Author Manuscript

Author Manuscript

Author Manuscript

Author Manuscript

HHS Public Access

Author manuscript

Cell Rep. Author manuscript; available in PMC 2023 October 23.

Published in final edited form as:

Cell Rep. 2023 February 28; 42(2): 112052. doi:10.1016/j.celrep.2023.112052.

Diverse logics and grammar encode notochord enhancers

Benjamin P. Song[1,2,3,5], **Michelle F. Ragsac**[1,2,4,5], **Krissie Tellez**[1,2], **Granton A. Jindal**[1,2], **Jessica L. Grudzien**[1,2], **Sophia H. Le**[1,2], **Emma K. Farley**[1,2,6,*]

[1]Department of Medicine, Health Sciences, University of California San Diego, La Jolla, CA 92093, USA

[2]Department of Molecular Biology, Biological Sciences, University of California San Diego, La Jolla, CA 92093, USA

[3]Biological Sciences Graduate Program, University of California San Diego, La Jolla, CA 92093, USA

[4]Bioinformatics and Systems Biology Graduate Program, University of California San Diego, La Jolla, CA 92093, USA

[5]These authors contributed equally

[6]Lead contact

SUMMARY

The notochord is a defining feature of all chordates. The transcription factors Zic and ETS regulate enhancer activity within the notochord. We conduct high-throughput screens of genomic elements within developing *Ciona* embryos to understand how Zic and ETS sites encode notochord activity. Our screen discovers an enhancer located near *Lama*, a gene critical for notochord development. Reversing the orientation of an ETS site within this enhancer abolishes expression, indicating that enhancer grammar is critical for notochord activity. Similarly organized clusters of Zic and ETS sites occur within mouse and human *Lama1* introns. Within a Brachyury (Bra) enhancer, FoxA and Bra, in combination with Zic and ETS binding sites, are necessary and sufficient for notochord expression. This binding site logic also occurs within other *Ciona* and vertebrate Bra enhancers. Collectively, this study uncovers the importance of grammar within notochord enhancers and discovers signatures of enhancer logic and grammar conserved across chordates.

This is an open access article under the CC BY-NC-ND license (http://creativecommons.org/licenses/by-nc-nd/4.0/).

*Correspondence: efarley@ucsd.edu.

AUTHOR CONTRIBUTIONS

E.K.F., B.P.S., M.F.R., K.T., and G.A.J. designed the experiments. B.P.S., K.T., J.L.G., and S.H.L. conducted the experiments. M.F.R. conducted bioinformatic analyses. E.K.F., B.P.S., and M.F.R. wrote the manuscript. All authors were involved in editing the manuscript.

SUPPLEMENTAL INFORMATION

Supplemental information can be found online at https://doi.org/10.1016/j.celrep.2023.112052.

DECLARATION OF INTERESTS

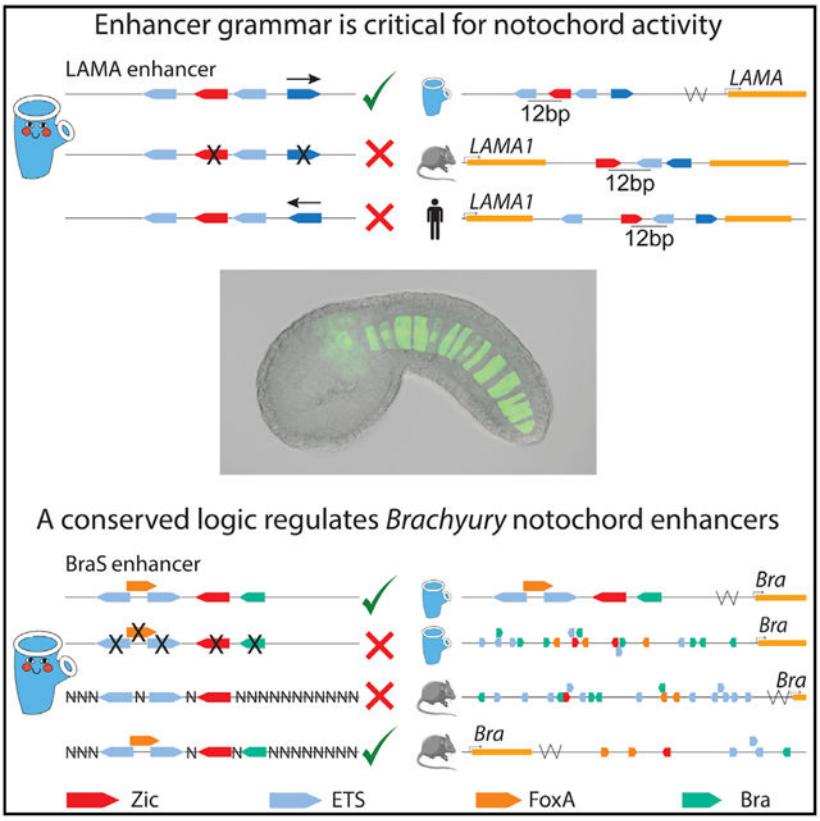
The authors declare no competing interests.

INCLUSION AND DIVERSITY

One or more of the authors in this paper self-identifies as an underrepresented ethnic minority in their field of research or within their geographical location. One or more of the authors of this paper received support from a program designed to increase minority representation in their field of research.

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

Graphical Abstract



In brief

Song et al. conduct high-throughput screens of genomic elements within developing embryos to understand how enhancers encode notochord expression. The orientation of binding sites, an aspect of enhancer grammar, is essential for notochord enhancer activity. Signatures of enhancer logic and grammar occur across chordates, suggesting a conserved enhancer grammar.

INTRODUCTION

Enhancers are genomic elements that act as switches to ensure the precise patterns of gene expression required for development.[1] Enhancers regulate the timing, locations, and levels of expression by binding of transcription factors (TFs) to sequences within the enhancer known as transcription factor binding sites (TFBSs).[2–6] This binding, along with protein-protein interactions, leads to recruitment of transcriptional machinery and activation of gene expression. While we know that TFBSs regulate enhancers and mediate tissue-specific expression, we have limited understanding of how the sequence of an enhancer encodes a particular expression pattern and what combinations of binding sites within enhancers are able to mediate enhancer activity. Given that the majority of variants associated with disease and phenotypic diversity lie within enhancers,[7–9] it is critical that we understand how the underlying enhancer sequence encodes tissue-specific expression and what types of

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

changes within an enhancer sequence can cause changes in expression, cellular identity, and phenotypes.

A set of grammatical rules that define how enhancer sequence encodes tissue-specific expression was suggested almost 30 years ago.[10–13] The hypothesis for grammatical rules is based on the physical properties of transcription factors and enhancer DNA. These physical constraints govern functional protein-DNA interactions and could be read out within the DNA sequence as constraints on the arrangement of the TFBSs within a functional enhancer. Enhancer grammar is composed of constraints on the number, type, and affinity of TFBSs within an enhancer and the syntax of these sites (orders, orientations, and spacings).[14]

We previously identified grammatical rules governing notochord enhancers regulated by Zic and ETS TFBSs.[15] We found that there was an interplay between affinity and organization of TFBSs, such that organization could compensate for poor affinity and vice versa. Using these rules, we discovered two notochord enhancers, Mnx and Brachyury Shadow (BraS). These enhancers use low-affinity ETS sites in combination with Zic sites to encode notochord expression.[15] Here, we focus on obtaining a deeper understanding of how enhancers regulated by Zic and ETS encode notochord expression.

Zic and ETS are co-expressed in the developing notochord of the marine chordate *Ciona intestinalis type A*, also known as *Ciona robusta* (*Ciona*), (Figure 1) and in vertebrates.[16,17] The notochord is a key feature of chordates and acts as a signaling center to pattern the neighboring neural tube, paraxial mesoderm, and gut.[18,19] Specification of the notochord by Brachyury (Bra), also known as T, is highly conserved across chordates.[20–23] Other conserved TFs important for activation of notochord gene expression include Zic,[16,17,24–28] ETS,[17,29–32] a TFdownstream of FGF signaling, and FoxA.[33–38]

Our study focuses on the marine chordate, *Ciona*, a member of the urochordates, the sister group to vertebrates.[39] Fertilized *Ciona* eggs can be electroporated with many enhancers in a single experiment, which allows for testing of many enhancers in whole, developing embryos.[40,41] Furthermore, these embryos are transparent and have defined cell lineages, making it easy to image and determine the location of enhancer activity. These advantages, along with the fast development of *Ciona* and the similarity of notochord development programs between *Ciona* and vertebrates,[40,42] make it an ideal organism to study the rules governing notochord enhancers during development.

Within the *Ciona* genome, we found 1,092 elements containing one Zic site and at least two ETS sites within 30 bp upstream or downstream of the Zic site. We tested 90 of these for expression in developing *Ciona* embryos. Only 10% of these regions drive notochord expression. These notochord enhancers fall into three categories: enhancers containing Zic and ETS sites, ones with Zic, ETS, and Bra sites, and ones with Zic, ETS, FoxA, and Bra sites. Within enhancers containing Zic and ETS sites, the organization of sites is important for activity, indicating that grammatical constraints on Zic and ETS encode enhancer activity. We find that one of the Zic and ETS enhancers is near an important notochord gene, *laminin alpha*.[43] The orientation of binding sites within this *laminin alpha* enhancer is critical for enhancer activity demonstrating the role of enhancer grammar. We

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

find similar clusters of Zic and ETS sites within the introns of *laminin alpha-1* in both mouse and human. Strikingly, we find the same 12 bp spacing between the Zic and ETS conserved across all three species. In addition, this study identifies two enhancers using a combination of Zic, ETS, FoxA, and Bra to encode notochord expression. One of these is the BraS enhancer. By creating a library of 45 million enhancer variants with the sequence, affinity, and position of the Zic, ETS, FoxA, and Bra sites fixed while all other nucleotides are randomized, we discover that these sites are necessary and sufficient for notochord expression. Other known *Bra* enhancers within *Ciona*[44] and vertebrates[45] also harbor this combination of TFs, suggesting that Zic, ETS, FoxA, and Bra is a common feature of *Bra* regulation in chordates. Collectively, our study finds that grammar is a key component of functional enhancers with signatures of this enhancer logic and grammar seen across chordates.

RESULTS

Searching for clusters of Zic and ETS sites within the *Ciona* genome

To better understand how Zic and ETS sites within enhancers encode notochord expression, we searched the *Ciona* genome (KH2012) for clusters of Zic and ETS sites. To do this, we first identified Zic motifs in the genome. We defined Zic motifs using EMSA and enhancer mutagenesis data from previous studies (see STAR Methods for motifs).[17,28,46] Using the Zic site as an anchor, we searched the 30 bp upstream and downstream of the Zic site for ETS sites, using the core motif GGAW (GGAA and GGAT) to consider all ETS sites regardless of affinity,[47,48] as we have previously found that low-affinity ETS sites are required to encode notochord-specific expression.[15] This search identified 1,092 genomic regions approximately 68 bp in length. We define these regions as ZEE elements.

Testing ZEE genomic elements for enhancer activity in developing *Ciona* embryos

We selected 90 ZEE elements (Figure S1A; Table S1) and synthesized these upstream of a minimal promoter (bpFog[49,50]) and a transcribable barcode to conduct an enhancer screen (experiment outlined in Figure 2A). Each enhancer was associated with, on average, six unique barcodes. Each different barcode is a distinct measurement of enhancer activity. We electroporated this library into fertilized *Ciona* eggs. We collected embryos at the late gastrula stage (5.5 h post-fertilization [hpf]) when notochord cells are developing[51] and both Zic and ETS are expressed.[52,53] At this time point, we isolated mRNA and DNA. To determine that all the enhancer plasmids got into the embryos, we isolated the plasmids from the embryos and sequenced the DNA barcodes. We detected barcodes associated with all 90 ZEE elements from the isolated plasmids, indicating that all elements were tested for activity within the developing *Ciona* embryos.

