I-SceI Meganuclease-mediated transgenesis in the acorn worm, *Saccoglossus kowalevskii*

Paul J. Minor[1*#], D. Nathaniel Clarke[1*], José M. Andrade López[1], Jens H. Fritzenwanker[3] Jessica Gray[2], Christopher J. Lowe[1#]

1. Hopkins Marine Station
   Department of Biology
   Stanford University
   120 Oceanview Blvd.
   Pacific Grove, CA 93950

2. Department of Systems Biology
   Harvard Medical School
   200 Longwood Ave
   Boston, MA 02115

3. Department of Biology
   Georgetown University
   411 Regents Hall
   37th and O Streets, NW
   Washington DC, 20057

* These authors contributed equally to this work

Authors for correspondence: PJM (pminor@stanford.edu), CJL (clowe@stanford.edu)

Abstract

Hemichordates are a phylum of marine invertebrate deuterostomes that are closely related to chordates, and represent one of the most promising models to provide insights into early deuterostome evolution. The genome of the hemichordate, *Saccoglossus kowalevskii,* reveals an extensive set of non-coding elements conserved across all three deuterostome phyla. Functional characterization and cross-phyla comparisons of these putative regulatory elements will enable a better understanding of enhancer evolution, and subsequently how changes in gene regulation give rise to morphological innovation. Here, we describe an efficient method of transgenesis for the characterization of non-coding elements in *S. kowalevskii*. We first test the capacity of an I-SceI transgenesis system to drive ubiquitous or regionalized gene expression, and to label specific cell types. Finally, we identified a minimal promoter that can be used to test the capacity of putative enhancers in *S. kowalevskii*. This work demonstrates that this I-SceI transgenesis technique, when coupled with an understanding of chromatin accessibility, can be a powerful tool for studying how evolutionary changes in gene regulatory mechanisms contributed to the diversification of body plans in deuterostomes.

Keywords: Hemichordate, transgenesis, enhancer, evolution

Introduction

Transcriptional enhancers direct the spatiotemporal expression of genes, thus playing a crucial role in development. Advances in genome-wide analysis of chromatin state and transcription factor binding have predicted hundreds of thousands of potential regulatory elements in the human genome (Neph et al., 2012; Rada-Iglesias et al., 2011), and comparative genomic studies have suggested that components of this *cis*-regulatory landscape may be broadly conserved across animal phyla (Simakov et al., 2015). However, performing comparative functional studies of *cis*-regulatory elements remains challenging due to a general lack of transgenesis techniques in taxonomically important non-model systems. A goal of evolutionary developmental biology is to differentiate between *cis*-regulatory elements with ancient origins versus those with more recent origins that are associated with the emergence of morphological novelties. To explore such questions of gene regulatory evolution, efficient transgenesis techniques need to be developed in organisms in key phylogenetic positions that are informative about phenotypic innovation.

The hemichordate *S. kowalevskii* is a promising model in which to interrogate the correspondence between the evolution of *cis*-regulatory elements and deuterostome morphological innovation, due to: (1) its key phylogenetic position as a close outgroup to chordates (Cannon et al., 2014; Simakov et al., 2015); and (2) the compelling similarities in axial patterning mechanisms it shares with vertebrates that are involved in patterning key aspects of the vertebrate body plan despite strong morphological differences between the two groups (Lowe et al., 2015; Lowe et al., 2006; Lowe et al., 2003; Yao et al., 2016) . Compared to vertebrates, *S. kowalevskii* has a modest body plan consisting of three major domains (proboscis, collar, and trunk), and a relatively simple nervous system composed of a diffuse basiepithelial nerve net, and dorsal and ventral nerve cords (Kaul and Stach, 2010; Kaul-Strehlow et al., 2015; Knight-Jones, 1952). Despite this simple anatomy, recent work has revealed close similarities in the gene regulatory network regulating early vertebrate brain regionalization and the anterior ectoderm of *S. kowalevskii* (Darras et al., 2018; Lowe et al., 2003; Pani et al., 2012). Furthermore, the genome of *S. kowalevskii* displays extensive synteny with the genome of the basal chordate amphioxus, and contains more than 6,500 conserved non-coding elements found in all deuterostomes (Simakov et al., 2015). This remarkable level of genomic conservation, in

addition to the common axial patterning mechanisms found between hemichordates and chordates despite the striking differences in their body plans, makes *S. kowalevskii* a valuable system for exploring the role of *cis*-regulatory elements in the evolution of these disparate morphologies. For this study, we sought to develop an efficient transgenesis technique that could be used to functionally characterize non-coding regulatory elements found within the *S. kowalevskii* genome.

