1 GATA-1 GATA-2
15 8 10 13
OLS 16 ETS-2 GATA-3 ETS-1 GATA-1 GATA-2
13 8 15 10
OLS 17 GATA-3 GATA-1 ETS-1 ETS-2 GATA-2
8 13 15 9
OLS 18 ETS-2 GATA-1 ETS-1 GATA-3 GATA-2
20 15 10 7
OLS 19 ETS-2 GATA-3 GATA-2 ETS-1 GATA-1
15 13 7 10
OLS 20 GATA-3 ETS-2 ETS-1 GATA-2 GATA-1
15 13 7 20
OLS 21 ETS-2 GATA-1 GATA-3 ETS-1 GATA-2
13 20 15 7
OLS 22 ETS-1 GATA-2 GATA-3 ETS-2 GATA-1
7 13 15 10
OLS 23 GATA-3 GATA-1 ETS-1 GATA-2 ETS-2
20 10 7 15
OLS 24 GATA-3 GATA-1 ETS-2 ETS-1 GATA-2
15 20 13 8
OLS 25 GATA-2 ETS-1 GATA-1 GATA-3 ETS-2
19 15 8 13
OLS 26 GATA-2 ETS-1 GATA-1 GATA-3 ETS-2
13 8 15 10
OLS 27 ETS-2 GATA-3 ETS-1 GATA-1 GATA-2
15 10 13 19
OLS 28 ETS-1 GATA-3 GATA-2 ETS-2 GATA-1
15 19 8 13
OLS 29 GATA-1 ETS-2 GATA-2 ETS-1 GATA-3
8 9 13 15
OLS 30 ETS-2 ETS-1 GATA-3 GATA-1 GATA-2
13 15 10 7
GATA-1 0.9
GATA-3 0.5
GATA-2 0.3
ETS-1 0.6
ETS-2 0.4

Figure 3: Schematics of Otx-a wild-type, (A) OLS grammatical variants 1-30, (B) OLS grammatical variants 31-60, and No ETS-2 OLS 47, 48, 57. Orange icons are GATA binding sites, blue icons are ETS binding sites. Darker colored icons infer higher affinity binding sites; lighter icons infer lower affinity binding sites. Orientation of the sites is indicated by the direction of the arrow. Spacing is the number between sites and counted from core to core of the motifs. A red X marks the non-functional motif in the No ETS-2 variants.

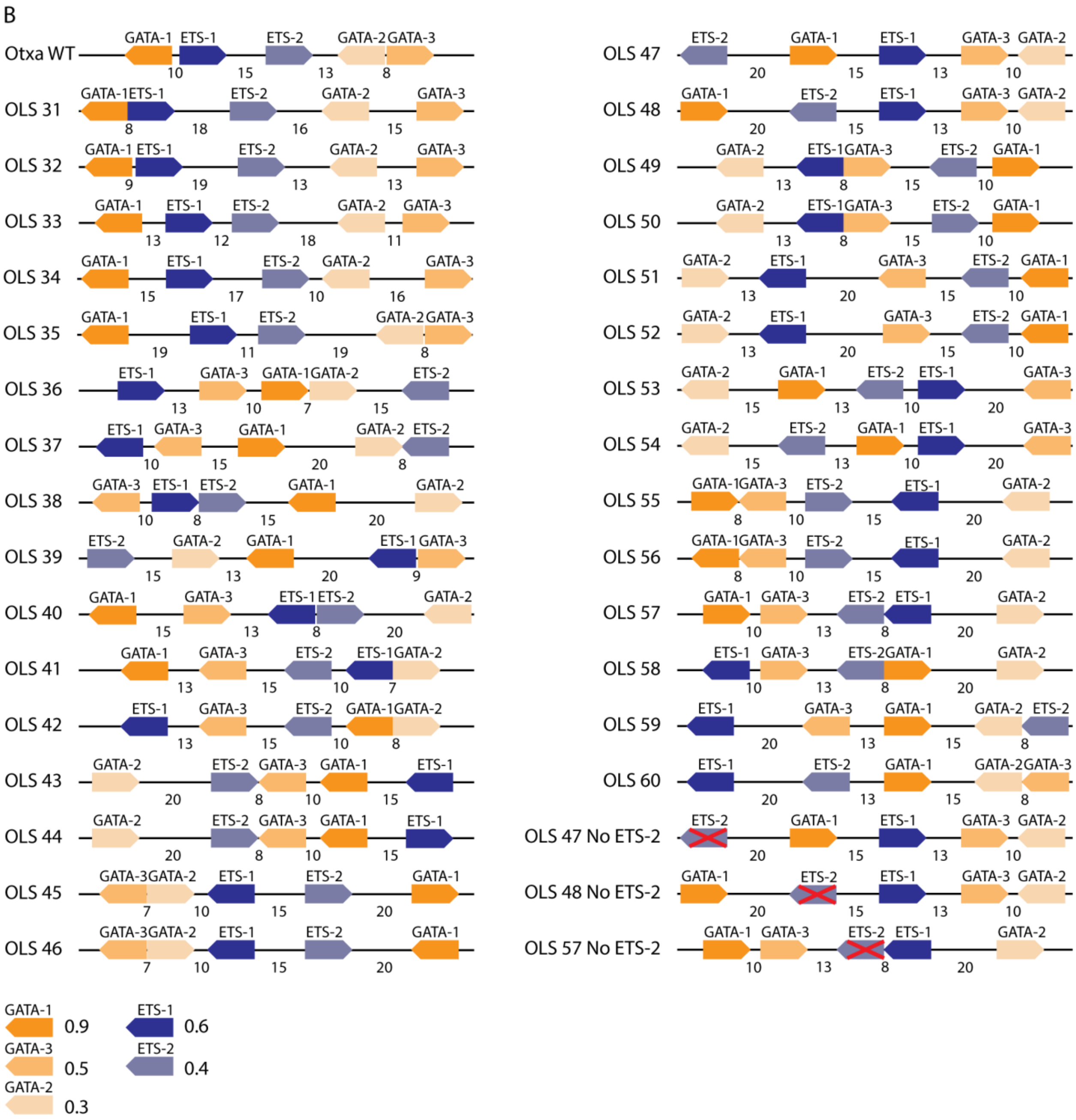
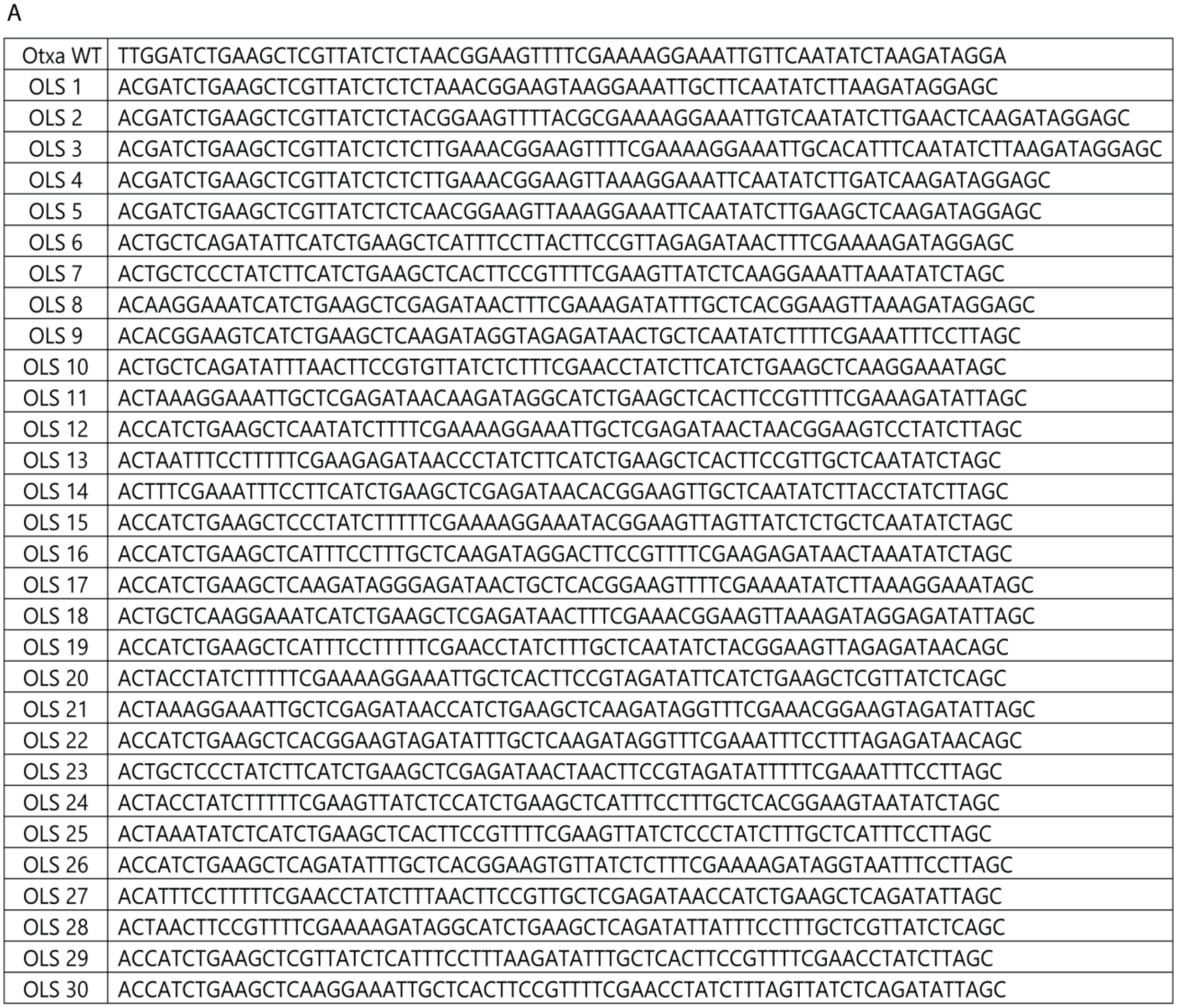