We next wanted to see how many of the 90 ZEE elements act as enhancers to drive transcription. Active enhancers will transcribe the GFP and the barcode into mRNA. To find the functional enhancers, we isolated the mRNA barcodes from our electroporated embryos and sequenced them. We analyzed the sequencing data and measured the reads per million (RPM) for each barcode. To calculate an average RNA RPM for a given enhancer, we averaged the RPM for each RNA barcode associated with an enhancer. To normalize the

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

enhancer activity to the differences in the amount of plasmid and therefore number of copies of the enhancer electroporated into embryos, we took the log2 of the average enhancer RNA RPM divided by the DNA RPM for the same enhancer to create an enhancer activity score. Enhancer activity scores below zero are non-functional, while elements with scores above zero are considered functional enhancers. The highest activity score is around four. The experiment was repeated in biological triplicate and there was a high correlation between all three biological replicates (Figures S1B and S1C).

Many genomic ZEE elements are not enhancers

As an internal, positive control in our enhancer screen, we included the BraS enhancer. This enhancer drives expression in the notochord and weak expression in the a6.5 lineage, both locations that express Zic and ETS.[15] The BraS enhancer activity score is 2.4 (Figure 2B), indicating that our library screen is detecting functional enhancers. Thirty-nine of the ZEE elements act as enhancers in our screen, while 51 of the ZEE elements drove no expression. This suggests that genomic elements containing a single Zic site and at least two ETS sites are not sufficient to drive expression in the notochord. To further validate our sequencing data and to determine the tissue-specific location of the functional enhancers, we selected 20 non-functional elements and 24 functional enhancers from our screen to test by an orthogonal approach. Each of these ZEE elements were cloned upstream of a minimal bpFog promoter and GFP. We electroporated each enhancer into fertilized eggs and analyzed the GFP expression of these ZEE elements under the microscope at 8 hpf in at least 150 embryos across three biological replicates. Collectively, we analyzed expression of these elements in over 6,600 embryos with this orthogonal approach.

All 20 ZEE elements defined as non-functional in our library drove no GFP expression, validating our enhancer activity score cut off that we defined for non-functional enhancers (Figure 2C). In the 24 enhancers detected as functional within the enhancer screen, 92% of these enhancers (22/24) showed GFP expression within the embryos when tested individually (Table S2). Nine ZEE elements drove expression in the notochord (Figure S2; Table S3). Four of these enhancers are active almost exclusively in the notochord (ZEE10, 13, 20, 27). The remaining five are active in the notochord with additional expression in the endoderm and/or nerve cord (b6.5 lineage). Twelve of the ZEE enhancers drove varying levels of expression in the a6.5 lineage, which gives rise to the neural cell types called the anterior sensory vesicle and the palps, but only one drove expression exclusively in this cell type (ZEE22). Thirteen ZEE elements drove expression in one or more for the following cell types: the nerve cord (b6.5 lineage), mesenchyme, and endoderm. The expression patterns seen for these active enhancers are consistent with the expression patterns of Zic and ETS, which are expressed in the muscle, endoderm, ectoderm, mesenchyme, notochord, a6.5 neural lineage, and b6.5 neural cell types.[54–58] The only cells to co-express both Zic and ETS are the notochord, a6.5, and a small number of mesenchyme cells (Figure 1). Therefore, enhancers under combinatorial control of Zic and ETS are likely to be active in the notochord and the a6.5 neural lineage.[17,58,59] Collectively these results indicate that our enhancer screen accurately detects functional enhancers, and our tissue-specific analysis provides detailed expression patterns for these enhancers.

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

Elucidating the logic of the enhancers driving notochord expression

Having seen that so few enhancers drive expression in the notochord, we were interested to better understand why these nine functional enhancers were active in the notochord. It is possible that they are functional due to the grammar of the Zic and ETS sites or because other TFBSs are required for notochord expression. To investigate these two hypotheses, we looked at the nine notochord enhancers in more detail. FoxA and Bra are two other TFs important for activation of notochord enhancers in chordates.[22,33–37,60] We therefore searched all 90 ZEE elements for FoxA and Bra sites. We used EMSA and crystal structure data to define TRTTTAY as the FoxA motif[36,37,61] and TNNCAC as the Bra motif.[60,62–65]

The nine elements that drive notochord expression contain three different combinations of TFs

Of the 90 genomic regions we tested, 42 had only Zic and ETS sites, 39 had Zic, ETS, and Bra sites, 4 had Zic, ETS, FoxA, and Bra sites, and 5 had Zic, ETS, and FoxA sites. Ten percent of the enhancers containing only Zic and ETS sites drive notochord expression (4/42). Eight percent (3/39) of the enhancers containing Zic, ETS, and Bra drive notochord expression. None of the enhancers (0/5) containing Zic, ETS, and FoxA drive notochord expression, while 50% (2/4) of the enhancers containing Zic, ETS, FoxA, and Bra are active in the notochord (Figures 3 and S3). Thus, there are three groups of notochord enhancers that contain: (1) Zic and ETS sites alone, (2) Zic, ETS, and Bra sites, or (3) Zic, ETS, FoxA, and Bra sites. Having found that only a few of the elements containing Zic and ETS sites alone were functional, we wanted to understand if the organization or grammar of sites within these enhancers was important.

Zic and ETS enhancer grammar encodes notochord *laminin alpha* expression

Four enhancers containing Zic and ETS sites only (ZEE13, 20, 27, and 85) drive notochord expression. ZEE13, 20, and 27 drive expression only in the notochord and have similar levels of expression. ZEE85 drives expression predominantly in the nerve cord (b6.5 lineage) with weak notochord expression. ZEE20, 27, and 85 are not in close proximity to known notochord genes, although it is possible that these elements regulate notochord genes further away. The ZEE13 enhancer is located close to *laminin alpha*, which is critical for notochord development[43] (Figure 4A). Given the proximity of this notochord-specific enhancer to *laminin alpha*, we decided to focus further analysis on this enhancer, which we renamed the Lama enhancer. Notably, this enhancer contains three ETS sites. To determine the affinity of these sites, we used protein binding microarray (PBM) data for mouse ETS-1,[48] as the binding specificity of ETS is highly conserved across bilaterians.[48,66] The consensus highest-affinity site has a score of 1.0, and all other 8-mer sequences have a score relative to the consensus. The Lama enhancer contains two ETS sites with exceptionally low affinities of 0.10, or 10% of the maximal binding affinity, while the most distal ETS site is a high-affinity site (0.73).

To determine if the Zic site and ETS sites are important for enhancer activity, we made a point mutation to ablate the ETS3 site, which we chose because it has the highest affinity (Figures 4B and S4A; Table S4). This led to a complete loss of notochord activity, indicating that this ETS site contributes to enhancer activity. Similarly, ablation of the Zic site results

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

in complete loss of enhancer activity, indicating that both Zic and ETS sites are necessary for activity of this Lama enhancer (Figures 4C and S4A; Table S4). We did not ablate the low-affinity ETS sites of the Lama enhancer. Previously, we saw that the organization of sites within enhancers, a component of enhancer grammar, is critical for enhancer activity in both the Mnx and Bra enhancer. To see if enhancer grammar is important for activity within the Lama enhancer, we altered the orientation of sites within this enhancer and measured the impact on enhancer activity. Reversing the orientation of the first ETS site, which has an affinity of 0.10, led to a dramatic reduction in notochord expression, suggesting that the orientation of this ETS site is important for enhancer activity. Similarly, reversing the orientation of the third ETS site (Lama RE3), which has an affinity of 0.73, also causes a loss of notochord expression (Figures 4D and S4A; Table S4). These two manipulations demonstrate that the orientation of these ETS sites within this enhancer is important for activity, and, thus, that there are some grammatical constraints on the *Ciona* Lama enhancer. It is likely that grammar is an important feature of enhancers regulated by Zic and ETS, as we have previously seen similar grammatical constraints on the orientation and spacing of binding sites within the Mnx and BraS enhancer, and because so few of the genomic ZEE elements containing these sites are functional.[15]

Vertebrate *laminin alpha-1* introns contain clusters of Zic and ETS with conserved spacing

The expression of laminin in the notochord is highly conserved between urochordates and vertebrates.[43,67,68] Indeed, laminins play a vital role in both urochordate and vertebrate notochord development, with mutations in laminins or components that interact with laminins causing notochord defects.[69–71] The *Ciona laminin alpha* is the ortholog of the vertebrate *laminin alpha 1/3/5* family. We therefore sought to determine if we could find a similar combination of Zic and ETS sites in proximity to vertebrate *laminin* genes, as both Zic[16,27] and ETS[72,73] are important in vertebrate notochord development. Strikingly, we find a cluster of Zic and ETS sites within the intron of both the mouse and human *laminin alpha-1* genes. The affinity of the ETS sites in all three species is also far from the consensus: the human cluster contains three ETS sites of 0.12, 0.17, and 0.25 affinity, while the putative mouse enhancer contains fewer, but higher-affinity, ETS sites (Figure 4E). We have previously seen that the spacing between Zic and adjacent ETS sites affects levels of expression, with spacings of 11 and 13 bp seen between ETS and Zic sites in the BraS enhancer and Mnx enhancer, respectively.[15] In line with this observation, the *laminin alpha-1* clusters in mouse and human and the *Ciona* Lama enhancer have a 12 bp spacing between the ETS and adjacent Zic site in all three species, suggesting that such spacings (11–13 bp) are a feature of some notochord enhancers regulated by Zic and ETS. The conservation of this combination of sites, the low-affinity ETS sites, and the conserved spacing hints at the conservation of enhancer grammar across chordates.

The Zic, ETS, FoxA, and Bra regulatory logic encodes notochord enhancer activity

The group of genomic elements most enriched in notochord expression was the group containing Zic, ETS, FoxA, and Bra binding sites, with two of the four driving notochord expression. Both of these enhancers are located near genes expressed in the notochord.[67] The first was our positive control BraS, while the second enhancer is in proximity of the

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

Lrig gene. Both of these enhancers drive strong notochord expression along with some neural a6.5 expression.

We previously identified the BraS enhancer through a search for rules governing Zic and ETS grammar that included number and type of TFBSs, along with the affinity, spacing, and orientation of TFBSs.[15] The BraS enhancer contains a Zic and two low-affinity ETS sites (0.14 and 0.25). We previously saw that changing the orientation of the lowest affinity ETS site, located 11 bp from the Zic site, leads to loss of expression, indicating that there are grammatical constraints on this enhancer and that the 0.14 affinity ETS site is important for expression.[15] To further confirm the role of the Zic and two ETS sites within BraS, we ablated these three sites (Zic and both ETS sites) with point mutations; this leads to complete loss of expression, demonstrating that these sites are necessary for notochord expression (Figures 5B and S5B; Table S4). To test if these sites are sufficient for notochord expression, we created a library of 24.5 million variants in which the Zic and two ETS sites were kept constant in sequence, affinity, and position while all other nucleotides were randomized. We electroporated this library into embryos and counted GFP expression in 8 hpf embryos. BraS has notochord expression in 73% of embryos, while the ZEE-randomized BraS enhancer (BraS rZE) has notochord expression in only 28% of embryos. Thus, BraS rZE drives expression within the notochord in significantly fewer embryos than BraS, indicating that there are other sites within the enhancer that are also important for tissue-specific expression (Figures 5C and S5B; Table S4). This experiment highlights the importance of understanding sufficiency in addition to necessity of sites.