Here, we present methods for I-SceI meganuclease-mediated transgenesis in *S. kowalevskii* embryos. I-SceI is an intron-encoded restriction endonuclease originally isolated from baker's yeast, *Saccharomyces cerevisiae* (Jacquier and Dujon, 1985). Previous work has shown I-SceI enables integrative transgenesis in Medaka, zebrafish, frogs, sea urchins, newts, lamprey, and the starlet sea anemone, *Nematostella vectensis* (Casco-Robles et al., 2010; Grabher et al., 2004; Ochiai et al., 2008; Ogino et al., 2006; Parker et al., 2014b; Renfer and Technau, 2017; Thermes et al., 2002). In this technique, I-SceI protein is co-injected into the fertilized zygote with a plasmid carrying a transgene flanked by 18-bp I-SceI restriction sites, which enable the enzyme to integrate the transgene into pseudo-sites within the genome (Thermes et al., 2002). To test this transgenesis method in *S. kowalevskii*, we demonstrate that meganuclease-mediated integration of promoter:GFP fusion cassettes can be used to: (1) drive ubiquitous expression; (2) efficiently replicate expression patterns for genes with restricted spatiotemporal profiles; and (3) label specific cell types. Further, we show that for some genes, a minimal chromosome immunoprecipitation sequencing (ChIP-seq) dataset was sufficient for the identification of putative enhancer and promoter elements, and describe a minimal promoter from the *S. kowalevskii* Gbx gene that can be used to characterize putative native enhancers found in *S. kowalevskii,* as well as to test exogenous enhancers from other animals. This work adds to the toolkit of experimental techniques available in *S. kowalevskii,* and demonstrates that it is now a model capable of providing insights into deuterostome evolution and chordate origins through the functional testing of *cis*-regulatory elements.

Materials and methods

Embryonic Cultures

Gravid *S. kowalevskii* adults were collected from the intertidal in Waquoit Bay, MA and maintained in flow-through seawater tables at the Marine Biological Laboratory, Woods Hole, MA. Spawning and animal maintenance were carried out as previously described (Lowe et al., 2004).

Hemichordate Transgenesis

I-SceI meganuclease-mediated transgenesis was adapted from previously established protocols from other organisms (Ogino et al., 2006; Parker et al., 2014b). Putative *S. kowalevskii* promoters and enhancers were identified by eye using a H3K27ac ChIP-seq dataset (see below) by selecting regions strongly enriched for H3K27ac signal, and these regions plus approximately 100 – 200 bp of flanking sequence were cloned into a reporter plasmid containing I-SceI meganuclease restriction sites (Parker et al., 2014a) upstream of an eGFP coding sequence and a SV40 late polyadenylation signal sequence using Gibson assembly (Gibson et al., 2009). In cases where no obvious H3K27ac enrichment was observed, 5 to 10 kb of genomic sequence immediately upstream of the predicted started codon was used for cloning instead (as for Synapsin, Supplemental Fig. 1D). All primer sequences used in this study can be found in Supplemental Table 1. To prepare injection mixtures, 10 µl restriction digests containing 5 units of I-SceI enzyme (NEB), 1µl of CutSmart buffer, and 130 ng of reporter plasmid (final concentration 13 ng/µl), were incubated at 37°C for 40 minutes. Upon completion, 0.5 µl of 50 mg/ml rhodamine was added to the restriction digest as a tracer and this mix was injected into *S. kowalevskii* embryos as previously described (Darras et al., 2011; Lowe et al., 2006).

Live Imaging

Injected animals were screened for fluorescence using a Zeiss Axio Zoom.V16 or Zeiss SteREO Discovery.V20 stereo microscope. Photographs and movies of fluorescent embryos were captured with Zeiss Axiocam MRc5 or MRm cameras using the ZEN software package.

ChIP-seq library preparation and sequencing

The chromatin immunoprecipitation protocol was adapted from previous reports (Buecker et al., 2014). Briefly, 80-100µL of 48hpf and 72hpf old embryos were crosslinked with 1% formaldehyde for 10min at RT, quenched with glycine at a final concentration of 0.125M, and

flash frozen and stored at -80˚C. Chromatin was sheared by sonication while emerged in an ice bath to 100-500bp fragments. Sonicated chromatin was incubated with 5–10 µg antibody and incubated overnight at 4 °C. Antibody was bound to Dynabeads by incubating 1mL of antibody bound chromatin with 100µL Dynabead solution at 4˚C for 2-4h. ChIP DNA were subject to end repair, A-tailing, adaptor ligation and cleavage with USER enzyme, followed by size selection to 250–500 bp and amplification with NEBNext sequencing primers (Cat#E7335S). Libraries were purified, quantified, multiplexed (with NEBNext Multiplex Oligos for Illumina kit, E7335S). Samples were sequenced with 40-bp single-end read (~150 million reads) on an Illumina HiSEQ2000-seq (BioMicroCenter). Data were processed using the Galaxy sequence analysis platform and aligned to the *S. kowalevskii* genome using BowTie2 (Goecks et al., 2010; Langmead and Salzberg, 2012; Simakov et al., 2015). The antibody used for ChIP-seq was H3K27ac from Active Motif (#39133, 39134).