B
Otxa WT GATA-1 ETS-1 ETS-2 GATA-2 GATA-3 10 15 13 8
OLS 31 GATA-1 ETS-1 ETS-2 GATA-2 GATA-3 8 18 16 15
OLS 32 GATA-1 ETS-1 ETS-2 GATA-2 GATA-3 9 19 13 13
OLS 33 GATA-1 ETS-1 ETS-2 GATA-2 GATA-3 13 12 18 11
OLS 34 GATA-1 ETS-1 ETS-2 GATA-2 GATA-3 15 17 10 16
OLS 35 GATA-1 ETS-1 ETS-2 GATA-2 GATA-3 19 11 19 8
OLS 36 ETS-1 GATA-3 GATA-1 GATA-2 ETS-2 13 10 7 15
OLS 37 ETS-1 GATA-3 GATA-1 GATA-2 ETS-2 10 15 20 8
OLS 38 GATA-3 ETS-1 ETS-2 GATA-1 GATA-2 10 8 15 20
OLS 39 ETS-2 GATA-2 GATA-1 ETS-1 GATA-3 15 13 20 9
OLS 40 GATA-1 GATA-3 ETS-1 ETS-2 GATA-2 15 13 8 20
OLS 41 GATA-1 GATA-3 ETS-2 ETS-1 GATA-2 13 15 10 7
OLS 42 ETS-1 GATA-3 ETS-2 GATA-1 GATA-2 13 15 10 8
OLS 43 GATA-2 ETS-2 GATA-3 GATA-1 ETS-1 20 8 10 15
OLS 44 GATA-2 ETS-2 GATA-3 GATA-1 ETS-1 20 8 10 15
OLS 45 GATA-3 GATA-2 ETS-1 ETS-2 GATA-1 7 10 15 20
OLS 46 GATA-3 GATA-2 ETS-1 ETS-2 GATA-1 7 10 15 20
OLS 47 ETS-2 GATA-1 ETS-1 GATA-3 GATA-2 20 15 13 10
OLS 48 GATA-1 ETS-2 ETS-1 GATA-3 GATA-2 20 15 13 10
OLS 49 GATA-2 ETS-1 GATA-3 ETS-2 GATA-1 13 8 15 10
OLS 50 GATA-2 ETS-1 GATA-3 ETS-2 GATA-1 13 8 15 10
OLS 51 GATA-2 ETS-1 GATA-3 ETS-2 GATA-1 13 20 15 10
OLS 52 GATA-2 ETS-1 GATA-3 ETS-2 GATA-1 13 20 15 10
OLS 53 GATA-2 GATA-1 ETS-2 ETS-1 GATA-3 15 13 10 20
OLS 54 GATA-2 ETS-2 GATA-1 ETS-1 GATA-3 15 13 10 20
OLS 55 GATA-1 GATA-3 ETS-2 ETS-1 GATA-2 8 10 15 20
OLS 56 GATA-1 GATA-3 ETS-2 ETS-1 GATA-2 8 10 15 20
OLS 57 GATA-1 GATA-3 ETS-2 ETS-1 GATA-2 10 13 8 20
OLS 58 ETS-1 GATA-3 ETS-2 GATA-1 GATA-2 10 13 8 20
OLS 59 ETS-1 GATA-3 GATA-1 GATA-2 ETS-2 20 13 15 8
OLS 60 ETS-1 ETS-2 GATA-1 GATA-2 GATA-3 20 13 15 8
OLS 47 No ETS-2 ETS-2 GATA-1 ETS-1 GATA-3 GATA-2 20 15 13 10
OLS 48 No ETS-2 GATA-1 ETS-2 ETS-1 GATA-3 GATA-2 20 15 13 10
OLS 57 No ETS-2 GATA-1 GATA-3 ETS-2 ETS-1 GATA-2 10 13 8 20
GATA-1 0.9
ETS-1 0.6
GATA-3 0.5
ETS-2 0.4
GATA-2 0.3

Figure 3: Schematics of Otx-a wild-type, (A) OLS grammatical variants 1-30, (B) OLS grammatical variants 31-60, and No ETS-2 OLS 47, 48, 57. Orange icons are GATA binding sites, blue icons are ETS binding sites. Darker colored icons infer higher affinity binding sites; lighter icons infer lower affinity binding sites. Orientation of the sites is indicated by the direction of the arrow. Spacing is the number between sites and counted from core to core of the motifs. A red X marks the non-functional motif in the No ETS-2 variants, continued.

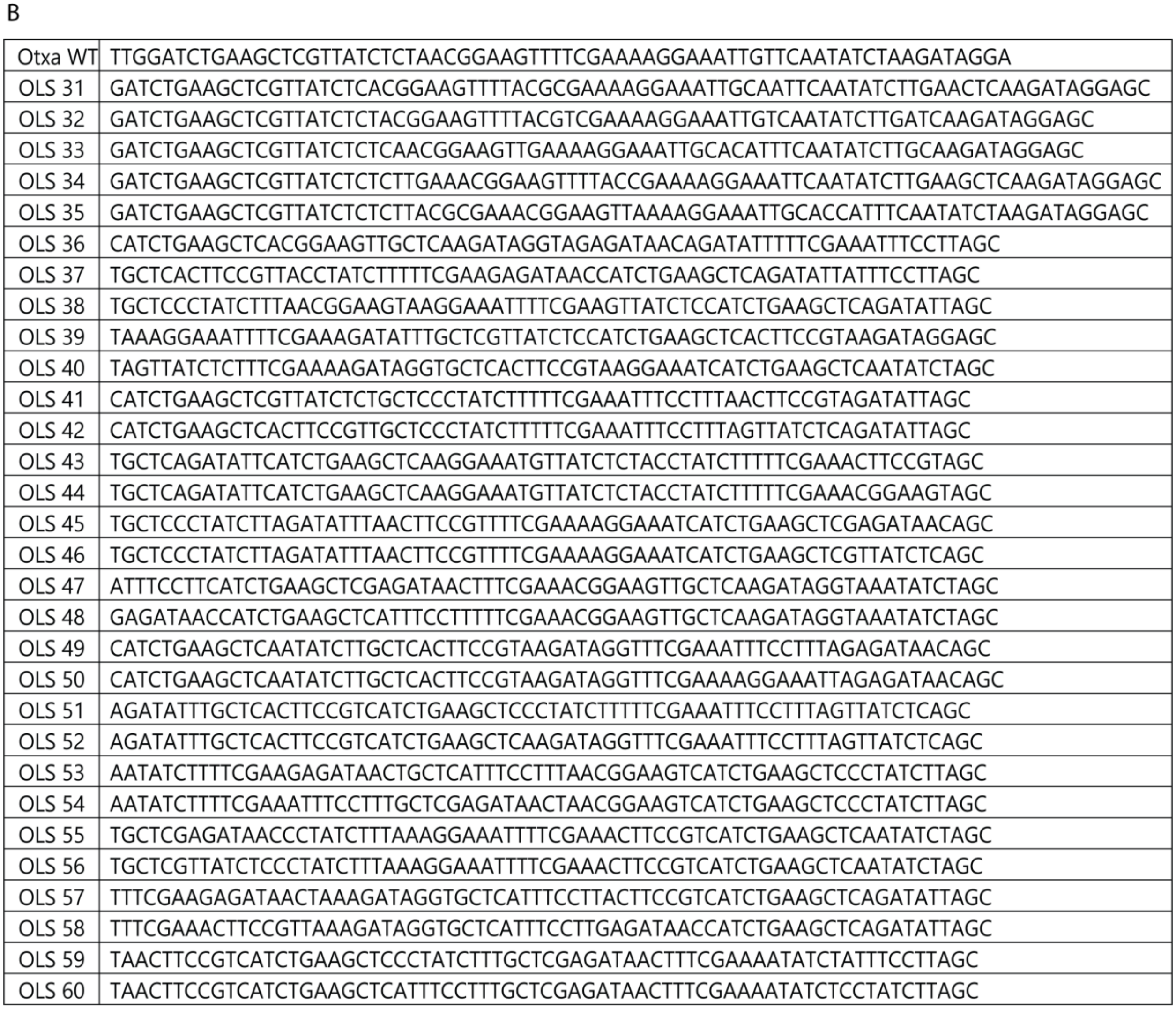
Table 1: Sequence of all enhancer constructs. (A) OLS 1-30, (B) OLS 31-60.



A

Otxa WT	TTGGATCTGAAGCTCGTTATCTCTAACGGAAGTTTTCGAAAAGGAAATTGTTCAATATCTAAGATAGGA
OLS 1	ACGATCTGAAGCTCGTTATCTCTCTAAACGGAAGTAAGGAAATTGCTTCAATATCTTAAGATAGGAGC
OLS 2	ACGATCTGAAGCTCGTTATCTCTACGGAAGTTTTACGCGAAAAGGAAATTGTCAATATCTTGAACTCAAGATAGGAGC
OLS 3	ACGATCTGAAGCTCGTTATCTCTCTTGAAACGGAAGTTTTCGAAAAGGAAATTGCACATTTCAATATCTTAAGATAGGAGC
OLS 4	ACGATCTGAAGCTCGTTATCTCTCTTGAAACGGAAGTTAAAGGAAATTCAATATCTTGATCAAGATAGGAGC
OLS 5	ACGATCTGAAGCTCGTTATCTCTCAACGGAAGTTAAAGGAAATTCAATATCTTGAAGCTCAAGATAGGAGC
OLS 6	ACTGCTCAGATATTCATCTGAAGCTCATTTCCTTACTTCCGTTAGAGATAACTTTCGAAAAGATAGGAGC
OLS 7	ACTGCTCCCTATCTTCATCTGAAGCTCACTTCCGTTTTCGAAGTTATCTCAAGGAAATTAAATATCTAGC
OLS 8	ACAAGGAAATCATCTGAAGCTCGAGATAACTTTCGAAAGATATTTGCTCACGGAAGTTAAAGATAGGAGC
OLS 9	ACACGGAAGTCATCTGAAGCTCAAGATAGGTAGAGATAACTGCTCAATATCTTTTCGAAATTTCCTTAGC
OLS 10	ACTGCTCAGATATTTAACTTCCGTGTTATCTCTTTCGAACCTATCTTCATCTGAAGCTCAAGGAAATAGC
OLS 11	ACTAAAGGAAATTGCTCGAGATAACAAGATAGGCATCTGAAGCTCACTTCCGTTTTCGAAAGATATTAGC
OLS 12	ACCATCTGAAGCTCAATATCTTTTCGAAAAGGAAATTGCTCGAGATAACTAACGGAAGTCCTATCTTAGC
OLS 13	ACTAATTTCCTTTTTCGAAGAGATAACCCTATCTTCATCTGAAGCTCACTTCCGTTGCTCAATATCTAGC
OLS 14	ACTTTCGAAATTTCCTTCATCTGAAGCTCGAGATAACACGGAAGTTGCTCAATATCTTACCTATCTTAGC
OLS 15	ACCATCTGAAGCTCCCTATCTTTTTCGAAAAGGAAATACGGAAGTTAGTTATCTCTGCTCAATATCTAGC
OLS 16	ACCATCTGAAGCTCATTTCCTTTGCTCAAGATAGGACTTCCGTTTTCGAAGAGATAACTAAATATCTAGC
OLS 17	ACCATCTGAAGCTCAAGATAGGGAGATAACTGCTCACGGAAGTTTTCGAAAATATCTTAAAGGAAATAGC
OLS 18	ACTGCTCAAGGAAATCATCTGAAGCTCGAGATAACTTTCGAAACGGAAGTTAAAGATAGGAGATATTAGC
OLS 19	ACCATCTGAAGCTCATTTCCTTTTTCGAACCTATCTTTGCTCAATATCTACGGAAGTTAGAGATAACAGC
OLS 20	ACTACCTATCTTTTTCGAAAAGGAAATTGCTCACTTCCGTAGATATTCATCTGAAGCTCGTTATCTCAGC
OLS 21	ACTAAAGGAAATTGCTCGAGATAACCATCTGAAGCTCAAGATAGGTTTCGAAACGGAAGTAGATATTAGC
OLS 22	ACCATCTGAAGCTCACGGAAGTAGATATTTGCTCAAGATAGGTTTCGAAATTTCCTTTAGAGATAACAGC
OLS 23	ACTGCTCCCTATCTTCATCTGAAGCTCGAGATAACTAACTTCCGTAGATATTTTTCGAAATTTCCTTAGC
OLS 24	ACTACCTATCTTTTTCGAAGTTATCTCCATCTGAAGCTCATTTCCTTTGCTCACGGAAGTAATATCTAGC
OLS 25	ACTAAATATCTCATCTGAAGCTCACTTCCGTTTTCGAAGTTATCTCCCTATCTTTGCTCATTTCCTTAGC
OLS 26	ACCATCTGAAGCTCAGATATTTGCTCACGGAAGTGTTATCTCTTTCGAAAAGATAGGTAATTTCCTTAGC
OLS 27	ACATTTCCTTTTTCGAACCTATCTTTAACTTCCGTTGCTCGAGATAACCATCTGAAGCTCAGATATTAGC
OLS 28	ACTAACTTCCGTTTTCGAAAAGATAGGCATCTGAAGCTCAGATATTATTTCCTTTGCTCGTTATCTCAGC
OLS 29	ACCATCTGAAGCTCGTTATCTCATTTCCTTTAAGATATTTGCTCACTTCCGTTTTCGAACCTATCTTAGC
OLS 30	ACCATCTGAAGCTCAAGGAAATTGCTCACTTCCGTTTTCGAACCTATCTTTAGTTATCTCAGATATTAGC

Table 1: Sequence of all enhancer constructs. (A) OLS 1-30, (B) OLS 31-60, continued.