Two obvious candidates for additional functional sites within BraS are the FoxA and Bra sites, which we detected in this enhancer. Both FoxA and Bra are TFs known to regulate notochord enhancers in urochordates and vertebrates.[26,35,37,59,74,75] To test if the Bra and FoxA sites contribute to expression, we ablated these sites. Ablating the Bra site within BraS leads to a significant reduction in expression, as does ablating the FoxA site (Figures 5D, 5E, and S4B; Table S4). These manipulations suggest that all five sites (Zic, FoxA, Bra, and two ETS sites) are necessary for enhancer activity, and that all four TFs contribute to the activity of BraS.

To test if the Zic, two ETS, FoxA and Bra sites are sufficient for notochord expression, we created another BraS randomization library with 45 million variants in which the Zic, ETS, FoxA, and Bra sites were fixed in sequence, position, and affinity, and all other nucleotides within the enhancer were randomized. When we electroporated this library into *Ciona*, the number of embryos showing notochord expression between the BraS Zic, ETS, FoxA, and Bra-randomized library (BraS rZEFB) and BraS WT was not significantly different (73% BraS versus 62% BraS rZEFB) (Figures 5F and S5B; Table S4), suggesting that these five sites together are sufficient to drive notochord expression in the BraS enhancer. While there is no significant difference in the number of embryos with notochord expression between the BraS rZEFB and BraS enhancers, we noticed that expression in the notochord was slightly weaker for BraS rZEFB ($p = 0.03$) (Figure S4C), suggesting that other elements within the randomized region may further augment the levels of notochord expression. We also noted that significantly fewer embryos drive expression in the a6.5 lineage in the BraS rZEFB relative to the BraS enhancer (14% versus 32% of embryos, respectively, $p < 0.01$) (Figure

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

S4D), suggesting that sequences within the randomized region are important for the neural a6.5 expression. Studies of enhancers often stop when mutation experiments demonstrate that a TF is necessary for enhancer activity. However, this falls short of a full understanding of enhancers. Our results highlight that finding necessary sites is not enough to identify the regulatory logic of an enhancer. These necessity and sufficiency experiments have uncovered a deeper understanding of the BraS enhancer, namely that it is regulated by Zic, ETS, FoxA, and Bra.

Zic, ETS, FoxA, and Bra may be a common regulatory logic for *Ciona* Brachyury enhancers

The first and most well-studied *Bra* enhancer is the Bra434 enhancer,[44,76] which drives strong expression in the notochord (Figure S5A). Bra434 enhancer contains Zic, ETS, FoxA, and Bra sites; ablating these sites within this enhancer leads to reduced expression, suggesting that these sites contribute to enhancer activity.[75,77] There are different reports regarding the number and location of ZEFB sites within the Bra434 enhancer depending on the method used to define sites.[44,77] Here, we annotate the Bra434 enhancer using crystal structure data, enhancer mutagenesis data, and EMSA and PBM data.[17,28,36,37,46–48,60–65]

Our approach identifies two Zic sites, six low-affinity ETS sites, three FoxA sites, and eight Bra sites (Figures 5G and S5B). Of these TFs, the least information is available regarding Zic; thus, it is possible that there are other more degenerate Zic sites that may be identified in future studies.[44,75–77] Bra434 has stronger expression in the notochord than BraS and this may be due to the longer length of the Bra434 enhancer and the presence of more Zic, ETS, FoxA, and Bra sites within Bra434 relative to BraS enhancer. Having seen that clusters of Zic, ETS, FoxA, and Bra are important in the BraS and Bra434 enhancers, we next wanted to see if this logic is found in *Bra* enhancers in vertebrates.

Vertebrate notochord enhancers contain clusters of Zic, ETS, FoxA, and Bra, suggesting that this is a common logic for regulation of *Brachyury* expression in the notochord

In mouse, the most well-defined notochord enhancer to date is within an intron of *T2*, 38 kb upstream of *T*, which is the mouse ortholog of *Bra*[45] (Figure 5H). This mouse *T* enhancer is required for *Bra/T* expression, notochord cell specification, and differentiation.[45] Homozygous deletion of this *Bra/T* enhancer in mouse leads to reduction of *Bra/T* expression, a reduction in the number of notochord cells, and halving of tail length. Bra/T and FoxA binding sites have previously been identified within this enhancer.[45] We find that this mouse *Bra/T* enhancer also contains Zic and ETS binding sites. Within this enhancer there are 12 ETS sites; 11 of these have affinities ranging from 0.09–0.14, while 1 site has an affinity of 0.65, indicating that this enhancer contains low-affinity ETS sites.

As we saw with the *Ciona* BraS and Bra434 enhancer, typically there are multiple enhancers that all regulate the same or similar patterns of expression.[78–80] This is thought to confer the transcriptional robustness required for successful development.[78,80–82] Following this reasoning, we continued to search the mouse *Bra/T* region to see if we could find other putative notochord enhancers that may regulate *Bra/T*. We identified a region located 2 kb downstream of *T* that contains a cluster of Zic, low-affinity ETS (0.11–0.12), FoxA, and Bra sites (Figure 5I). This putative enhancer occurs within an open chromatin region in mouse

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

E8.25 notochord cells,[83] suggesting that this may be another mouse *T* enhancer. Similarly in zebrafish, a notochord enhancer located 2.1 kb upstream of the *Bra* ortholog *ntl*[84] also contains a cluster of Zic, ETS, FoxA, and Bra sites (Table S6). The presence of these four TFs in *Ciona*, zebrafish, and mouse *Bra* enhancers suggests that the use of Zic, ETS, FoxA, and Bra could be a common enhancer logic regulating expression of the key notochord specification gene *Bra* in chordates.

DISCUSSION

In this study we sought to understand the regulatory logic of notochord enhancers by taking advantage of high-throughput studies within the marine chordate *Ciona*. Within the *Ciona* genome, there are 1,092 genomic regions containing a Zic site within 30 bp of 2 ETS sites. We tested 90 of these ZEE genomic regions for expression in developing *Ciona* embryos. Surprisingly, only nine of the regions drove notochord expression. Among these nine, we identified a *laminin alpha* enhancer that was highly dependent on grammatical constraints for proper expression. We found a similar cluster of Zic and ETS sites within the intron of the mouse and human *laminin alpha-1* gene; strikingly, these clusters and the *Ciona* laminin enhancer have the same spacing between the Zic and ETS sites. Within the BraS enhancer, although Zic and ETS are necessary for enhancer activity, randomization of the BraS enhancer keeping only the Zic and ETS sites constant in a sea of 24.5 million variants reveals that these sites are not sufficient for notochord activity. FoxA and Bra sites are also necessary for notochord expression. Indeed, creating a library of 45 million BraS variants in which all five TFBSs are kept constant in position and affinity, while all other nucleotides are randomized, leads to notochord expression in a similar proportion of embryos as the WT BraS, which indicates that these sites are sufficient for notochord expression. We find that the combination of Zic, ETS, FoxA, and Bra occurs within other *Bra* enhancers in *Ciona* and vertebrates suggesting that this combination of TFs may be a common logic regulating *Bra* expression. Our study discovers developmental enhancers, demonstrates the importance of enhancer grammar within developmental enhancers, and provides a deeper understanding of the regulatory logic governing *Bra*. Our findings of the same clusters of sites within vertebrates hint at the conserved role of grammar and logic across chordates.

Very few genomic regions containing Zic and two ETS sites are functional enhancers

Our analysis of 90 genomic elements all containing at least one Zic site in combination with two ETS sites strikingly demonstrate that clusters of sites are not sufficient to drive expression. Only 39 of the 90 (43%) elements tested drove any expression and, even more surprisingly, only 15 of these drove expression in lineages that co-express Zic and ETS, namely the a6.5 (anterior sensory vesicle and palps) and/or notochord. These findings indicate that searching for clusters of TFs is only minimally effective in identification of enhancers and suggests that the organization of sites is also important for rendering a cluster of binding sites a functional enhancer. Our findings are in agreement with the work from King et al.,[85] that found only 28% of the genomic elements they tested for enhancer function in ESCs drove enhancer activity, despite the fact that these genomic elements contain TF motifs and bound these TFs in ChIP-seq assays. Our study and that of King et al.[85] suggest

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

that having motifs, or even TF binding, is not sufficient to drive expression and suggests that the grammar of these sites is critical for rendering a cluster of TFBSs a functional enhancer.

Grammar is a key constraint of the Lama and BraS enhancers

Zic and ETS are necessary for activity of the Lama enhancer. Within the Lama enhancer, the orientation of binding sites relative to each other was critical for expression, providing evidence that enhancer grammar is a critical feature of functional enhancers regulated by Zic and ETS. Flipping the orientation of either the first or last ETS sites relative to the Zic site led to loss of enhancer activity in the *Ciona* Lama enhancer. This mirrors the results of flipping the orientation of the ETS sites within the BraS enhancer.[15] *Laminin alpha* is a key gene involved in notochord development in both *Ciona* and vertebrates.[43,71] Intriguingly, we find that both the human and mouse *laminin alpha-1* have introns that harbor a similar cluster of Zic and ETS sites to those seen within *Ciona*. There is a conservation of 12 bp spacing between the Zic and ETS sites across all three chordate enhancers, similar to the spacing we have observed between Zic and ETS sites within the notochord enhancers Mnx and BraS.[15] We note that the vertebrate regions do not drive notochord expression in *Ciona*. It is possible that grammar is subtly tweaked between different species. Alternatively, the lack of activity could be due to promoter incompatibility across species, as in our assay we tested the mouse and human Lama enhancers with a Ciona promoter. Reporter assays within mouse embryos could further investigate the functionality of the mouse and human Lama putative enhancers and the role of the 12 bp spacing within these elements.

Necessity of sites does not mean sufficiency—A deeper understanding of the BraS enhancer

Our study of the BraS enhancer highlights the importance of testing sufficiency of sites to investigate if we fully understand the regulatory logic of an enhancer. We previously demonstrated that reversing the orientation of an ETS site led to loss of notochord expression in the BraS enhancer. Here, in this study, we show via point mutations that both Zic and ETS sites are required for enhancer activity. However, randomization of the BraS enhancer to create 24.5 million variants in which only the Zic and ETS sites are constant demonstrates that these sites are not sufficient for enhancer activity, as the randomized BraS enhancer (BraS rZE) only drives notochord expression in less than half the number of embryos as the BraS enhancer. Having discovered that Zic and ETS alone were not sufficient, we find that both FoxA and Bra sites also contribute to the enhancer activity. In a library of 45 million variants in which the Zic, ETS, Bra, and FoxA sites are kept constant in sequence, affinity, and position within a randomized backbone (BraS rZEFB), we see no significant difference in the number of embryos with notochord expression. This indicates that these five sites are necessary and sufficient for enhancer activity. However, the neural expression seen with the BraS enhancer appears to depend on some features within the randomized backbone, as the rZEFB library drives significantly less neural expression. We also note that the BraS rZEFB enhancer drives slightly weaker levels of notochord expression. These findings illustrate that enhancers are densely encoded with many features that contribute to expression. This is in line with recent work suggesting that enhancers contain far more regulatory information than previously appreciated.[86] It is possible that degenerate Zic, ETS, FoxA, or Bra sites could be present or that other TFBS are also

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

contributing to this logic. Further analysis conducting MPRAs with these two libraries (BraS rZE and BraS rZEFB) will determine what other features are contributing to notochord and neural expression. Sufficiency experiments are rarely done, and we are unaware of another study that has tested sufficiency across the entirety of an enhancer in developing embryos. Our experiments demonstrate the importance of testing sufficiency to determine all the features contributing to enhancer function and illustrate the dense encoding of regulatory information within enhancers.