Fixation, antibody staining, and confocal microscopy

72-hour embryos were fixed and stained as described previously (Gillis et al., 2012; Lowe et al., 2004). Briefly, embryos were fixed for 30 minutes at room temperature in fixation buffer (3.7% formaldehyde, 0.1 M MOPS pH 7.5, 0.5 M NaCl, 2 mM EGTA pH 8.0, 1 mM MgCl2, 1X PBS), and subsequently stored in ethanol at -20°C. For antibody staining, embryos were rehydrated into 1x PBS + 0.1% Triton X-100 (PBSTr), rinsed 3 x 10 minutes in PBSTr, and placed into a blocking solution of PBSTr + 5% goat serum for 2 hours at room temperature. Embryos were incubated with anti-GFP antibody (Rabbit anti-GFP serum; Life Technologies #A-6455) at a 1:500 dilution overnight at 4 ˚C. After primary antibody incubation, embryos were washed 4 x 30 minutes in PBSTr, and then incubated for 4 hours at room temperature with secondary antibody (Alexa-Fluor 555 goat anti-mouse IgG, ThermoFisher #A28180) diluted 1:500 in blocking solution. Samples were then washed 4 x 30 minutes in PBSTr and transitioned into 100% glycerol. Images were caputured on a LSM 700 confocal microscope with a 10x objective using the Zen software package (Carl Zeiss).

Results

I-SceI meganuclease is an efficient means of transgenesis in *S. kowalevskii*

To test the efficiency and mosaicism of I-SceI-mediated transgensis in *S. kowalevskii,* we sought to create a transgene capable of driving ubiquitous expression in the developing embryo. Translation elongation factor 1-alpha (EF1α) is an essential component of the ribosome-associated translational machinery in all eukaryotic cells, and its promoter element has been used to drive ubiquitous transgene expression in diverse organisms ranging from plants to vertebrates (Curie et al., 1993; Krieg et al., 1989). We identified an EF1α ortholog in the *S. kowalevskii* genome. To aid in the identification of candidate promoters and enhancers, we performed a ChIP-seq experiment using an antibody to detect Histone 3 Lysine 27 acetylation (H3K27ac), a conserved marker of active chromatin (Mundade et al., 2014; Schwaiger et al., 2014). From this dataset we identified a 1.2-kb fragment upstream of its translational start site enriched for active chromatin marks during embryonic stages (Supplemental Fig. 1A). This region was amplified from genomic DNA and cloned into the I-SceI plasmid (Figure 1A).

When microinjected into fertilized eggs, the EF1α:GFP construct produced transgenic embryos with high efficiency (92%; Figure 1B, Table 1, Supplemental Movie 1). No embryos with completely ubiquitous GFP expression were observed, but rather all embryos displayed a range of mosaic expression patterns. The early cleavage patterns of *S. kowalevskii* can be used to determine the timing of plasmid integration (Figure 1C). The first division defines the left-right axis whereas the second defines the dorsal-ventral axis; therefore, integration of the plasmid at the one, two, or four-cell stage should lead to expression in the whole, one half, or one quadrant of the embryo, respectively. As predicted, the expression patterns of EF1α:GFP could largely be sorted into three categories based on quadrants of GFP expression: mosaic expression restricted to one quadrant of the embryo (Figure 1Di), restricted to one half (Figure 1Dii), or throughout the whole embryo (Figure 1Diii) . GFP expression remained persistent through development, and strong GFP expression was maintained past hatching into juvenile stages (Figure 1E-I). At 7 days post-fertilization, mosaic GFP expression was observed to continue in a number of adult tissues, including musculature in the proboscis (Figure 1H, Supplemental Movie 2) and the ventral nerve cord (Figure 1I). GFP expression was sustained for the duration of experimentation and observation, up to 21 days post-fertilization.