B

Otxa WT	TTGGATCTGAAGCTCGTTATCTCTAACGGAAGTTTTCGAAAAGGAAATTGTTCAATATCTAAGATAGGA
OLS 31	GATCTGAAGCTCGTTATCTCACGGAAGTTTTACGCGAAAAGGAAATTGCAATTCAATATCTTGAACTCAAGATAGGAGC
OLS 32	GATCTGAAGCTCGTTATCTCTACGGAAGTTTTACGTCGAAAAGGAAATTGTCAATATCTTGATCAAGATAGGAGC
OLS 33	GATCTGAAGCTCGTTATCTCTCAACGGAAGTTGAAAAGGAAATTGCACATTTCAATATCTTGCAAGATAGGAGC
OLS 34	GATCTGAAGCTCGTTATCTCTCTTGAAACGGAAGTTTTACCGAAAAGGAAATTCAATATCTTGAAGCTCAAGATAGGAGC
OLS 35	GATCTGAAGCTCGTTATCTCTCTTACGCGAAACGGAAGTTAAAAGGAAATTGCACCATTTCAATATCTAAGATAGGAGC
OLS 36	CATCTGAAGCTCACGGAAGTTGCTCAAGATAGGTAGAGATAACAGATATTTTTCGAAATTTCCTTAGC
OLS 37	TGCTCACTTCCGTTACCTATCTTTTTCGAAGAGATAACCATCTGAAGCTCAGATATTATTTCCTTAGC
OLS 38	TGCTCCCTATCTTTAACGGAAGTAAGGAAATTTTCGAAGTTATCTCCATCTGAAGCTCAGATATTAGC
OLS 39	TAAAGGAAATTTTCGAAAGATATTTGCTCGTTATCTCCATCTGAAGCTCACTTCCGTAAGATAGGAGC
OLS 40	TAGTTATCTCTTTCGAAAAGATAGGTGCTCACTTCCGTAAGGAAATCATCTGAAGCTCAATATCTAGC
OLS 41	CATCTGAAGCTCGTTATCTCTGCTCCCTATCTTTTTCGAAATTTCCTTTAACTTCCGTAGATATTAGC
OLS 42	CATCTGAAGCTCACTTCCGTTGCTCCCTATCTTTTTCGAAATTTCCTTTAGTTATCTCAGATATTAGC
OLS 43	TGCTCAGATATTCATCTGAAGCTCAAGGAAATGTTATCTCTACCTATCTTTTTCGAAACTTCCGTAGC
OLS 44	TGCTCAGATATTCATCTGAAGCTCAAGGAAATGTTATCTCTACCTATCTTTTTCGAAACGGAAGTAGC
OLS 45	TGCTCCCTATCTTAGATATTTAACTTCCGTTTTCGAAAAGGAAATCATCTGAAGCTCGAGATAACAGC
OLS 46	TGCTCCCTATCTTAGATATTTAACTTCCGTTTTCGAAAAGGAAATCATCTGAAGCTCGTTATCTCAGC
OLS 47	ATTTCCTTCATCTGAAGCTCGAGATAACTTTCGAAACGGAAGTTGCTCAAGATAGGTAAATATCTAGC
OLS 48	GAGATAACCATCTGAAGCTCATTTCCTTTTTCGAAACGGAAGTTGCTCAAGATAGGTAAATATCTAGC
OLS 49	CATCTGAAGCTCAATATCTTGCTCACTTCCGTAAGATAGGTTTCGAAATTTCCTTTAGAGATAACAGC
OLS 50	CATCTGAAGCTCAATATCTTGCTCACTTCCGTAAGATAGGTTTCGAAAAGGAAATTAGAGATAACAGC
OLS 51	AGATATTTGCTCACTTCCGTCATCTGAAGCTCCCTATCTTTTTCGAAATTTCCTTTAGTTATCTCAGC
OLS 52	AGATATTTGCTCACTTCCGTCATCTGAAGCTCAAGATAGGTTTCGAAATTTCCTTTAGTTATCTCAGC
OLS 53	AATATCTTTTCGAAGAGATAACTGCTCATTTCCTTTAACGGAAGTCATCTGAAGCTCCCTATCTTAGC
OLS 54	AATATCTTTTCGAAATTTCCTTTGCTCGAGATAACTAACGGAAGTCATCTGAAGCTCCCTATCTTAGC
OLS 55	TGCTCGAGATAACCCTATCTTTAAAGGAAATTTTCGAAACTTCCGTCATCTGAAGCTCAATATCTAGC
OLS 56	TGCTCGTTATCTCCCTATCTTTAAAGGAAATTTTCGAAACTTCCGTCATCTGAAGCTCAATATCTAGC
OLS 57	TTTCGAAGAGATAACTAAAGATAGGTGCTCATTTCCTTACTTCCGTCATCTGAAGCTCAGATATTAGC
OLS 58	TTTCGAAACTTCCGTTAAAGATAGGTGCTCATTTCCTTGAGATAACCATCTGAAGCTCAGATATTAGC
OLS 59	TAACTTCCGTCATCTGAAGCTCCCTATCTTTGCTCGAGATAACTTTCGAAAATATCTATTTCCTTAGC
OLS 60	TAACTTCCGTCATCTGAAGCTCATTTCCTTTGCTCGAGATAACTTTCGAAAATATCTCCTATCTTAGC

Figure 4: Percentage of embryos with tissue-specific expression for OLS 31-60. Grammatical variants were organized from low to high percentage of well-developed embryos with GFP expression. Blue, green, purple, pink, and yellow columns represent anterior sensory vesicle, dorsal nerve cord, mesenchyme, endoderm, and notochord tissue expression, respectively. (A) Total embryos with GFP expression 0-10%, (B) 11-30%, (C) 31-60%, (D) 61-100%. All grammatical variants had a minimum of two biological repeats performed with a minimum of n=50 well-developed embryos counted. The only exception was OLS 33, which had one experiment with n=47. Error bars represent the standard error of the mean.

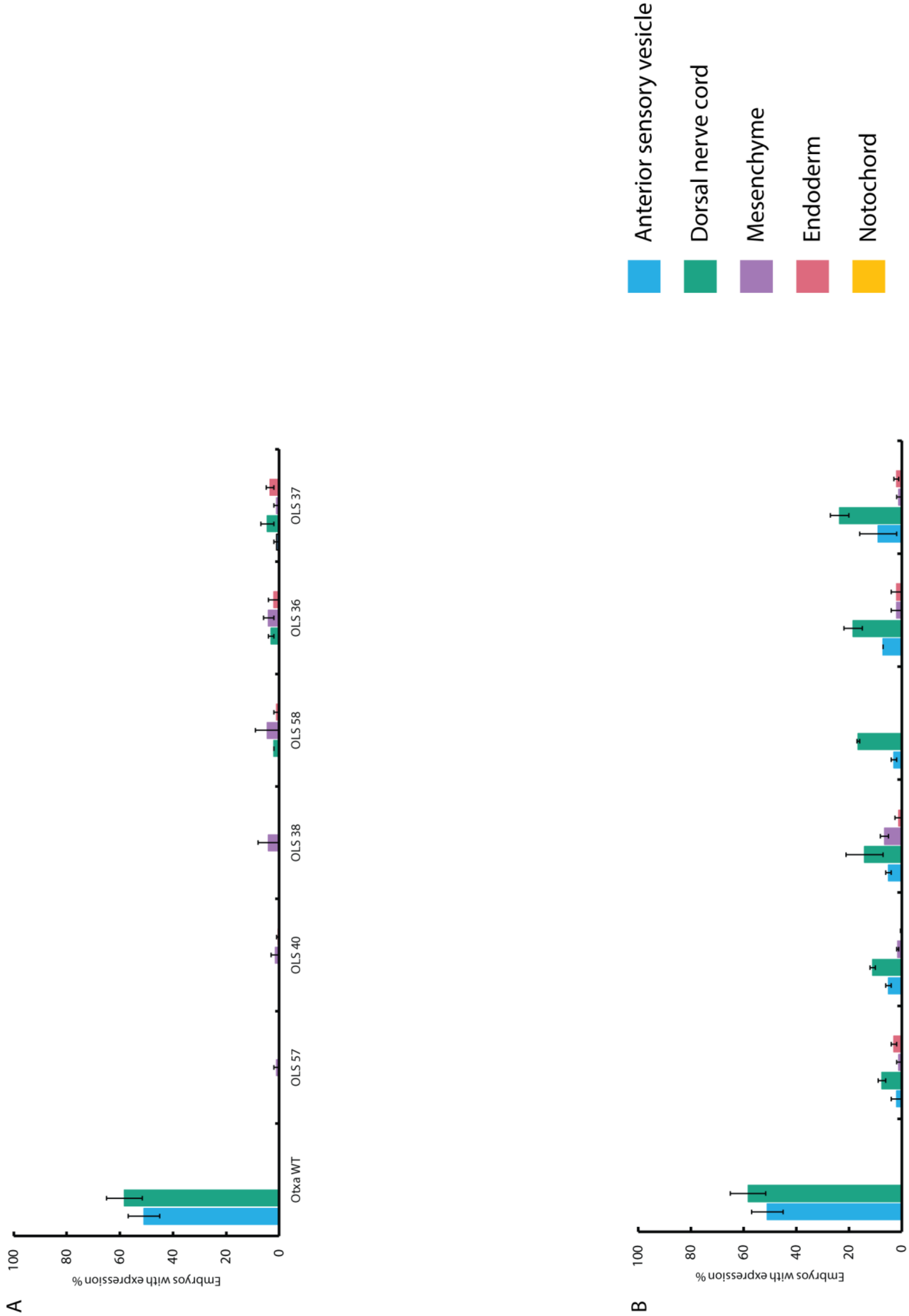

A
Embryos with expression %
100
80
60
40
20
0
Otxa WT
OLS 57
OLS 40
OLS 38
OLS 58
OLS 36
OLS 37
B
Embryos with expression %
100
80
60
40
20
0
Otxa WT
OLS 59
OLS 41
OLS 55
OLS 31
OLS 52
OLS 60
Anterior sensory vesicle
Dorsal nerve cord
Mesenchyme
Endoderm
Notochord

Figure 4: Percentage of embryos with tissue-specific expression for OLS 31-60. Grammatical variants were organized from low to high percentage of well-developed embryos with GFP expression. Blue, green, purple, pink, and yellow columns represent anterior sensory vesicle, dorsal nerve cord, mesenchyme, endoderm, and notochord tissue expression, respectively. (A) Total embryos with GFP expression 0-10%, (B) 11-30%, (C) 31-60%, (D) 61-100%. All grammatical variants had a minimum of two biological repeats performed with a minimum of n=50 well-developed embryos counted. The only exception was OLS 33, which had one experiment with n=47. Error bars represent the standard error of the mean, continued.