Partial grammatical rules can provide signatures that identify enhancers, but improved understanding could lead to more accurate predictions

We were able to find the BraS enhancer using grammatical constraints on organization and spacing between Zic and ETS sites and affinity of ETS sites.[15] Interestingly, we did not have all the features required for enhancer activity. As such, this suggests that partial knowledge of grammatical constraints, or partial signatures of grammar, could be used to identify functional enhancers. Our previous strategy searched for these grammatical constraints in proximity of known notochord genes, which may be why we were successful in identification of the Mnx and BraS enhancer with only partial grammar rules. Current genomic screens that use TFBSs and biochemical markers, such as histone modifications and co-factor binding, have varied success in identifying functional regulatory elements.[85] Understanding the dependency between all features within an enhancer will likely enable greater success in identification of functional enhancers. Until then, our current knowledge of grammatical constraints may still be useful for pointing us toward putative enhancers.

Zic, ETS, FoxA, and Bra may be a common logic upstream of *Brachyury* in chordates

The Bra434 enhancer also contains the same combination of sites as the BraS enhancer; therefore, it is possible that this is a common logic for regulating *Bra*. Interestingly, we find these sites within mouse and zebrafish Bra enhancers.[45,84] While there are differences in expression dynamics of these factors in vertebrates and ascidians, it is striking to see this combination of sites in validated notochord enhancers across these species. Indeed, our study in both the *laminin* enhancers and *Bra* enhancers provides hints of a conserved regulatory logic across chordates, although future tests of these putative enhancers within mouse are required to see if these are truly conserved enhancers with similar grammar signatures. Our study focuses on conservation of grammatical signatures rather than sequence conservation. A recent study searching for conserved enhancers in syntenic regions suggests that there may be much more conservation of enhancer function than expected based on sequence conservation.[87] Our approach searching for grammatical signatures rather than sequence conservation may allow for identification of such functionally conserved enhancers.

Approaches to understanding dependency grammar of notochord expression

Searching for grammatical rules governing enhancers requires comparison of functional enhancers with the same features. Although we thought we had the same features in all 90 regions, we actually had at least three distinct types of enhancers within our screen. This illustrates a common problem in mining genomic data for patterns, as the assumption that we are comparing like with like is often an incorrect one. Other screens mining

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

genomic elements have hit similar roadblocks, with only a few functional genomic examples being uncovered and thus limiting the ability to find grammatical rules.[85] To uncover the grammatical constraints on enhancers, we need to not only understand the number and types of sites within an enhancer, but also the dependency between these sites, such as affinity, spacing, and orientation.[14]

Massively or gigantic parallel reporter assays with increased size and complexity and that combine both synthetic enhancers and genomic elements will likely be required to pinpoint the rules governing enhancer activity within genomes. However, integrating synthetic screens with genomic screens is a major challenge as synthetic screens often have limited application within the context of the genome.[85] Another approach is to study entirely random sequences for enhancer activity, which has been done in the context of promoters in bacteria and yeast.[88,89] Indeed, the conclusions of these studies mirror our own findings that grammar and low-affinity sites are critical components of functional regulatory elements. However, as 83% of the random sequences within yeast drove expression, it is unclear how well random sequences mirror the regulatory landscape within the genome that has been shaped by evolutionary constraints over millions of years. Nonetheless, testing random sequences within the context of developing embryos could provide another source of data to understand how enhancers encode tissue-specific expression.[90] In the future, integration of genomic regions, synthetic designed, and random sequences will contribute to our understanding of enhancer grammar. Despite the complexity of studying enhancers in developing embryos, our study demonstrates that enhancer grammar is critical for encoding notochord activity and our observation of the same logics and grammar signatures in both *Ciona* and vertebrates hints at conservation of these grammatical constraints across chordates.

Limitations of the study

In this study, we screened 90 ZEE elements for functionality; however, only 10% were active in the notochord. We anticipate that discovering more notochord enhancers regulated by Zic or ETS, or regulated by Zic, ETS, FoxA, and Bra, could better inform our understanding of notochord grammar. Toward this end, testing all 1,092 ZEE elements we identified within the *Ciona* genome could strengthen this study. However, this would likely only yield 100 notochord enhancers, which would still not be enough to define grammatical rules. As discussed above, combining assays of genomic regions with synthetic and random enhancer screens could help gain enough data to determine the grammar of notochord enhancers.

Another limitation relates to our identification of conserved enhancer logic and grammar across chordates. While we identified similar signatures with the Lama enhancers in *Ciona*, mouse and humans, we did not test the mouse Lama enhancer for activity in mouse, nor did we functionally interrogate the importance of the 12 bp spacing within this enhancer in the context of *Ciona* or mouse. Conducting these studies would deepen our understanding of the conservation of grammar across chordates. We also identified a common logic of Zic, ETS, FoxA, and Bra sites within Bra enhancers. While we know that deletion of the mouse *Bra* TNE enhancer does lead to loss of notochord in mouse, it would strengthen the study to

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

manipulate the Zic, ETS, FoxA, and Bra sites within the context of the mouse and zebrafish Bra/T enhancers to determine if the conservation of this logic is important for regulation of *Bra*.

STAR★METHODS

RESOURCE AVAILABILITY

Lead contact—Further information and requests for resources and reagents should be directed to and will be fulfilled by the lead contact, Emma Farley (efarley@ucsd.edu).

Materials availability—Plasmids generated in this study are available upon request.

Data and code availability

- All ZEE screen sequencing data can be found on the SRA database. Accession number PRJNA861319: https://www.ncbi.nlm.nih.gov/sra/PRJNA861319. Microscopy and scoring data reported in this paper will be shared by the lead contact upon request.
- All original code can be found at: https://github.com/farleylab/Diverse-Logics-Notochord-Study
- Any additional information required to reanalyze the data reported in this paper is available from the lead contact upon request.

EXPERIMENTAL MODEL AND SUBJECT DETAILS

Tunicates—Adult *C. intestinalis* type A aka *Ciona robusta* (obtained from M-Rep) were maintained under constant illumination in seawater (obtained from Reliant Aquariums) at 18°C. *Ciona* are hermaphroditic, therefore there is only one possible sex for individuals. Age or developmental stage of the embryos studied are indicated in the main text.

METHOD DETAILS

Library construction—The genomic regions were ordered from Agilent Technologies with adapters containing BseRI sites. This was cloned into the customdesigned SEL-Seq (Synthetic Enhancer Library-Sequencing) vector using type II restriction enzyme BseRI. After cloning, the library was transformed into bacteria (MegaX DHB10 electrocompetent cells), and the culture was grown up until an OD of 1 was reached. DNA was extracted using the Macherey-Nagel Nucleobond Xtra Midi kit. A 30bp barcode with adapters containing Esp3I sites was cloned into this library using type II restriction enzyme Esp3I. The library was transformed into bacteria (MegaX DHB10 electrocompetent cells) and grown up until an OD of 2 was reached. The DNA library was extracted from the bacteria using the Macherey-Nagel Nucleobond Xtra Midi kit.

Electroporation—*Ciona* eggs were dissected from the egg duct and dechorionated in 1% sodium thioglycolate, 0.05% Pronase E, and 0.042N NaOH. Dechorionated eggs were washed in seawater with 0.1mg glycine twice and then washed in seawater two more times. Sperm was dissected from the sperm duct and diluted in seawater 1:1000. 168μL of sperm

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

was dispersed over the dechorionated eggs and allowed to fertilize the eggs for 4 min. The fertilized embryos were washed twice with seawater and electroporated with DNA using the Gene Pulser Xcell electroporator (Bio-Rad) with the following settings: 50V, 1000μF, Infinite resistance, and 4mm cuvette length.

GFP reporter assays—70 μg DNA was resuspended in 100 μL water and added to 400 μL of 0.96 M D-mannitol. Typically for each electroporation, eggs and sperm were collected from 10 adults. Embryos were fixed at the appropriate developmental stage for 15 min in 3.7% formaldehyde. The tissue was then cleared in a series of washes of 0.3% Triton X- in PBS and then of 0.01% Triton X- in PBS. Samples were mounted in Prolong Gold. GFP images were obtained with an Olympus FV3000, using the 40X objective. All constructs were electroporated in three biological replicates.

ZEE MPRA screen—50 μg of the ZEE library was electroporated into ~5000 fertilized eggs. Embryos developed until 5hrs 30 min at 22°C. Embryos put into TriZol, and RNA was extracted following the manufacturer's 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 used in an RT reaction that was specifically selected for the barcoded mRNA (Transcriptor High Fidelity, Roche). The RT product was PCR amplified and size selected using Agencourt AMPure beads (Beckman Coulter), then checked for quality and size on the 2100 Bioanalyzer (Agilent) and sent for sequencing on the NovaSeq S4 PE100 mode (Illumina). Three biological replicates were sent for sequencing.

The DNA was extracted by mixing the phenol-chloroform and interphase of TriZol extraction with 500uL of Back Extraction Buffer (4M guanidine thiocyanate, 50mM sodium citrate, and 1M Tris-base). DNA was treated with RnaseA (Thermo Fisher). DNA was cleaned up with phenol:chloroform:isoamyl alcohol (25:24:1) (Life Technologies). The DNA was PCR amplified and size selected using Agencourt AMPure beads (Beckman Coulter), then checked for quality and size on the 2100 Bioanalyzer (Agilent) and sent for sequencing on the NovaSeq S4 PE100 mode (Illumina). Three biological replicates were sent for sequencing.

Counting embryos—For each experiment, once embryos had been mounted on slides, slide labels were covered with thick tape and randomly numbered by a laboratory member not involved in this project. Expression of GFP within embryos on each slide was counted blind. In each experiment, all comparative constructs were present, along with a slide with BraS as a reference. The X-Cite was turned on for 1hr before analysis to ensure the illumination intensity was constant. To determine levels of expression, high expression was set as visible with less than 25% power on X-Cite illuminator. Fifty embryos were counted for each biological replicate.

Acquisition of images—For enhancers being compared, images were taken from electroporations performed on the same day using identical settings. For representative images, embryos were chosen that represented the average from counting data. All images

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

are subsequently cropped to an appropriate size. In each figure, the same exposure time for each image is shown to allow direct comparison.

Identification of putative notochord enhancers—We developed a script that allows for the input of any organism's genome in the fasta file format. The script first looks for an exact match of one of seven canonical Zic family binding sites and their reverse complements. We used the following sites in our search: CAGCTGTG (Zic1/2/3), CCGCAGT (Zic7/3/1), CCGCAGTC (Zic6), CCCGCTGTG (Zic1), CCAGCTGTG (Zic3), CCGCTGTG (Zic2/ZicC), and CCCGCAGTC (Zic5) as these have been identified as functional in previous studies.[17,28] Next, we drew a window of 30 bp from either end of the canonical Zic family binding site and determine if there are at least two Ets binding site cores (i.e., either GGAA or GGAT and their respective reverse complement sequences) present within the window. The location of all regions containing at least a single Zic family binding site and two Ets binding sites are saved as part of the genome search.

Scoring relative affinities of binding sites—We calculated the relative ETS binding affinity using the median signal intensity of the universal protein binding microarray (PBM) data for mouse Ets-1 proteins from the UniProbe database (http://thebrain.bwh.harvard.edu/uniprobe/index.php).[101] Previous studies have shown that the specificity of ETS family members is highly conserved even from flies to humans,[48,66] and thus ETS-1 is a good proxy for binding affinity in *Ciona* ETS-1 which has a conserved DNA binding domain.[41] The relative affinity score represents the fractional binding of median signal intensities of the native 8-mer motifs compared to the optimal 8-mer motifs for optimal Ets, which we defined as the CCGGAAGT motif and its corresponding reverse complement.