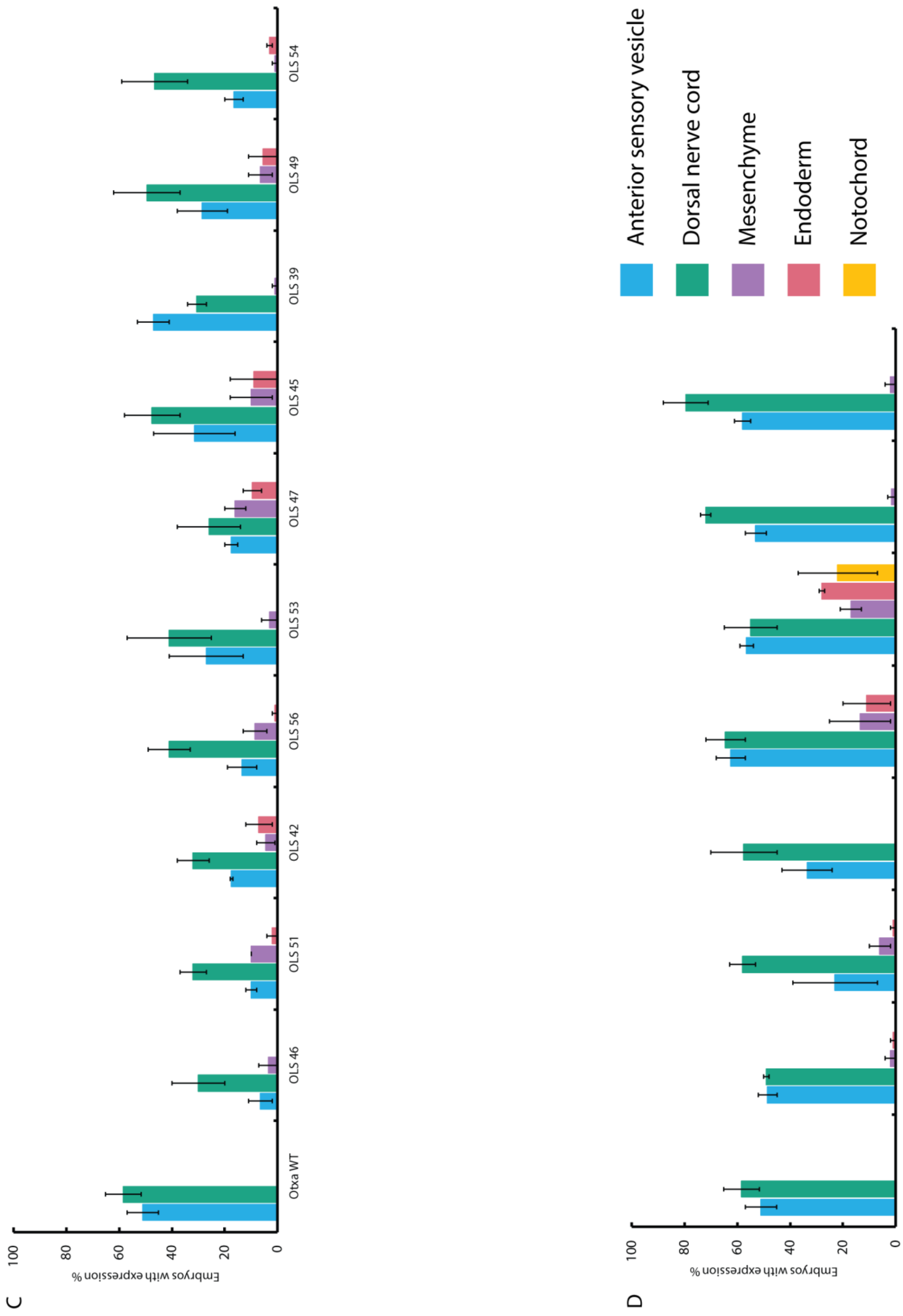
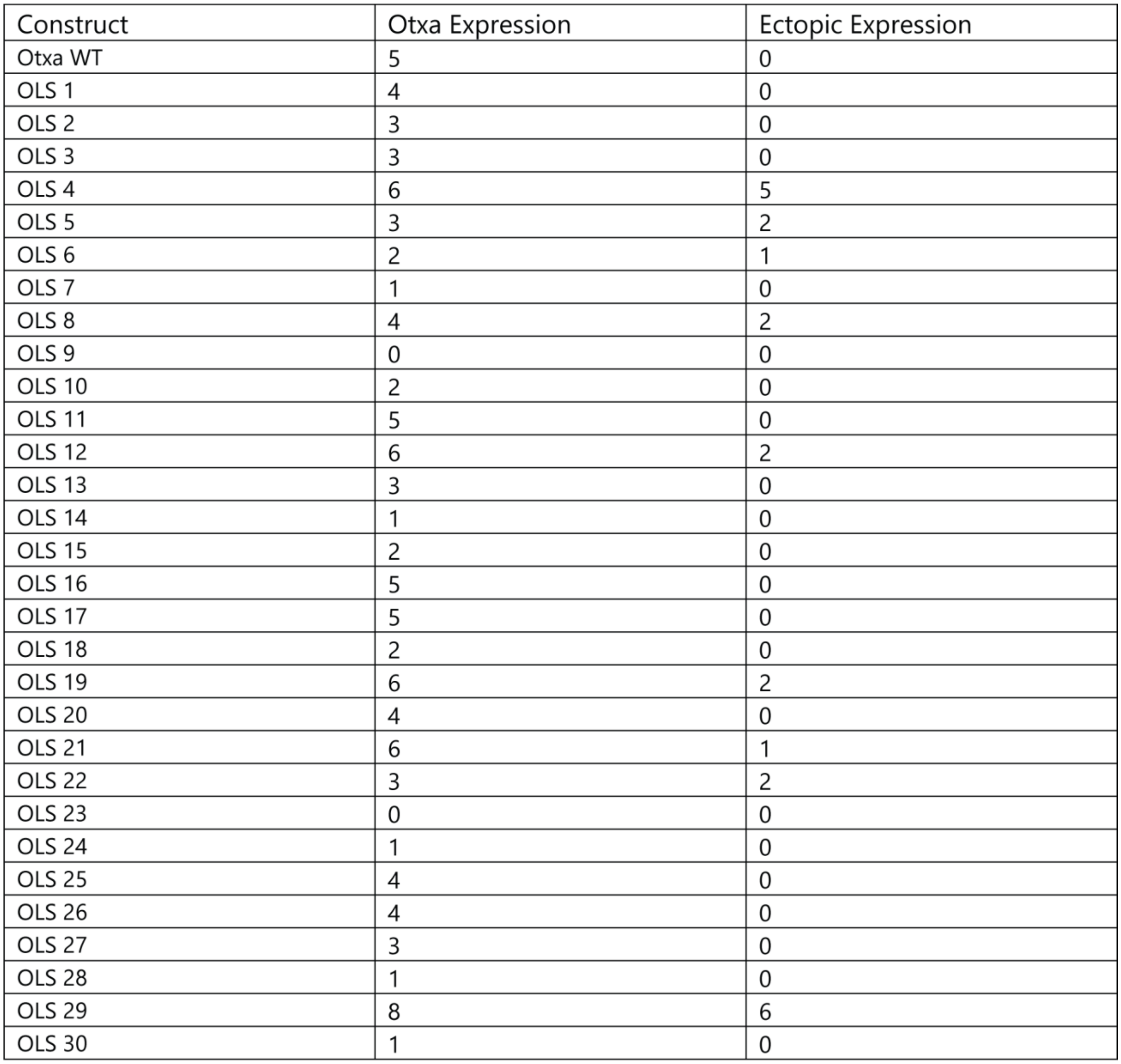

C
Embryos with expression %
100
80
60
40
20
0
Otxa WT
OLS 46
OLS 51
OLS 42
OLS 56
OLS 53
OLS 47
OLS 45
OLS 39
OLS 49
OLS 54
D
Embryos with expression %
100
80
60
40
20
0
Otxa WT
OLS 44
OLS 35
OLS 32
OLS 43
OLS 48
OLS 33
OLS 34
Anterior sensory vesicle
Dorsal nerve cord
Mesenchyme
Endoderm
Notochord

Table 2: Expression levels of grammatical variants OLS 1-30 from experiments done by rotation members in the laboratory. Levels were set on a scale of 1-10 with wild-type expression set at 5. Otx-a expression includes anterior sensory vesicle and dorsal nerve cord expression. Ectopic expression includes any other tissue that is not seen in normal wild-type Otx-a expression.



Construct	Otxa Expression	Ectopic Expression
Otxa WT	5	0
OLS 1	4	0
OLS 2	3	0
OLS 3	3	0
OLS 4	6	5
OLS 5	3	2
OLS 6	2	1
OLS 7	1	0
OLS 8	4	2
OLS 9	0	0
OLS 10	2	0
OLS 11	5	0
OLS 12	6	2
OLS 13	3	0
OLS 14	1	0
OLS 15	2	0
OLS 16	5	0
OLS 17	5	0
OLS 18	2	0
OLS 19	6	2
OLS 20	4	0
OLS 21	6	1
OLS 22	3	2
OLS 23	0	0
OLS 24	1	0
OLS 25	4	0
OLS 26	4	0
OLS 27	3	0
OLS 28	1	0
OLS 29	8	6
OLS 30	1	0

Figure 5: Principal Component Analysis of order and orientation syntax on tissue-specific expression. (A) Every order tested in OLS 31-60 and wild-type Otx-a plotted. (B) Every orientation of binding sites tested in OLS 31-60 and wild-type Otx-a plotted. Orientation labels correspond with the order of binding sites in wild-type Otx-a (e.g. orientation of GATA-1, ETS-1, etc.), regardless of the order of binding sites within the grammatical variant. Order and orientation combinations with a PCA coefficient less than 0.07 were not plotted to increase graph visibility. Each PCA biplot has the same expression attributes plotted, which allow us to find correlations between expression and syntax. Red dots show PCA scores, which are each grammatical variant's data projected on the principal components.