Enhancer to barcode assignment & dictionary analysis—We constructed a dictionary of unique barcode tag-enhancer pairs by not allowing for any mismatches in the ~68 bp enhancers in our library and by not allowing barcode tag-enhancer pairs to have a read count of fewer than 150 reads. Additionally, we required all barcode tags to be 29 bp or 30 bp in length. If more than one barcode tag was associated with a single enhancer, we included all associated barcode tags that met the aforementioned barcode length and read count requirements. Within our dictionary, we did not find barcode tags that were matched to multiple enhancers. In total, the dictionary contains 90 enhancers that were uniquely mapped to one or more barcode tags, and a total of 640 barcode tag-enhancer pairs.

SEL-seq data analysis—For the whole embryo library, we sequenced barcode tags from the DNA and RNA libraries on the Illumina HiSeq 4000. Reads that perfectly matched barcode tags in our barcode tag-enhancer dictionary were included in the subsequent analysis.

We extracted all of the read sequences from the sequencing libraries and collapse them based on unique sequences, tabulating the number of times a unique sequence appears in the library. Next, we perform preliminary filtering on the unique sequences, filtering out sequences that (i) have N's present, (ii) are missing the GFP sequence after our expected location of the barcode tag, (iii) contain a barcode that is not an exact match to our

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

enhancer-barcode tag dictionary, (iv) did not meet the minimum read cutoff of 25 reads. For the preliminary filtering step, all DNA and RNA libraries were processed separately.

We normalize our data into RPM. We filter our data to only include the set of barcode tags and enhancers that appear in DNA across all replicates and consolidate the expression for each enhancer by taking the average RPM value across barcode tags. For determining if an enhancer was active, we calculated an "enhancer activity score." This score is calculated by averaging the log2(RNA/DNA) value across a given enhancer's biological replicates.

QUANTIFICATION AND STATISTICAL ANALYSES

To assess statistical differences between enhancer expression, Fischer's exact test was used with the fisher.test function in R. To assess statistical differences between enhancer expression levels, chi-squared test was used with the CHISQ.TEST function in Excel.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

ACKNOWLEDGMENTS

We thank the Farley lab and Dennis Schifferl for helpful discussions. We thank Janet H.T. Song for her critical reading of the manuscript. We thank the IGM Genomics Center for their assistance with sequencing. We thank the San Diego Supercomputer Center for providing computational resources through the Triton Shared Computer Cluster.[91] B.P.S. was supported by NIH (T32 GM133351). M.F.R. was supported by T32 (GM008666). K.T. is supported by NSF (2109907 and 3DP2HG010013-01S1). G.A.J. was supported by a Hartwell Fellowship and NIH (T32HL007444). E.K.F., B.P.S., M.F.R., K.T., G.A.J., J.L.G., and S.H.L. were supported by NIH (DP2HG010013).

REFERENCES

1. Levine M (2010). Transcriptional enhancers in animal development and evolution. Curr. Biol 20, R754–R763. [PubMed: 20833320]
2. Heinz S, Benner C, Spann N, Bertolino E, Lin YC, Laslo P, Cheng JX, Murre C, Singh H, and Glass CK (2010). Simple combinations of lineage-determining transcription factors prime cis-regulatory elements required for macrophage and B cell identities. Mol. Cell 38, 576–589. [PubMed: 20513432]
3. Liu F, and Posakony JW (2012). Role of architecture in the function and specificity of two notch-regulated transcriptional enhancer modules. PLoS Genet. 8, e1002796. [PubMed: 22792075]
4. Small S, Blair A, and Levine M (1992). Regulation of even-skipped stripe 2 in the Drosophila embryo. EMBO J. 11, 4047–4057. [PubMed: 1327756]
5. Spitz F, and Furlong EEM (2012). Transcription factors: from enhancer binding to developmental control. Nat. Rev. Genet 13, 613–626. [PubMed: 22868264]
6. Swanson CI, Evans NC, and Barolo S (2010). Structural rules and complex regulatory circuitry constrain expression of a notch- and EGFR-regulated eye enhancer. Dev. Cell 18, 359–370. [PubMed: 20230745]
7. Maurano MT, Humbert R, Rynes E, Thurman RE, Haugen E, Wang H, Reynolds AP, Sandstrom R, Qu H, Brody J, et al. (2012). Systematic localization of common disease-associated variation in regulatory DNA. Science 337, 1190–1195. [PubMed: 22955828]
8. Tak YG, and Farnham PJ (2015). Making sense of GWAS: using epigenomics and genome engineering to understand the functional relevance of SNPs in non-coding regions of the human genome. Epigenet. Chromatin 8, 57.
9. Visel A, Rubin EM, and Pennacchio LA (2009). Genomic views of distant-acting enhancers. Nature 461, 199–205. [PubMed: 19741700]

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

10. Arnone MI, and Davidson EH (1997). The hardwiring of development: organization and function of genomic regulatory systems. Development 124, 1851–1864. [PubMed: 9169833]
11. Barolo S (2016). How to tune an enhancer. Proc. Natl. Acad. Sci. USA 113, 6330–6331. [PubMed: 27247414]
12. Levo M, and Segal E (2014). In pursuit of design principles of regulatory sequences. Nat. Rev. Genet 15, 453–468. [PubMed: 24913666]
13. Thanos D, and Maniatis T (1995). Virus induction of human IFN beta gene expression requires the assembly of an enhanceosome. Cell 83, 1091–1100. [PubMed: 8548797]
14. Jindal GA, and Farley EK (2021). Enhancer grammar in development, evolution, and disease: dependencies and interplay. Dev. Cell 56, 575–587. [PubMed: 33689769]
15. Farley EK, Olson KM, Zhang W, Rokhsar DS, and Levine MS (2016). Syntax compensates for poor binding sites to encode tissue specificity of developmental enhancers. Proc. Natl. Acad. Sci. USA 113, 6508–6513. [PubMed: 27155014]
16. Dykes IM, Szumska D, Kuncheria L, Puliyadi R, Chen CM, Papanayotou C, Lockstone H, Dubourg C, David V, Schneider JE, et al. (2018). A requirement for Zic2 in the regulation of nodal expression underlies the establishment of left-sided identity. Sci. Rep 8, 10439. [PubMed: 29992973]
17. Matsumoto J, Kumano G, and Nishida H (2007). Direct activation by Ets and Zic is required for initial expression of the Brachyury gene in the ascidian notochord. Dev. Biol 306, 870–882. [PubMed: 17459364]
18. Herrmann BG, and Kispert A (1994). The T genes in embryogenesis. Trends Genet. 10, 280–286. [PubMed: 7940757]
19. Stemple DL (2005). Structure and function of the notochord: an essential organ for chordate development. Development 132, 2503–2512. [PubMed: 15890825]
20. Chesley P (1935). Development of the short-tailed mutant in the house mouse. J. Exp. Zool 70, 429–459.
21. Chiba S, Jiang D, Satoh N, and Smith WC (2009). Brachyury null mutant-induced defects in juvenile ascidian endodermal organs. Development 136, 35–39. [PubMed: 19019990]
22. Wilkinson DG, Bhatt S, and Herrmann BG (1990). Expression pattern of the mouse T gene and its role in mesoderm formation. Nature 343, 657–659. [PubMed: 1689462]
23. Yasuo H, and Satoh N (1993). Function of vertebrate T gene. Nature 364, 582–583.
24. Elms P, Scurry A, Davies J, Willoughby C, Hacker T, Bogani D, and Arkell R (2004). Overlapping and distinct expression domains of Zic2 and Zic3 during mouse gastrulation. Gene Expr. Patterns 4, 505–511. [PubMed: 15261827]
25. Imai KS, Satou Y, and 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, 2723–2732. [PubMed: 12015299]
26. Kumano G, Yamaguchi S, and Nishida H (2006). Overlapping expression of FoxA and Zic confers responsiveness to FGF signaling to specify notochord in ascidian embryos. Dev. Biol 300, 770–784. [PubMed: 16950241]
27. Warr N, Powles-Glover N, Chappell A, Robson J, Norris D, and Arkell RM (2008). Zic2 -associated holoprosencephaly is caused by a transient defect in the organizer region during gastrulation. Hum. Mol. Genet 17, 2986–2996. [PubMed: 18617531]
28. Yagi K, Satou Y, and Satoh N (2004). A zinc finger transcription factor, ZicL, is a direct activator of *Brachyury* in the notochord specification of *Ciona intestinalis*. Development 131, 1279–1288. [PubMed: 14993185]
29. Imai KS, Satoh N, and Satou Y (2002). Early embryonic expression of *FGF4/6/9* gene and its role in the induction of mesenchyme and notochord in *Ciona savignyi* embryos. Development 129, 1729–1738. [PubMed: 11923208]
30. Miya T, and Nishida H (2003). An Ets transcription factor, HrEts, is target of FGF signaling and involved in induction of notochord, mesenchyme, and brain in ascidian embryos. Dev. Biol 261, 25–38. [PubMed: 12941619]
31. Schulte-Merker S, and Smith JC (1995). Mesoderm formation in response to Brachyury requires FGF signalling. Curr. Biol 5, 62–67. [PubMed: 7535172]

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

32. Yasuo H, and 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. [PubMed: 17022960]
33. Ang S-L, and Rossant J (1994). HNF-3β is essential for node and notochord formation in mouse development. Cell 78, 561–574. [PubMed: 8069909]
34. Dal-Pra S, Thisse C, and Thisse B (2011). FoxA transcription factors are essential for the development of dorsal axial structures. Dev. Biol 350, 484–495. [PubMed: 21172337]
35. José-Edwards DS, Oda-Ishii I, Kugler JE, Passamaneck YJ, Katikala L, Nibu Y, and Di Gregorio A (2015). Brachyury, Foxa2 and the cis-regulatory origins of the notochord. PLoS Genet. 11, e1005730. [PubMed: 26684323]
36. Katikala L, Aihara H, Passamaneck YJ, Gazdoiu S, José-Edwards DS, Kugler JE, Oda-Ishii I, Imai JH, Nibu Y, and Di Gregorio A (2013). Functional Brachyury binding sites establish a temporal read-out of gene expression in the Ciona notochord. PLoS Biol. 11, e1001697. [PubMed: 24204212]
37. Passamaneck YJ, Katikala L, Perrone L, Dunn MP, Oda-Ishii I, and Di Gregorio A (2009). Direct activation of a notochord cis-regulatory module by Brachyury and FoxA in the ascidian Ciona intestinalis. Development 136, 3679–3689. [PubMed: 19820186]
38. Weinstein DC, Ruiz i Altaba A, Chen WS, Hoodless P, Prezioso VR, Jessell TM, and Darnell JE Jr. (1994). The winged-helix transcription factor HNF-3β is required for notochord development in the mouse embryo. Cell 78, 575–588. [PubMed: 8069910]
39. Delsuc F, Brinkmann H, Chourrout D, and Philippe H (2006). Tunicates and not cephalochordates are the closest living relatives of vertebrates. Nature 439, 965–968. [PubMed: 16495997]
40. Davidson B, and Christiaen L (2006). Linking chordate gene networks to cellular behavior in ascidians. Cell 124, 247–250. [PubMed: 16439196]
41. Farley EK, Olson KM, Zhang W, Brandt AJ, Rokhsar DS, and Levine MS (2015). Suboptimization of developmental enhancers. Science 350, 325–328. [PubMed: 26472909]
42. Di Gregorio A (2020). The notochord gene regulatory network in chordate evolution: conservation and divergence from Ciona to vertebrates. Curr. Top. Dev. Biol 139, 325–374. [PubMed: 32450965]
43. Veeman MT, Nakatani Y, Hendrickson C, Ericson V, Lin C, and Smith WC (2008). Chongmague reveals an essential role for laminin-mediated boundary formation in chordate convergence and extension movements. Development 135, 33–41. [PubMed: 18032448]
44. Corbo JC, Levine M, and Zeller RW (1997). Characterization of a notochord-specific enhancer from the Brachyury promoter region of the ascidian, Ciona intestinalis. Development 124, 589–602. [PubMed: 9043074]
45. Schifferl D, Scholze-Wittler M, Wittler L, Veenvliet JV, Koch F, and Herrmann BG (2021). A 37 kb region upstream of *brachyury* comprising a notochord enhancer is essential for notochord and tail development. Development 148, dev200059. [PubMed: 34822716]
46. Takahashi H, Mitani Y, Satoh G, and Satoh N (1999). Evolutionary alterations of the minimal promoter for notochord-specific Brachyury expression in ascidian embryos. Development 126, 3725–3734. [PubMed: 10433903]
47. Lamber EP, Vanhille L, Textor LC, Kachalova GS, Sieweke MH, and Wilmanns M (2008). Regulation of the transcription factor Ets-1 by DNA-mediated homo-dimerization. EMBO J. 27, 2006–2017. [PubMed: 18566588]
48. Wei G-H, Badis G, Berger MF, Kivioja T, Palin K, Enge M, Bonke M, Jolma A, Varjosalo M, Gehrke AR, et al. (2010). Genome-wide analysis of ETS-family DNA-binding in vitro and in vivo. EMBO J. 29, 2147–2160. [PubMed: 20517297]
49. Rothbächer U, Bertrand V, Lamy C, and Lemaire P (2007). A combinatorial code of maternal GATA, Ets and β-catenin-TCF transcription factors specifies and patterns the early ascidian ectoderm. Development 134, 4023–4032. [PubMed: 17965050]
50. Stolfi A, Ryan K, Meinertzhagen IA, and Christiaen L (2015). Migratory neuronal progenitors arise from the neural plate borders in tunicates. Nature 527, 371–374. [PubMed: 26524532]
51. Jiang D, and Smith WC (2007). Ascidian notochord morphogenesis. Dev. Dynam 236, 1748–1757.

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

52. Imai KS, Hino K, Yagi K, Satoh N, and Satou Y (2004). Gene expression profiles of transcription factors and signaling molecules in the ascidian embryo: towards a comprehensive understanding of gene networks. Development 131, 4047–4058. [PubMed: 15269171]
53. Winkley KM, Reeves WM, and Veeman MT (2021). Single-cell analysis of cell fate bifurcation in the chordate Ciona. BMC Biol. 19, 180. [PubMed: 34465302]
54. Hudson C, Sirour C, and Yasuo H (2016). Co-expression of Foxa.a, Foxd and Fgf9/16/20 defines a transient mesendoderm regulatory state in ascidian embryos. Elife 5, e14692. [PubMed: 27351101]
55. Hudson C, Lotito S, and Yasuo H (2007). Sequential and combinatorial inputs from Nodal, Delta2/Notch and FGF/MEK/ERK signalling pathways establish a grid-like organisation of distinct cell identities in the ascidian neural plate. Development 134, 3527–3537. [PubMed: 17728350]
56. Imai KS, Levine M, Satoh N, and Satou Y (2006). Regulatory blueprint for a chordate embryo. Science 312, 1183–1187. [PubMed: 16728634]
57. Picco V, Hudson C, and Yasuo H (2007). Ephrin-Eph signalling drives the asymmetric division of notochord/neural precursors in Ciona embryos. Development 134, 1491–1497. [PubMed: 17344225]
58. Wagner E, and Levine M (2012). FGF signaling establishes the anterior border of the Ciona neural tube. Development 139, 2351–2359. [PubMed: 22627287]
59. Ikeda T, and Satou Y (2017). Differential temporal control of *Foxa.a* and *Zinc-r.b* specifies brain versus notochord fate in the ascidian embryo. Development 144, 38–43. 10.1242/dev.142174. [PubMed: 27888196]
60. Casey ES, O'Reilly MA, Conlon FL, and Smith JC (1998). The T-box transcription factor Brachyury regulates expression of eFGF through binding to a non-palindromic response element. Development 125, 3887–3894. [PubMed: 9729496]
61. Li J, Dantas Machado AC, Guo M, Sagendorf JM, Zhou Z, Jiang L, Chen X, Wu D, Qu L, Chen Z, et al. (2017). Structure of the fork-head domain of FOXA2 bound to a complete DNA consensus site. Biochemistry 56, 3745–3753. [PubMed: 28644006]
62. Conlon FL, Fairclough L, Price BM, Casey ES, and Smith JC (2001). Determinants of T box protein specificity. Development 128, 3749–3758. [PubMed: 11585801]
63. Di Gregorio A, and Levine M (1999). Regulation of Ci-tropomyosin-like, a Brachyury target gene in the ascidian, Ciona intestinalis. Development 126, 5599–5609. [PubMed: 10572037]
64. Dunn MP, and Di Gregorio A (2009). The evolutionarily conserved leprecan gene: its regulation by Brachyury and its role in the developing Ciona notochord. Dev. Biol 328, 561–574. [PubMed: 19217895]
65. Müller CW, and Herrmann BG (1997). Crystallographic structure of the T domain–DNA complex of the Brachyury transcription factor. Nature 389, 884–888. [PubMed: 9349824]
66. Nitta KR, Jolma A, Yin Y, Morgunova E, Kivioja T, Akhtar J, Hens K, Toivonen J, Deplancke B, Furlong EEM, and Taipale J (2015). Conservation of transcription factor binding specificities across 600 million years of bilateria evolution. Elife 4, e04837. [PubMed: 25779349]
67. Reeves WM, Wu Y, Harder MJ, and Veeman MT (2017). Functional and evolutionary insights from the Ciona notochord transcriptome. Development 144, 3375–3387. [PubMed: 28928284]
68. Scott A, and Stemple DL (2005). Zebrafish notochordal basement membrane: signaling and structure. Curr. Top. Dev. Biol 65, 229–253. [PubMed: 15642386]
69. Machingo QJ, Fritz A, and Shur BD (2006). A beta1,4-galactosyl-transferase is required for convergent extension movements in zebrafish. Dev. Biol 297, 471–482. [PubMed: 16904099]
70. Parsons MJ, Campos I, Hirst EMA, and Stemple DL (2002). Removal of dystroglycan causes severe muscular dystrophy in zebrafish embryos. Development 129, 3505–3512. [PubMed: 12091319]
71. Pollard SM, Parsons MJ, Kamei M, Kettleborough RNW, Thomas KA, Pham VN, Bae MK, Scott A, Weinstein BM, and Stemple DL (2006). Essential and overlapping roles for laminin alpha chains in notochord and blood vessel formation. Dev. Biol 289, 64–76. [PubMed: 16321372]
72. Barnett MW, Old RW, and Jones EA (1998). Neural induction and patterning by fibroblast growth factor, notochord and somite tissue in Xenopus. Dev. Growth Differ 40, 47–57. [PubMed: 9563910]

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

73. Olivera-Martinez I, Harada H, Halley PA, and Storey KG (2012). Loss of FGF-dependent mesoderm identity and rise of endogenous retinoid signalling determine cessation of body Axis elongation. PLoS Biol. 10, e1001415. [PubMed: 23118616]
74. Lolas M, Valenzuela PDT, Tjian R, and Liu Z (2014). Charting Brachyury-mediated developmental pathways during early mouse embryogenesis. Proc. Natl. Acad. Sci. USA 111, 4478–4483. [PubMed: 24616493]
75. Reeves WM, Shimai K, Winkley KM, and Veeman MT (2021). Brachyury controls Ciona notochord fate as part of a feed-forward network. Development 148, dev195230. [PubMed: 33419874]
76. Fujiwara S, Corbo JC, and Levine M (1998). The snail repressor establishes a muscle/notochord boundary in the Ciona embryo. Development 125, 2511–2520. [PubMed: 9609834]
77. Shimai K, and Veeman M (2021). Quantitative dissection of the proximal Ciona brachyury enhancer. Front. Cell Dev. Biol 9, 804032. [PubMed: 35127721]
78. Frankel N, Davis GK, Vargas D, Wang S, Payre F, and Stern DL (2010). Phenotypic robustness conferred by apparently redundant transcriptional enhancers. Nature 466, 490–493. [PubMed: 20512118]
79. Hong J-W, Hendrix DA, and Levine MS (2008). Shadow enhancers as a source of evolutionary novelty. Science 321, 1314. [PubMed: 18772429]
80. Perry MW, Boettiger AN, Bothma JP, and Levine M (2010). Shadow enhancers foster robustness of Drosophila gastrulation. Curr. Biol 20, 1562–1567. [PubMed: 20797865]
81. Antosova B, Smolikova J, Klimova L, Lachova J, Bendova M, Kozmikova I, Machon O, and Kozmik Z (2016). The gene regulatory network of lens induction is wired through meis-dependent Shadow enhancers of Pax6. PLoS Genet. 12, e1006441. [PubMed: 27918583]
82. Osterwalder M, Barozzi I, Tissières V, Fukuda-Yuzawa Y, Mannion BJ, Afzal SY, Lee EA, Zhu Y, Plajzer-Frick I, Pickle CS, et al. (2018). Enhancer redundancy provides phenotypic robustness in mammalian development. Nature 554, 239–243. [PubMed: 29420474]
83. Pijuan-Sala B, Wilson NK, Xia J, Hou X, Hannah RL, Kinston S, Calero-Nieto FJ, Poirion O, Preissl S, Liu F, and Göttgens B (2020). Single-cell chromatin accessibility maps reveal regulatory programs driving early mouse organogenesis. Nat. Cell Biol 22, 487–497. [PubMed: 32231307]
84. Harvey SA, Tümpel S, Dubrulle J, Schier AF, and Smith JC (2010). No tail integrates two modes of mesoderm induction. Development 137, 1127–1135. [PubMed: 20215349]
85. King DM, Hong CKY, Shepherdson JL, Granas DM, Maricque BB, and Cohen BA (2020). Synthetic and genomic regulatory elements reveal aspects of cis-regulatory grammar in mouse embryonic stem cells. Elife 9, e41279. [PubMed: 32043966]
86. Fuqua T, Jordan J, van Breugel ME, Halavatyi A, Tischer C, Polidoro P, Abe N, Tsai A, Mann RS, Stern DL, and Crocker J (2020). Dense and pleiotropic regulatory information in a developmental enhancer. Nature 587, 235–239. [PubMed: 33057197]
87. Wong ES, Zheng D, Tan SZ, Bower NL, Garside V, Vanwalleghem G, Gaiti F, Scott E, Hogan BM, Kikuchi K, et al. (2020). Deep conservation of the enhancer regulatory code in animals. Science 370, eaax8137. [PubMed: 33154111]
88. Yona AH, Alm EJ, and Gore J (2018). Random sequences rapidly evolve into de novo promoters. Nat. Commun 9, 1530. [PubMed: 29670097]
89. de Boer CG, Vaishnav ED, Sadeh R, Abeyta EL, Friedman N, and Regev A (2020). Deciphering eukaryotic gene-regulatory logic with 100 million random promoters. Nat. Biotechnol 38, 56–65. [PubMed: 31792407]
90. Galupa R, Alvarez-Canales G, Borst NO, Fuqua T, Gandara L, Misunou N, Richter K, Alves MR, Karumbi E, Perkins ML, and Kocijan T (2022). Enhancer architecture and chromatin accessibility constrain phenotypic space during development. Preprint at bioRxiv. 10.1101/2022.06.02.494376.
91. Center SDS (2022). Triton shared computing cluster. 10.57873/T34W2R.
92. Satou Y, Nakamura R, Yu D, Yoshida R, Hamada M, Fujie M, Hisata K, Takeda H, and Satoh N (2019). A nearly complete genome of Ciona intestinalis type A (C. Robusta) reveals the contribution of inversion to chromosomal evolution in the genus Ciona. Genome Biol. Evol 11, 3144–3157. [PubMed: 31621849]