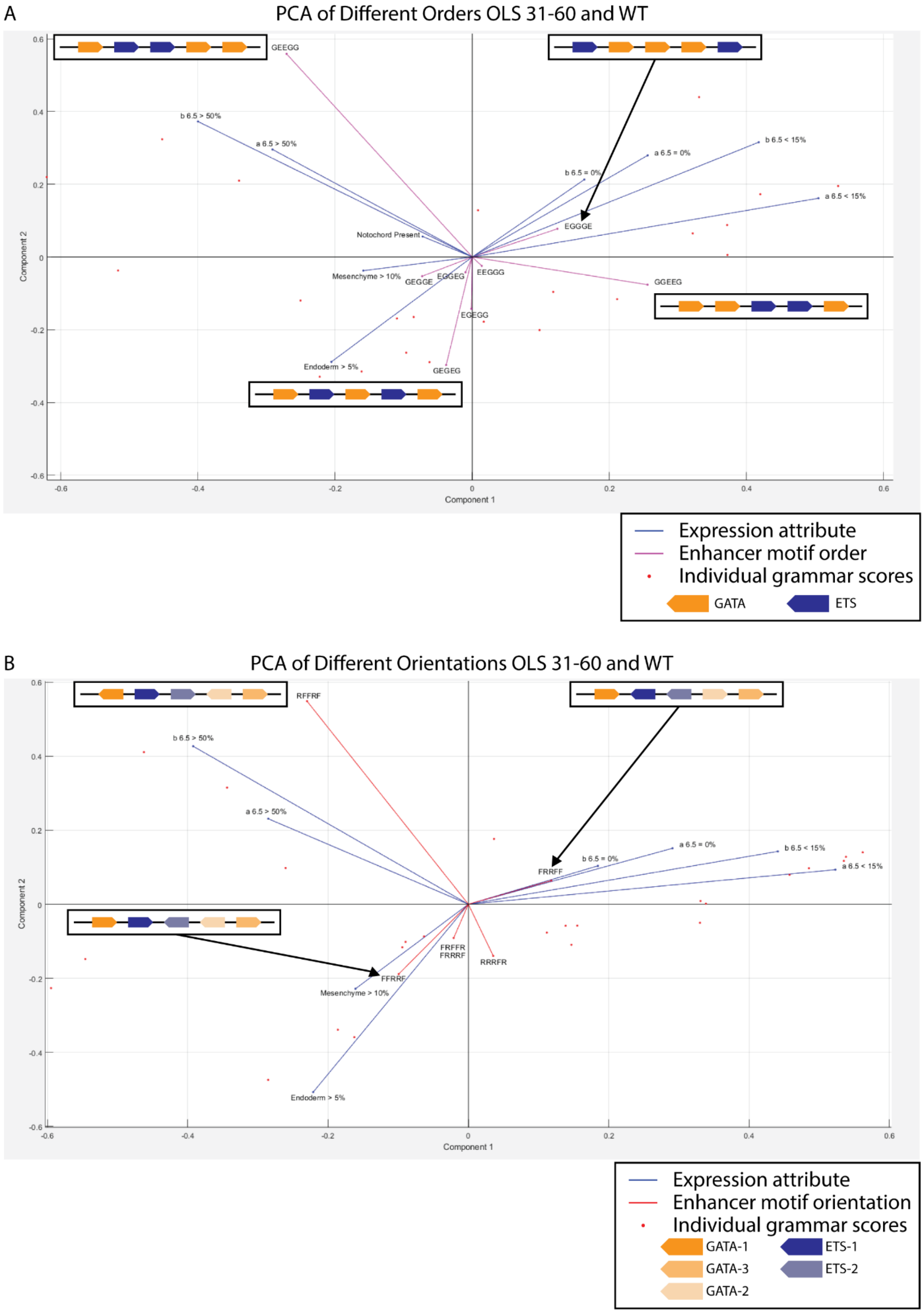

A
PCA of Different Orders OLS 31-60 and WT
GEEGG
b 6.5 > 50%
a 6.5 > 50%
Notochord Present
Mesenchyme > 10%
GEGGE
EGGEG
EEGGG
EGEGG
Endoderm > 5%
GEGEG
b 6.5 = 0%
a 6.5 = 0%
b 6.5 < 15%
a 6.5 < 15%
EGGGE
GGEEG
Component 1
Component 2
0.6
0.4
0.2
0
-0.2
-0.4
-0.6
Expression attribute
Enhancer motif order
Individual grammar scores
GATA
ETS
B
PCA of Different Orientations OLS 31-60 and WT
RFFRF
b 6.5 > 50%
a 6.5 > 50%
FRRFF
b 6.5 = 0%
a 6.5 = 0%
b 6.5 < 15%
a 6.5 < 15%
FRFFR
FRRRF
RRRFR
FFRRF
Mesenchyme > 10%
Endoderm > 5%
Component 1
Component 2
Expression attribute
Enhancer motif orientation
Individual grammar scores
GATA-1
GATA-3
GATA-2
ETS-1
ETS-2

Figure 6: Enhancer grammatical variants that have no anterior sensory vesicle or dorsal nerve cord expression. Enhancer schematics show the order, orientation, and spacing (core to core) of ETS and GATA binding sites within the WT enhancer along with grammatical variants that had no expression in anterior sensory vesicle or dorsal nerve cord. (A) Max intensity projection (MIP) of embryo electroporated with WT Otx-a enhancer showing the endogenous expression pattern. (B, D) DIC image of the embryo, focused on the neural plane to show notochord and anterior brain structure. (C) MIP of embryo electroporated with construct OLS 57, showing no expression in any tissue. This image is representative of OLS 40 and 38, which had no a 6.5 or b 6.5 expression as well. (E) Percentage of well-developed embryos with expression in tissue types compared to wild-type Otx-a. Blue, green, purple, and pink columns represent anterior sensory vesicle, dorsal nerve cord, mesenchyme, and endoderm tissue expression, respectively. All grammatical variants had a minimum of two biological repeats performed, with a minimum of n=50 well-developed embryos in each count. Error bars represent the standard error of the mean. (F) Expression level of grammatical variants with ETS sites touching within OLS 1-30 dataset. Levels were set on a scale of 1-10 with wild-type expression set at 5. Otx-a expression includes anterior sensory vesicle and dorsal nerve cord expression. Ectopic expression includes any other tissue that is not seen in normal wild-type Otx-a expression.

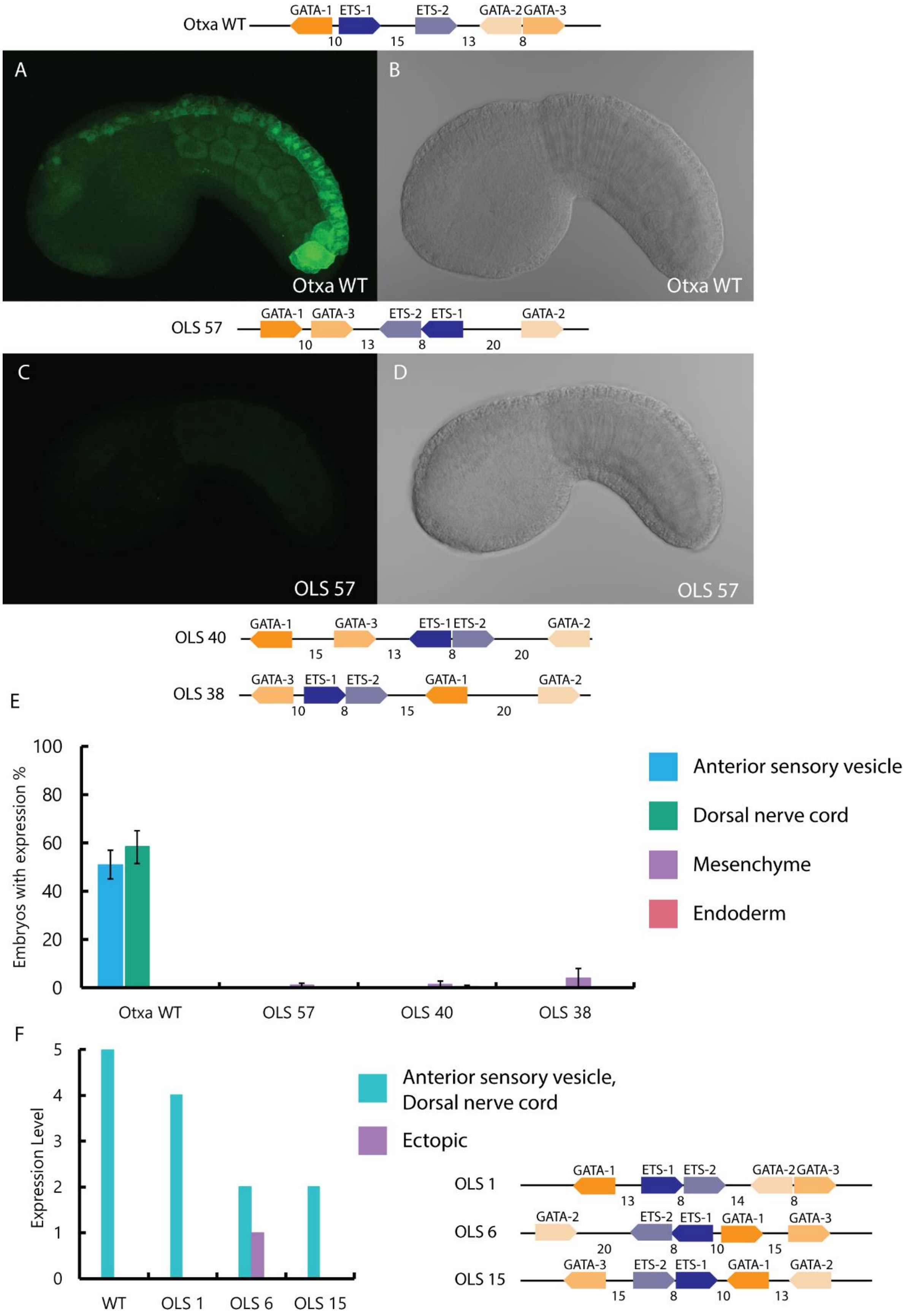
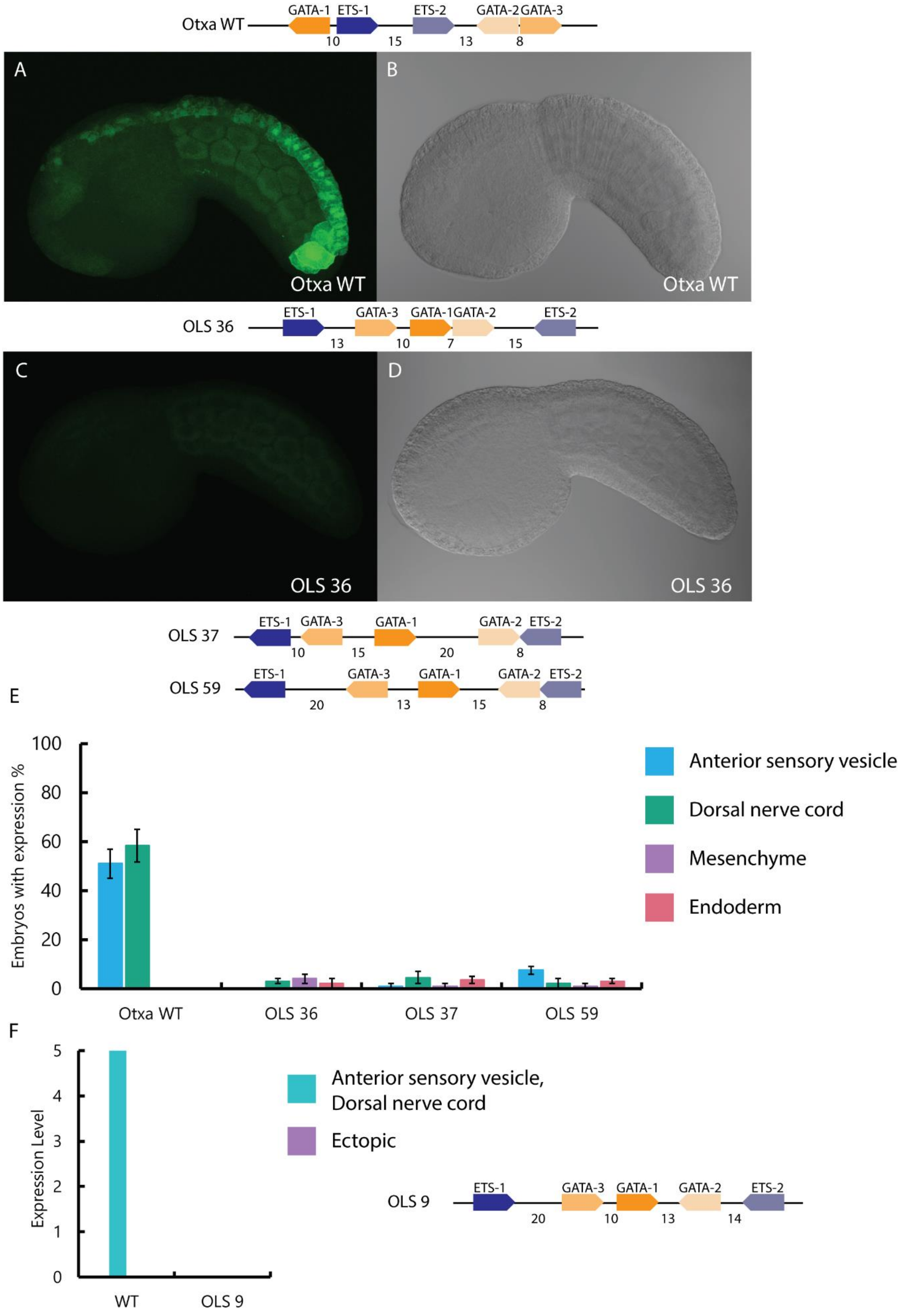
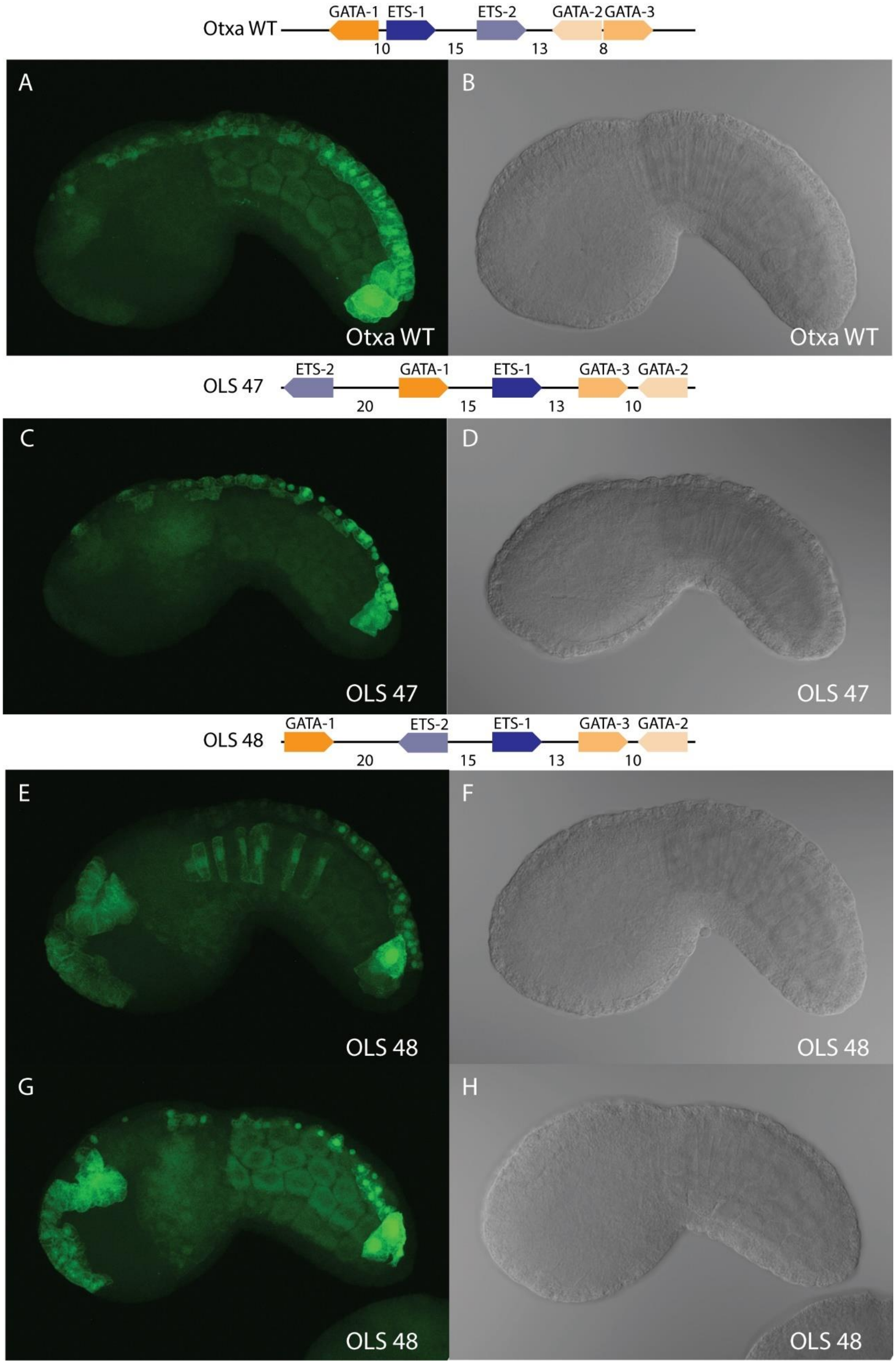

Otxa WT
GATA-1
ETS-1
ETS-2
GATA-2
GATA-3
10
15
13
8
A
Otxa WT
B
Otxa WT
OLS 57
GATA-1
GATA-3
ETS-2
ETS-1
GATA-2
10
13
8
20
C
OLS 57
D
OLS 57
OLS 40
GATA-1
GATA-3
ETS-1
ETS-2
GATA-2
15
13
8
20
OLS 38
GATA-3
ETS-1
ETS-2
GATA-1
GATA-2
10
8
15
20
E
Embryos with expression %
100
80
60
40
20
0
Otxa WT
OLS 57
OLS 40
OLS 38
Anterior sensory vesicle
Dorsal nerve cord
Mesenchyme
Endoderm
F
Expression Level
5
4
3
2
1
0
WT
OLS 1
OLS 6
OLS 15
Anterior sensory vesicle, Dorsal nerve cord
Ectopic
OLS 1
GATA-1
ETS-1
ETS-2
GATA-2
GATA-3
13
8
14
8
OLS 6
GATA-2
ETS-2
ETS-1
GATA-1
GATA-3
20
8
10
15
OLS 15
GATA-3
ETS-2
ETS-1
GATA-1
GATA-2
15
8
10
13

Figure 7: Enhancer grammatical variants that have very low anterior sensory vesicle and dorsal nerve cord expression. Enhancer schematics show the order, orientation, and spacing (core to core) of ETS and GATA binding sites within the WT enhancer along with variants that had very low expression in anterior sensory vesicle or dorsal nerve cord. (A) Max intensity projection (MIP) of embryo electroporated with WT Otx-a enhancer showing the endogenous expression pattern. (B, D) DIC image of the embryo, focused on the neural plane to show notochord and anterior brain structure. (C) MIP of embryo electroporated with grammatical variant OLS 36, showing no expression in any tissue. This image is representative of OLS 37 and 59, which had very few embryos with a 6.5 or b 6.5 expression as well. (E) Percentage of well-developed embryos with expression in tissue types compared to wild-type Otx-a. Blue, green, purple, and pink columns represent anterior sensory vesicle, dorsal nerve cord, mesenchyme, and endoderm tissue expression, respectively. All constructs had a minimum of two biological repeats performed, with a minimum of n=50 well-developed embryos in each count. Error bars represent the standard error of the mean**.** (F) Expression level of grammatical variants with ETS sites on ends of enhancer within OLS 1-30 dataset. Levels were set on a scale of 1-10 with wild-type expression set at 5. Otx-a expression includes anterior sensory vesicle and dorsal nerve cord expression. Ectopic expression includes any other tissue that is not seen in normal wild-type Otx-a expression.



Otxa WT
GATA-1
ETS-1
ETS-2
GATA-2
GATA-3
10
15
13
8
A
Otxa WT
B
Otxa WT
OLS 36
ETS-1
GATA-3
GATA-1
GATA-2
ETS-2
13
10
7
15
C
OLS 36
D
OLS 36
OLS 37
ETS-1
GATA-3
GATA-1
GATA-2
ETS-2
10
15
20
8
OLS 59
ETS-1
GATA-3
GATA-1
GATA-2
ETS-2
20
13
15
8
E
Embryos with expression %
100
80
60
40
20
0
Otxa WT
OLS 36
OLS 37
OLS 59
Anterior sensory vesicle
Dorsal nerve cord
Mesenchyme
Endoderm
F
Expression Level
5
4
3
2
1
0
WT
OLS 9
Anterior sensory vesicle, Dorsal nerve cord
Ectopic
OLS 9
ETS-1
GATA-3
GATA-1
GATA-2
ETS-2
20
10
13
14

Figure 8: Enhancer grammatical variant with notochord expression. Enhancer schematics show the order, orientation, and spacing (core to core) of ETS and GATA binding sites within the WT enhancer along with the variant with notochord expression, OLS 48, and its pair OLS 47 that is very similar. (A) Max intensity projection (MIP) of embryo electroporated with WT Otx-a enhancer showing the endogenous expression pattern. (B, D, F, H) DIC image of the embryo, focused on the neural plane to show notochord and anterior brain structure. (C) MIP of embryo electroporated with construct OLS 47, showing an embryo with a 6.5, b 6.5, and endoderm tissue expression. (E) MIP of embryo electroporated with construct OLS 48, showing an embryo with notochord expression. (G) MIP of embryo electroporated with construct OLS 48, showing the average tissue expression present in a 6.5, b 6.5, and mesenchyme tissues. (I) Percentage of well-developed embryos with expression in tissue types compared to wild-type Otx-a. Blue, green, purple, and pink columns represent anterior sensory vesicle, dorsal nerve cord, mesenchyme, and endoderm tissue expression, respectively. All constructs had a minimum of two biological repeats performed, with a minimum of n=50 well-developed embryos in each count. Error bars represent the standard error of the mean**.**


Otxa WT
GATA-1
ETS-1
ETS-2
GATA-2
GATA-3
10
15
13
8
A
Otxa WT
B
Otxa WT
OLS 47
ETS-2
GATA-1
ETS-1
GATA-3
GATA-2
20
15
13
10
C
OLS 47
D
OLS 47
OLS 48
GATA-1
ETS-2
ETS-1
GATA-3
GATA-2
20
15
13
10
E
OLS 48
F
OLS 48
G
OLS 48
H
OLS 48

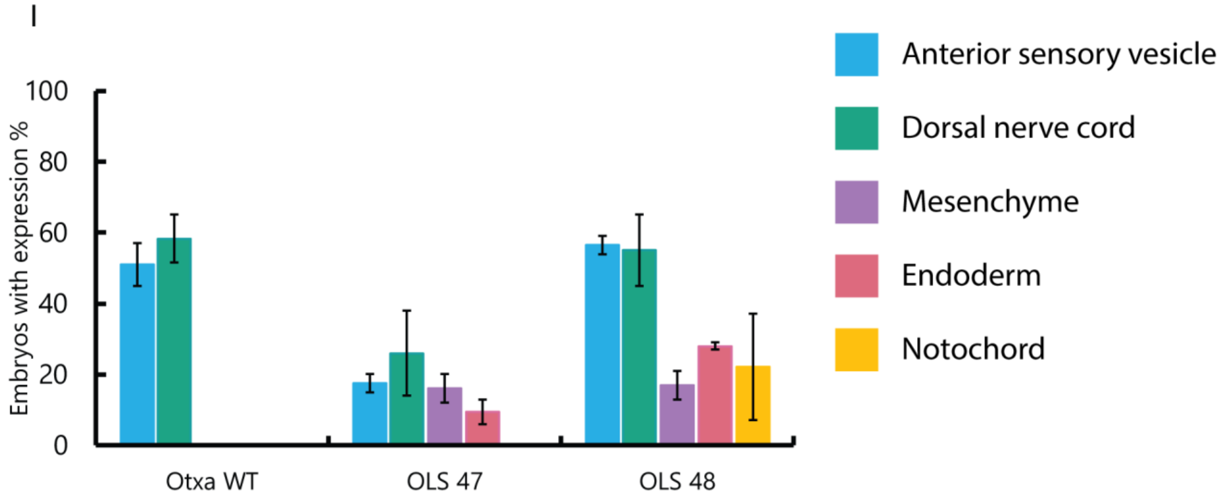


Figure 8: Enhancer grammatical variant with notochord expression. Enhancer schematics show the order, orientation, and spacing (core to core) of ETS and GATA binding sites within the WT enhancer along with the variant with notochord expression, OLS 48, and its pair OLS 47 that is very similar. (A) Max intensity projection (MIP) of embryo electroporated with WT Otx-a enhancer showing the endogenous expression pattern. (B, D, F, H) DIC image of the embryo, focused on the neural plane to show notochord and anterior brain structure. (C) MIP of embryo electroporated with construct OLS 47, showing an embryo with a 6.5, b 6.5, and endoderm tissue expression. (E) MIP of embryo electroporated with construct OLS 48, showing an embryo with notochord expression. (G) MIP of embryo electroporated with construct OLS 48, showing the average tissue expression present in a 6.5, b 6.5, and mesenchyme tissues. (I) Percentage of well-developed embryos with expression in tissue types compared to wild-type Otx-a. Blue, green, purple, pink, and yellow columns represent anterior sensory vesicle, dorsal nerve cord, mesenchyme, endoderm, and notochord tissue expression, respectively. All constructs had a minimum of two biological repeats performed, with a minimum of n=50 well-developed embryos in each count. Error bars represent the standard error of the mean, continued.