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

93. Grüning B, Dale R, Sjödin A, Chapman BA, Rowe J, Tomkins-Tinch CH, Valieris R, and Köster J; Bioconda Team (2018). Bioconda: sustainable and comprehensive software distribution for the life sciences. Nat. Methods 15, 475–476. [PubMed: 29967506]

94. Cock PJA, Antao T, Chang JT, Chapman BA, Cox CJ, Dalke A, Friedberg I, Hamelryck T, Kauff F, Wilczynski B, and de Hoon MJL (2009). Biopython: freely available Python tools for computational molecular biology and bioinformatics. Bioinformatics 25, 1422–1423. [PubMed: 19304878]

95. Ewels P, Magnusson M, Lundin S, and Käller M (2016). MultiQC: summarize analysis results for multiple tools and samples in a single report. Bioinformatics 32, 3047–3048. [PubMed: 27312411]

96. Magoč T, and Salzberg SL (2011). FLASH: fast length adjustment of short reads to improve genome assemblies. Bioinformatics 27, 2957–2963. [PubMed: 21903629]

97. Harris CR, Millman KJ, van der Walt SJ, Gommers R, Virtanen P, Cournapeau D, Wieser E, Taylor J, Berg S, Smith NJ, et al. (2020). Array programming with NumPy. Nature 585, 357–362. [PubMed: 32939066]

98. Hunter JD (2007). A 2D graphics environment. Comput. Sci. Eng 9, 90–95.

99. Pedregosa F, Varoquaux G, Gramfort A, Michel V, Thirion B, Grisel O, Blondel M, Prettenhofer P, Weiss R, Dubourg V, and Vanderplas J (2011). Scikit-learn: machine learning in Python. J. Mach. Learn. Res 12, 2825–2830.

100. Waskom M (2021). seaborn: statistical data visualization. J. Open Source Softw 6, 3021.

101. Hume MA, Barrera LA, Gisselbrecht SS, and Bulyk ML (2015). UniPROBE, update 2015: new tools and content for the online database of protein-binding microarray data on protein–DNA interactions. Nucleic Acids Res. 43, D117–D122. [PubMed: 25378322]

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

Highlights

- Screens of genomic elements in developing embryos discover notochord enhancers
- Diverse logics/combinations of binding sites can encode notochord enhancer activity
- Enhancer grammar is critical for notochord enhancer activity
- Signatures of enhancer logic and grammar are conserved across chordates

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

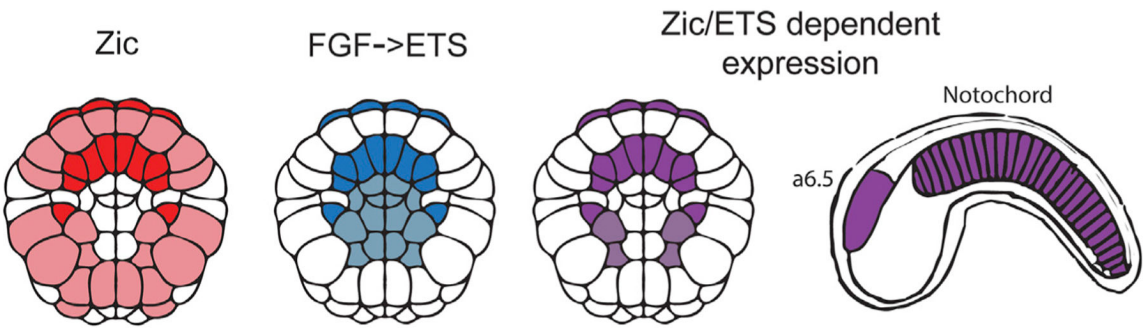


Figure 1. Zic and ETS expression in the 110-cell stage *Ciona* embryo

Co-expression of Zic (red) and ETS (blue) in 110-cell stage *Ciona* embryos is shown in purple and occurs in the notochord, a6.5 lineage, which gives rise to the anterior sensory vesicle and palps, and four mesenchyme cells shown in light purple. A schematic of the tailbud embryo shows the notochord and a6.5 cell types later in development. Dark coloring represents a6.5 and notochord lineages, and light coloring represents other tissues with expression of Zic and/or ETS.

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

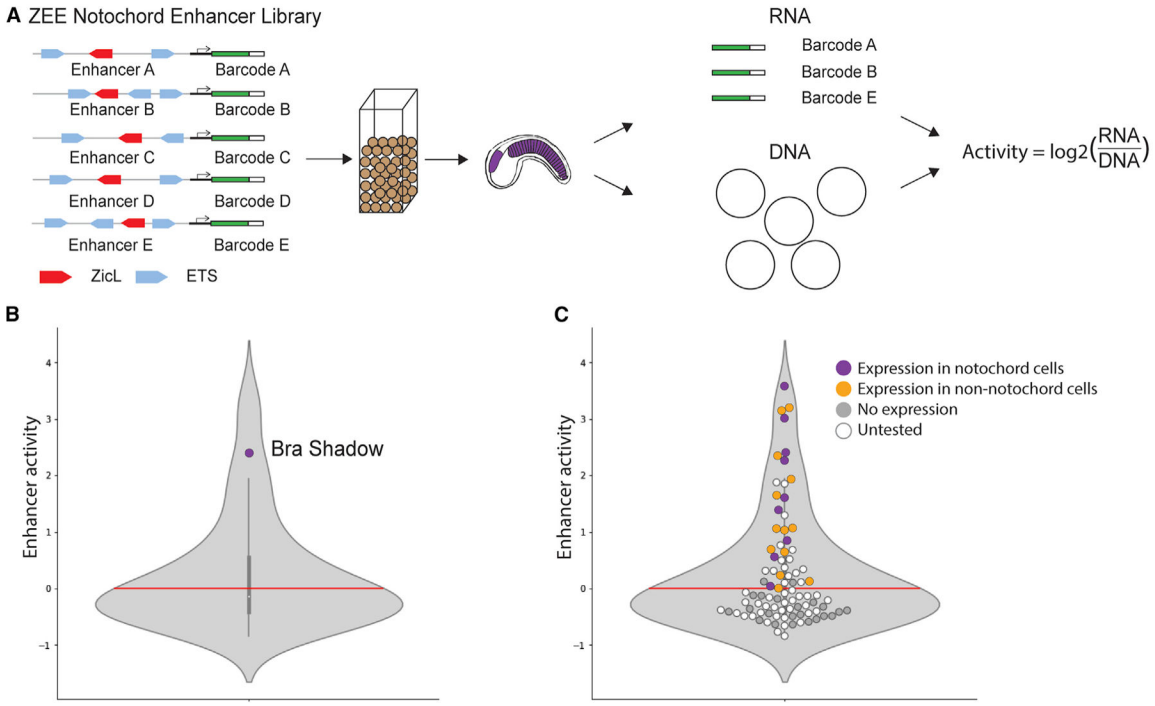


Figure 2. Screening Zic and ETS genomic elements in *Ciona*

(A) Schematic of enhancer screen. Ninety ZEE genomic regions, each associated with on average six unique barcodes, were electroporated into fertilized *Ciona* eggs. mRNA and plasmid DNA were extracted from 5.5 hpf embryos (tailbud embryo shown to highlight tissues with predicted expression). The mRNA and DNA barcodes were sequenced, and a normalized enhancer activity score was calculated for each enhancer by taking the log2 of the mRNA activity for a given enhancer divided by the number of copies of the plasmid.

(B) Violin plot showing the distribution of enhancer activity. The Bra Shadow enhancer served as a positive control and is labeled. The red line indicates the cutoff for non-functional elements at zero.

(C) Same plot as (B), but with all 90 ZEE elements plotted as dots. Dots are colored by the results of an orthogonal screen, where we measured the GFP expression in 150 embryos per enhancer to determine the location of expression (3 biological replicates of 50 embryos). Enhancers driving notochord expression are shown in purple, enhancers with expression but no notochord expression are shown in orange. ZEE elements that do not drive expression are gray and untested enhancers are shown in white.

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

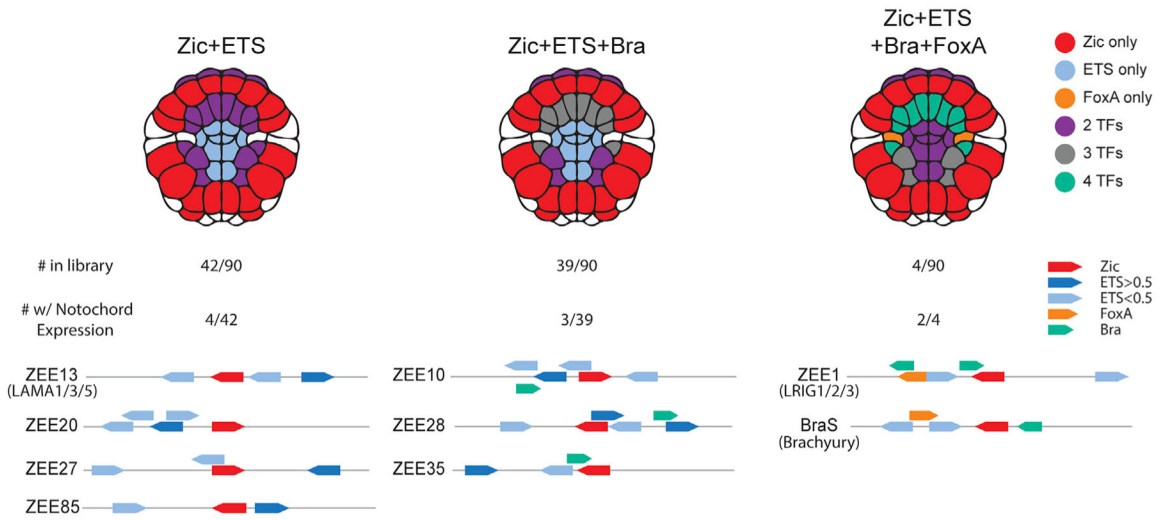


Figure 3. Combinations of transcription factors in ZEE enhancers that drive notochord expression

Notochord-expressing ZEE elements were grouped by the combination of transcription factor binding sites present in each element. For each combination, an embryo schematic shows the overlapping region of expression for that given combination. Below the embryo schematic, the number of ZEE elements, the number of ZEE elements with notochord expression and schematics of the ZEE elements with notochord expression within each group. Zic (red), ETS (blue), FoxA (orange), and Bra (green) sites are annotated. Dark blue ETS sites have an affinity of greater than 0.5, light blue sites have an affinity of less than 0.5.