Figure 9: Gene expression of OLS 47 and 48 grammatical variants with knocked out ETS site. Enhancer schematics show the order, orientation, and spacing (core to core) of ETS and GATA binding sites within the WT enhancer along with OLS grammatical variants with and without ETS-2 binding sites. (A) Max intensity projection (MIP) of embryo electroporated with WT Otx-a enhancer showing the endogenous expression pattern. (B, D, F, H, J, L) DIC image of the embryo, focused on the neural plane to show notochord and anterior brain structure. (C) MIP of embryo electroporated with construct OLS 47, showing an embryo with a 6.5, b 6.5, and endoderm tissue expression, and its pair (E) with a knocked out ETS-2 site. (G) MIP of embryo electroporated with construct OLS 48, showing an embryo with notochord expression, (I) average expression, and its pair (K) with a knocked-out ETS-2 site. (M) Percentage of well-developed embryos with expression in tissue types compared to wild-type Otx-a. Blue, green, purple, and pink columns represent anterior sensory vesicle, dorsal nerve cord, mesenchyme, and endoderm tissue expression, respectively. All grammatical variants had a minimum of two biological repeats performed, with a minimum of n=50 well-developed embryos in each count. Error bars represent the standard error of the mean**.**

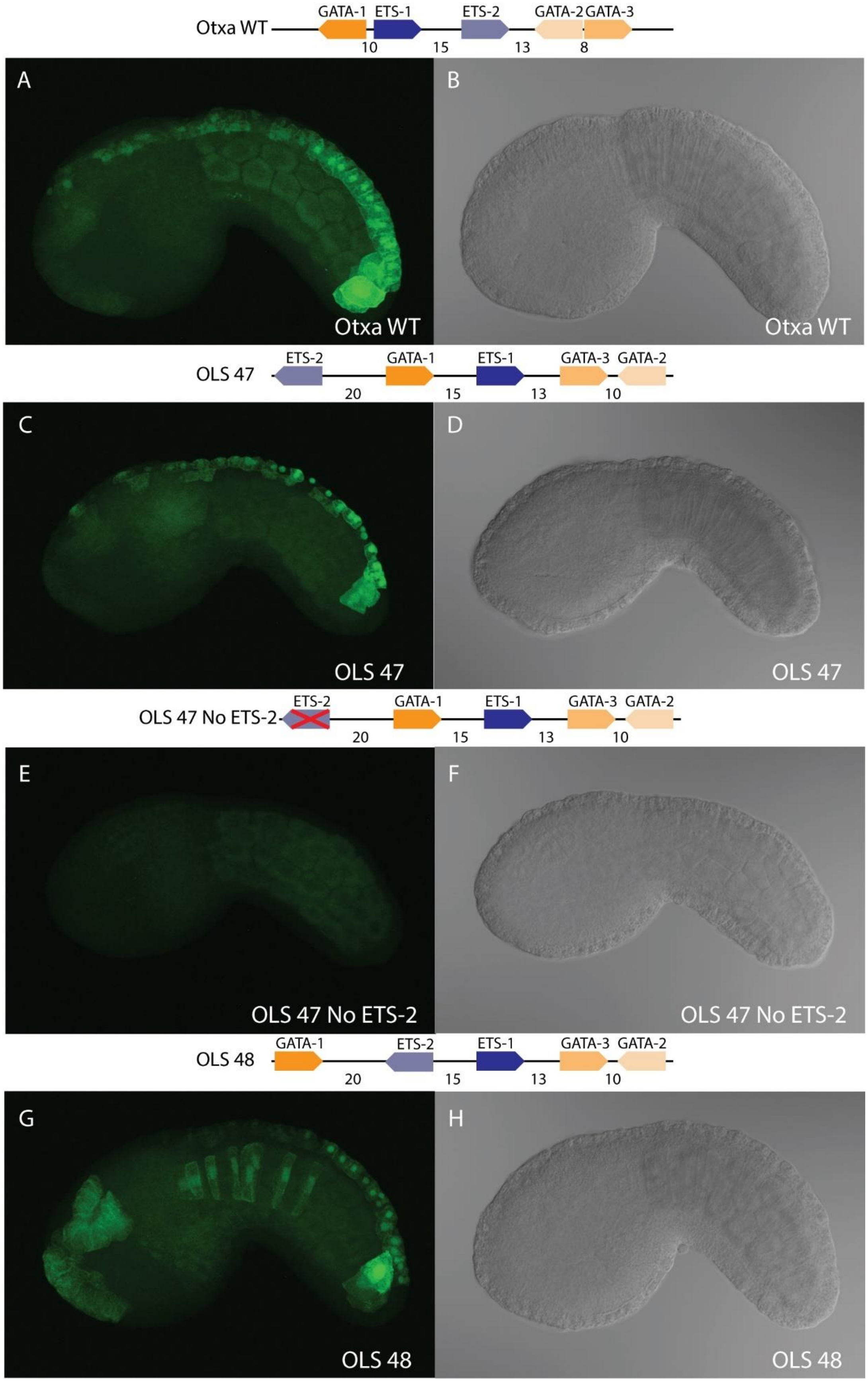

Otxa WT
GATA-1
ETS-1
ETS-2
GATA-2
GATA-3
10
15
13
8
A
Otxa WT
B
Otxa WT
OLS 47
ETS-2
GATA-1
ETS-1
GATA-3
GATA-2
20
15
13
10
C
OLS 47
D
OLS 47
OLS 47 No ETS-2
ETS-2
GATA-1
ETS-1
GATA-3
GATA-2
20
15
13
10
E
OLS 47 No ETS-2
F
OLS 47 No ETS-2
OLS 48
GATA-1
ETS-2
ETS-1
GATA-3
GATA-2
20
15
13
10
G
OLS 48
H
OLS 48

Figure 9: Gene expression of OLS 47 and 48 grammatical variants with knocked out ETS site. Enhancer schematics show the order, orientation, and spacing (core to core) of ETS and GATA binding sites within the WT enhancer along with OLS grammatical variants with and without ETS-2 binding sites. (A) Max intensity projection (MIP) of embryo electroporated with WT Otx-a enhancer showing the endogenous expression pattern. (B, D, F, H, J, L) DIC image of the embryo, focused on the neural plane to show notochord and anterior brain structure. (C) MIP of embryo electroporated with construct OLS 47, showing an embryo with a 6.5, b 6.5, and endoderm tissue expression, and its pair (E) with a knocked out ETS-2 site. (G) MIP of embryo electroporated with construct OLS 48, showing an embryo with notochord expression, (I) average expression, and its pair (K) with a knocked-out ETS-2 site. (M) Percentage of well-developed embryos with expression in tissue types compared to wild-type Otx-a. Blue, green, purple, and pink columns represent anterior sensory vesicle, dorsal nerve cord, mesenchyme, and endoderm tissue expression, respectively. All grammatical variants had a minimum of two biological repeats performed, with a minimum of n=50 well-developed embryos in each count. Error bars represent the standard error of the mean**,** continued.

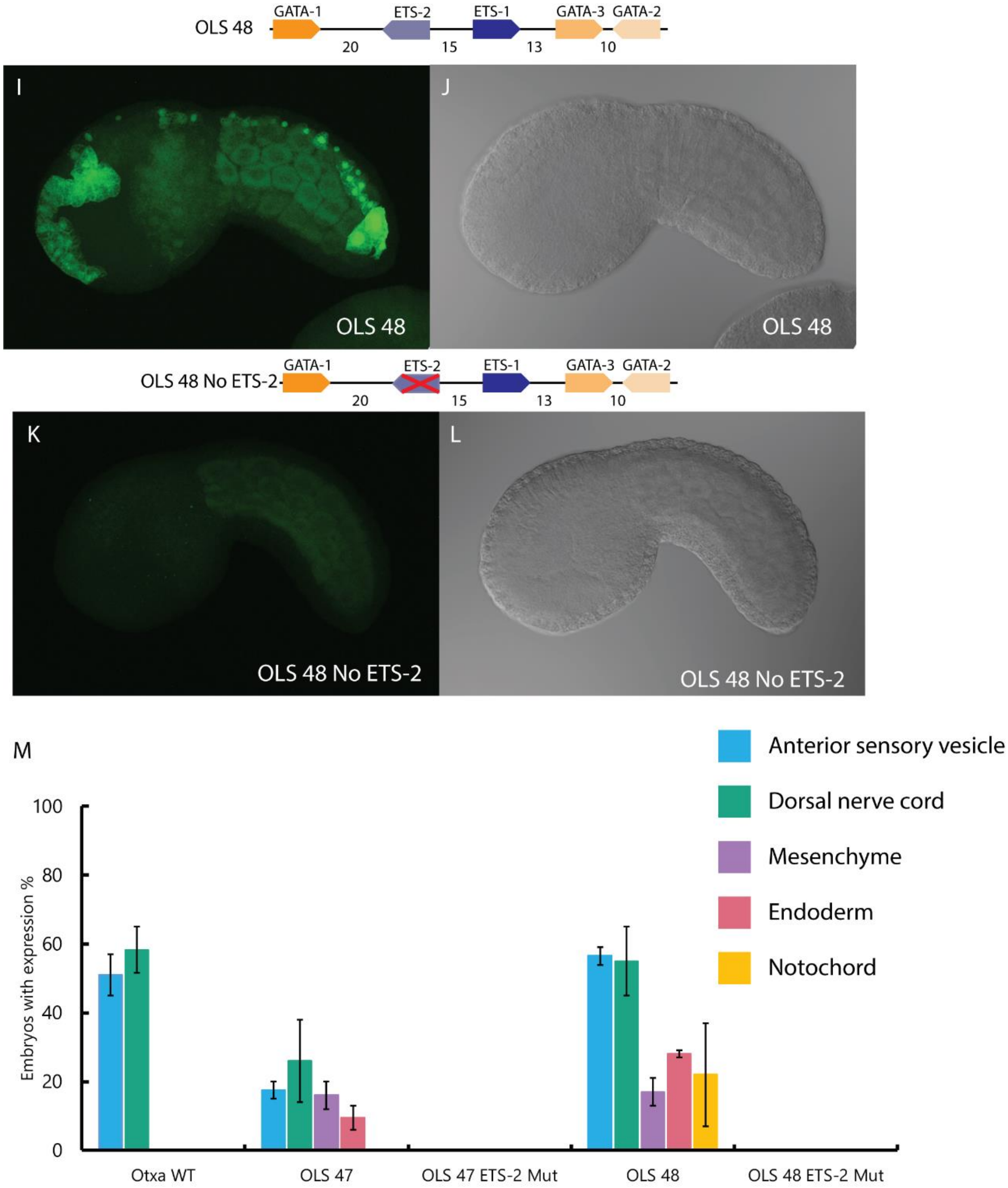
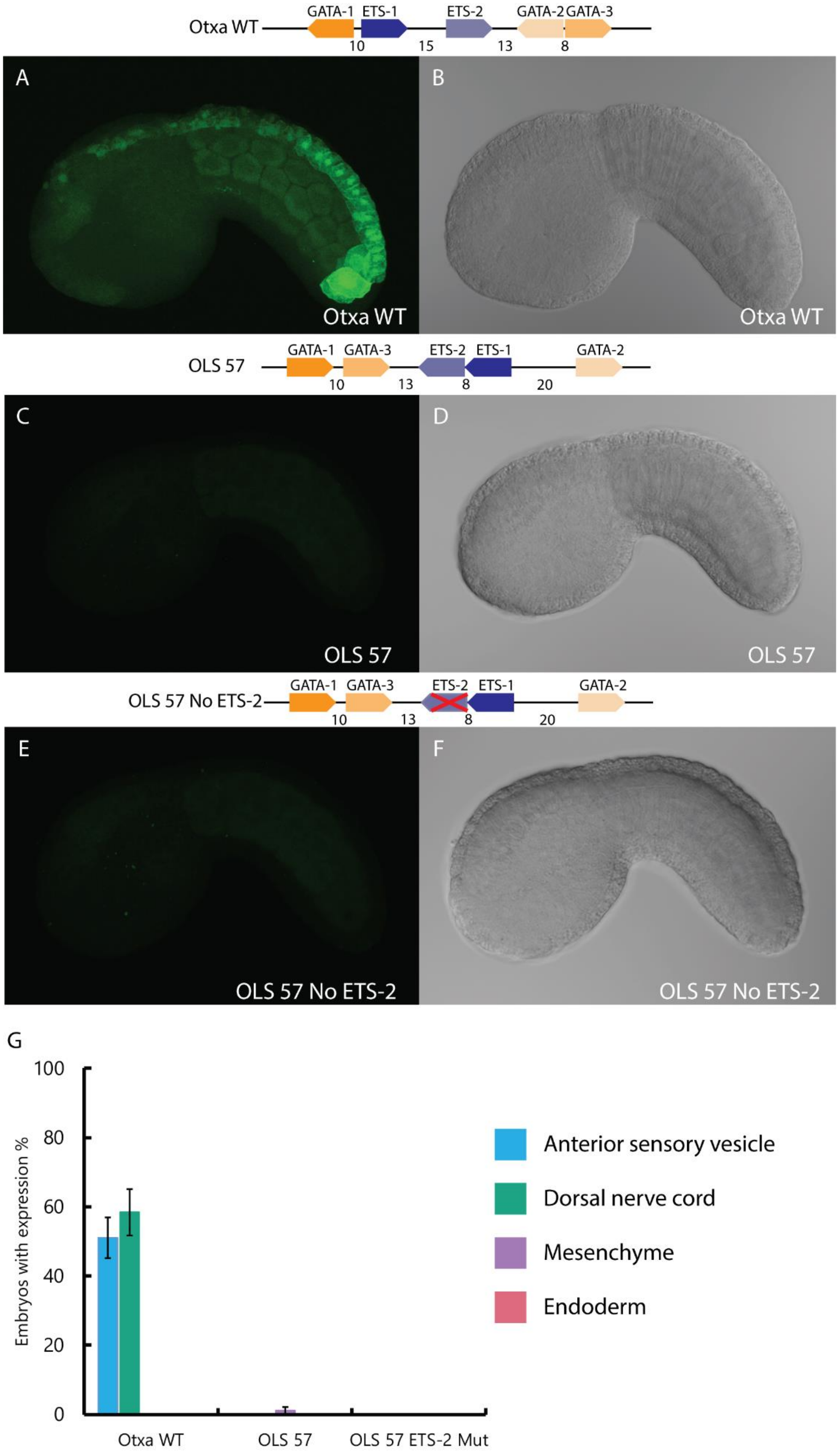

OLS 48
GATA-1
ETS-2
ETS-1
GATA-3
GATA-2
20
15
13
10
I
OLS 48
J
OLS 48
OLS 48 No ETS-2
GATA-1
ETS-2
ETS-1
GATA-3
GATA-2
20
15
13
10
K
OLS 48 No ETS-2
L
OLS 48 No ETS-2
M
Embryos with expression %
100
80
60
40
20
0
Otxa WT
OLS 47
OLS 47 ETS-2 Mut
OLS 48
OLS 48 ETS-2 Mut
Anterior sensory vesicle
Dorsal nerve cord
Mesenchyme
Endoderm
Notochord

Figure 10: Gene expression of OLS 57 grammatical variant with knocked out ETS site. Enhancer schematics show the order, orientation, and spacing (core to core) of ETS and GATA binding sites within the WT enhancer along with OLS grammatical variants with and without ETS-2 binding sites. (A) Max intensity projection (MIP) of embryo electroporated with WT Otx-a enhancer showing the endogenous expression pattern. (B, D, F) DIC image of the embryo, focused on the neural plane to show notochord and anterior brain structure. (C) MIP of embryo electroporated with construct OLS 57, showing an embryo with no tissue expression, and its pair (E) with a knocked out ETS-2 site. (G) Percentage of well-developed embryos with expression in tissue types compared to wild-type Otx-a. Blue, green, purple, and pink columns represent anterior sensory vesicle, dorsal nerve cord, mesenchyme, and endoderm tissue expression, respectively. All grammatical variants had a minimum of two biological repeats performed, with a minimum of n=50 well-developed embryos in each count. Error bars represent the standard error of the mean**.**


Otxa WT
GATA-1
ETS-1
ETS-2
GATA-2
GATA-3
10
15
13
8
A
Otxa WT
B
Otxa WT
OLS 57
GATA-1
GATA-3
ETS-2
ETS-1
GATA-2
10
13
8
20
C
OLS 57
D
OLS 57
OLS 57 No ETS-2
GATA-1
GATA-3
ETS-2
ETS-1
GATA-2
10
13
8
20
E
OLS 57 No ETS-2
F
OLS 57 No ETS-2
G
Embryos with expression %
100
80
60
40
20
0
Otxa WT
OLS 57
OLS 57 ETS-2 Mut
Anterior sensory vesicle
Dorsal nerve cord
Mesenchyme
Endoderm

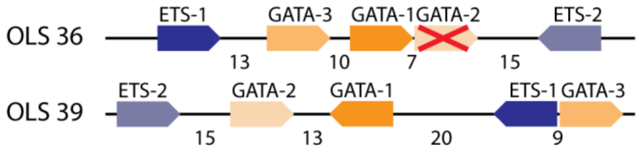


Figure 11: Enhancer schematic of non-functional GATA site in OLS 36. OLS 39 is shown below OLS 36 to show the similarities in structure between ETS sites when GATA-2 is not present.

References

Acampora, D., Annino, A., Tuorto, F., Puelles, E., Lucchesi, W., Papalia, A., Simeone, A., 2005. Otx genes in the evolution of the vertebrate brain. Brain Res. Bull. 66, 410–420.

Arnosti, D.N., Kulkarni, M.M., 2005. Transcriptional enhancers: Intelligent enhanceosomes or flexible billboards? J. Cell. Biochem. 94, 890–898.

Barolo, S., 2016. How to tune an enhancer. Proc. Natl. Acad. Sci. 113, 6330–6331.

Bertrand, V., Hudson, C., Caillol, D., Popovici, C., Lemaire, P., 2003. 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.

Christiaen, L., Wagner, E., Shi, W., Levine, M., Christiaen, L., Wagner, E., Shi, W., Levine, M., 2014. Electroporation of Transgenic DNAs in the Sea Squirt Ciona Protocol Electroporation of Transgenic DNAs in the Sea Squirt Ciona. Cold Spring Harb Protoc 2009.

Delsuc, F., Brinkmann, H., Chourrout, D., Philippe, H., 2006. Tunicates and not cephalochordates are the closest living relatives of vertebrates. Nature 439, 965–968.

Deplancke, B., Alpern, D., Gardeux, V., 2016. The Genetics of Transcription Factor DNA Binding Variation. Cell 166, 538–554.

Farley, E. K., Olson, K. M., & Levine, M. S., 2015. Regulatory Principles Governing Tissue Specificity of Developmental Enhancers. Cold Spring Harbor Symposia on Quantitative Biology, 80, 27–32.

Farley, E.K., Olson, K.M., Zhang, W., Brandt, A.J., Rokhsar, D.S., Levine, M.S., 2015. Suboptimization of developmental enhancers. Science (80-.). 350, 325–328.

Farley, E.K., Olson, K.M., Zhang, W., Rokhsar, D.S., Levine, M.S., 2016. Syntax compensates for poor binding sites to encode tissue specificity of developmental enhancers. Proc. Natl. Acad. Sci. 113, 6508–6513.

Guss, K.A., 2001. Control of a Genetic Regulatory Network by a Selector Gene. Science (80-.). 292, 1164–1167.

Imai, K.S., Satou, Y., Satoh, N., 2002. Multiple functions of a Zic-like gene in the differentiation of notochord, central nervous system and muscle in Ciona savignyi embryos. Development 129, 2723–32.

José-Edwards, D.S., Oda-Ishii, I., Kugler, J.E., Passamaneck, Y.J., Katikala, L., Nibu, Y., Di Gregorio, A., 2015. Brachyury, Foxa2 and the cis-Regulatory Origins of the Notochord. PLoS Genet. 11, 1–16.

Lettice, L.A., Devenney, P., De Angelis, C., Hill, R.E., 2017. The Conserved Sonic Hedgehog Limb Enhancer Consists of Discrete Functional Elements that Regulate Precise Spatial Expression. Cell Rep. 20, 1396–1408.

Long, H.K., Prescott, S.L., Wysocka, J., 2016. Ever-Changing Landscapes: Transcriptional Enhancers in Development and Evolution. Cell 167, 1170–1187.

Passamaneck, Y.J., Di Gregorio, A., 2005. Ciona intestinalis: Chordate development made simple. Dev. Dyn. 233, 1–19.

Rhee, H.S., Pugh, B.F., 2011. Comprehensive genome-wide protein-DNA interactions detected at single-nucleotide resolution. Cell 147, 1408–1419.

Rothbacher, U., Bertrand, V., Lamy, C., Lemaire, P., 2007. A combinatorial code of maternal GATA, Ets and -catenin-TCF transcription factors specifies and patterns the early ascidian ectoderm. Development 134, 4023–4032.

Shi, W., Levine, M., 2008. Ephrin signaling establishes asymmetric cell fates in an endomesoderm lineage of the Ciona embryo. Development 135, 931–940.

Shlyueva, D., Stampfel, G., Stark, A., 2014. Transcriptional enhancers: From properties to genome-wide predictions. Nat. Rev. Genet. 15, 272–286.

Yagi, K., 2004. A zinc finger transcription factor, ZicL, is a direct activator of Brachyury in the notochord specification of Ciona intestinalis. Development 131, 1279–1288.

Yasuo, H., Hudson, C., 2007. FGF8/17/18 functions together with FGF9/16/20 during formation of the notochord in Ciona embryos. Dev. Biol. 302, 92–103.