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

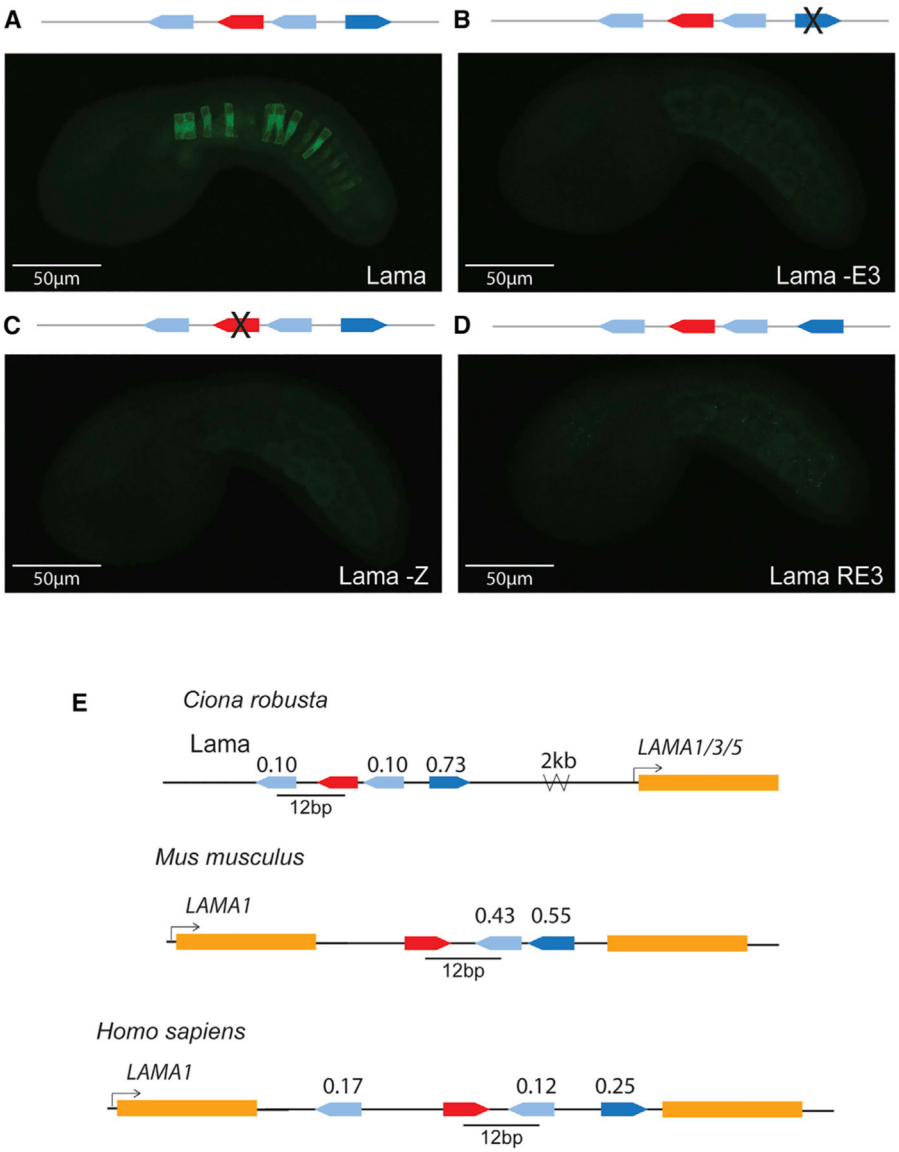


Figure 4. Zic and ETS grammar encodes a notochord *laminin alpha* enhancer

(A) Embryo electroporated with the Lama enhancer (ZEE13); GFP expression can be seen in the notochord.

(B) Embryo electroporated with Lama -E3, where ETS3 was mutated to be non-functional; no GFP expression detected.

(C) Embryo electroporated with Lama -Z, where the Zic was mutated to be non-functional; no GFP expression detected.

(D) Embryo electroporated with Lama RE3, where the sequence of ETS3 was reversed; no GFP expression detected. Comparable results were seen when ETS1 was reversed.

In (A)–(D), for each enhancer, three biological replicates were performed with 50 embryos per replicate (see Table S4). Each image in this figure is representative of the expression observed from three biological replicates. Scale bars, 50μm.

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

(E) Schematics of Zic and ETS clusters near *laminin alpha* in the genome of *Ciona*, mouse, and human. All three *laminin alpha* clusters have a spacing of 12 bp between an ETS and Zic site and all contain non-consensus ETS sites. ETS site affinity scores are noted above each site. Dark blue ETS sites have an affinity of greater than 0.5, light blue sites have an affinity of less than 0.5.

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

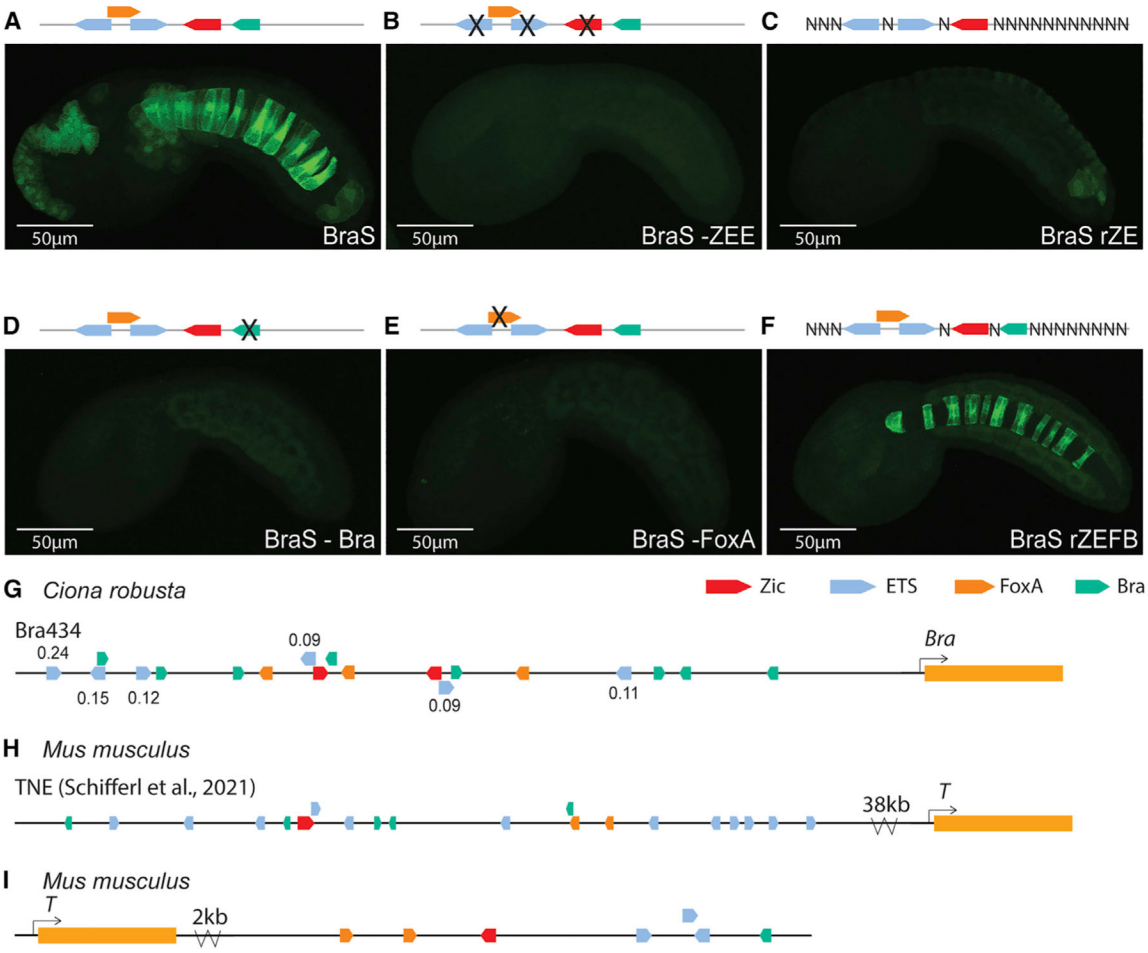


Figure 5. Zic, ETS, FoxA, and Bra may be a common regulatory logic for Brachyury enhancers

(A) Embryo electroporated with the Bra Shadow (BraS) enhancer; GFP expression can be seen in the notochord.

(B) Embryo electroporated with BraS -ZEE, where the Zic and two ETS sites were mutated to be non-functional; no GFP expression was detected.

(C) Embryo electroporated with BraS rZE, where the Zic and two ETS sites were fixed, and all other nucleotides were randomized; GFP expression was greatly diminished.

(D) Embryo electroporated with BraS -Bra, where the sequence of Bra was mutated to be non-functional; GFP expression was greatly diminished.

(E) Embryo electroporated with BraS -FoxA, where the sequence of FoxA was mutated to be non-functional; GFP expression was greatly diminished.

(F) Embryo electroporated with BraS rZEFB, where the Zic, two ETS, FoxA, and Bra sites were fixed, and all other nucleotides were randomized; GFP expression can be seen in the notochord. In (A)–(F), for each enhancer, two biological replicates were performed with 50 embryos per replicate (see Table S4).

Each image in this figure is representative of the expression observed from two biological replicates. Scale bars, 50µm.

(G–I) Schematics of Zic (red), ETS (blue), FoxA (orange), and Bra (green) clusters near *Bra* in the genomes of *Ciona* and mouse.

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

KEY RESOURCES TABLE

REAGENT or RESOURCE	SOURCE	IDENTIFIER
Deposited data		
snATACseq mouse E8.25	Pijuan-Sala et al.[83]	GEO: GSE133244
FACS-sorted notochord RNA-Seq	Reeves et al.[67]	N/A
Human reference genome NCBI build 38, GRCh38	Genome Reference Consortium	https://www.ncbi.nlm.nih.gov/grc/human
Mouse reference genome NCBI build 39,	Genome Reference Consortium	https://www.ncbi.nlm.nih.gov/grc/mouse
Ciona robusta genome	Satou et al.[92]	N/A
mouse ETS-1 universal PBM data	Wei et al.[48]	https://thebrain.bwh.harvard.edu/uniprobe/index.php
ZEE library screen	This paper	SRA: PRJNA861319; https://www.ncbi.nlm.nih.gov/sra/PRJNA861319
Experimental models: organisms/strains		
Ciona intestinalis type A (Ciona robusta)	M-Rep	N/A
Oligonucleotides		
Oligonucleotides for library screen, see Table S1	This paper	N/A
Oligonucleotides for mutagenesis, see Table S4	This paper	N/A
Recombinant DNA		
Plasmid: BraS bpFog > GFP	Farley lab	N/A
Plasmid: BraS -ZEE bpFog > GFP	This paper	N/A
Plasmid: BraS rZE bpFog > GFP	This paper	N/A
Plasmid: BraS -FoxA bpFog > GFP	This paper	N/A
Plasmid: BraS -Bra bpFog > GFP	This paper	N/A
Plasmid: BraS rZEFB bpFog > GFP	This paper	N/A
Plasmid: Lama1 bpFog > GFP	This paper	N/A
Plasmid: Lama1 bpFog > GFP	This paper	N/A
Plasmid: Lama1 -E3 bpFog > GFP	This paper	N/A
Plasmid: Lama1 -Z bpFog > GFP	This paper	N/A
Plasmid: Lama1 RE3 bpFog > GFP	This paper	N/A
Software and algorithms		
Python (version 3.8.6)	Python Software Foundation	https://www.python.org
Conda (version 4.9.2)	Anaconda, Inc.	https://docs.conda.io/projects/conda/en/latest/
Bioconda	Grüning et al.[93]	https://bioconda.github.io
Biopython (version 1.78)	Cock et al.[94]	https://biopython.org
FastQC (version 0.11.9)	Babraham Bioinformatics, Babraham Institute	https://www.bioinformatics.babraham.ac.uk/projects/fastqc/
MultiQC (version 1.8)	Ewels et al.[95]	https://multiqc.info
FLASH (version 1.2.11)	Magoč et al.[96]	http://www.cbcb.umd.edu/software/flash
pandas (version 1.2.1)	NumFOCUS	https://pandas.pydata.org
NumPy (version 1.20.3)	Harris et al.[97]	https://numpy.org
Matplotlib (version 3.2.2)	Hunter[98]	https://matplotlib.org/stable/index.html
scikit-learn (version 0.24.1)	Pedregosa et al.[99]	https://scikit-learn.org/stable/index.html

REAGENT or RESOURCE	SOURCE	IDENTIFIER
seaborn (version 0.11.1)	Waskom et al.[100]	https://seaborn.pydata.org/index.html
Diverse-Logics-Notochord-Study	Code used in this paper	https://github.com/farleylab/Diverse-Logics-Notochord-Study

Author Manuscript Author Manuscript Author Manuscript Author Manuscript