In some transgenesis systems, such as Bacterial Artificial Chromosome (BAC)-mediated transgenesis (Gong et al., 2002; Jeong et al., 2006), injection of linearized DNA fragments is

sufficient to drive genomic incorporation of transgenes. To test whether the inclusion of I-SceI protein in the injection mixture is necessary for transgenesis, the EF1α:GFP construct was linearized, purified, and microinjected into fertilized eggs without supplemental enzyme. The linearized construct alone was capable of driving GFP expression, but the efficiency of transgenesis was much lower and the level of mosaicism was higher compared to injections containing I-SceI protein, thus illustrating the efficiency and efficacy of the enzyme in creating transgenic *S. kowalevskii* embryos (17%; Table 1). We next tested the PCR-amplified reporter fragment containing EF1α:GFP and found that it was capable of driving expression, but with much lower efficiency and higher mosaicism (4%; Table 1). Finally, we sought to test the necessity of preincubating the I-SceI reaction for 40 minutes at 37°C prior to injection. Injection of the EF1α:GFP plasmid and I-SceI without preincubation again resulted in a much lower efficiency and high mosaicism (3%; Table 1). Together these results show the necessity of preincubating the I-SceI reaction prior to injection in order to efficiently drive GFP expression in *S. kowalevskii* embryos. A second ubiquitous expression construct using the *S. kowalevskii* Actin-1 gene was also tested, but was observed to be less efficient and more mosaic than EF1α, and therefore was not further characterized (Table 1, Supp. Figure 1B).

To further assess the utility of I-SceI transgenesis in *S. kowalevskii*, we sought to test whether the system could be used to drive spatiotemporally restricted gene expression without excessive ectopic expression. To this end, we selected two well-characterized transcription factor genes with highly regionalized expression patterns along the anteroposterior axis, Engrailed and Hox11/13c (Lowe et al., 2003). Upstream regions of these genes containing one or more putative enhancers and the native promoter were identified using the ChIP-seq dataset, and cloned into the I-SceI vector. Both the Engrailed:GFP and Hox11/13c:GFP constructs drove GFP expression with high efficiency (52% and 62%, respectively; Table 1), and successfully recapitulated the expression patterns detectable by *in situ* hybridization (Fig. 2, Supplemental Movies 3-6). The GFP expression observed with the Hox11/13c:GFP construct (Fig. 2D) extended slightly more anteriorly than the endogenous expression pattern detected (Fig. 2C), but was generally in the appropriate position along the anteroposterior axis (far posterior). Further, although mosaic, little ectopic GFP expression for both the Engrailed:GFP and Hox11/13c:GFP constructs was detected outside of the expected domains, thus showing I-SceI provides a simple and effective tool to produce accurate gene expression in *S. kowalevskii*.

I-SceI can be used to label individual cell types

We next tested the efficacy of I-SceI-mediated transgenesis for labelling specific cell types. A cell type of interest in *S. kowalevskii* are neurons, due to longstanding questions about the similarities between hemichordate and chordate nervous systems (Lowe et al., 2015). The ability to label individual neurons or specific neuronal subtypes in *S. kowalevskii* would increase our capacity to study the organization of the hemichordate nervous system. As a general proof-of-concept experiment, we chose a gene that could likely serve as a general neuronal marker, the structural neural gene Synapsin. Synapsins are a family of neuronal phosphoproteins that function in synapse formation and plasticity and play a role in the regulation of synaptic transmissions (Cesca et al., 2010). Synapsins are expressed in both the central and peripheral nervous system and are generally absent or expressed at very low levels in non-neuronal cells (De Camilli et al., 1983a; De Camilli et al., 1983b; De Camilli et al., 1979). To this end, we cloned an 8kb region upstream of the *S. kowalevskii* Synapsin ortholog into the I-SceI plasmid to test the hypothesis that the regulatory information within this upstream region should drive GFP expression specifically in neurons.

After the onset of neurogenesis, the Synapsin:GFP construct produced strong GFP expression in neuronal cell bodies and axonal projections in developing embryos at 72 hrs (Figure 3). The high level of GFP expression, in combination with the inherent mosaicism of I-SceI transgenesis, was sufficient to allow for the observation of individual neurons and their processes, which was previously not possible using either *in situ* hybridization or immunofluorescence with common neurotransmitter antibodies. Using this technique, we were able to trace axonal projections of individual neurons projecting between distant body regions, such as from cell bodies in the posterior tip of the trunk across the collar and into the proboscis (Figure 3A; arrows). Further, it was possible to observe the detailed morphology of individual neural cell bodies, which will allow for characterization of neuronal cell types (Figure 3B).

A minimal promoter from the *S. kowalevskii* Gbx can be used to characterize putative enhancer elements

We next wanted to determine if the I-SceI transgenesis method could be used to test the activity of enhancer elements in *S. kowalevskii*. To this end, it was necessary to identify a minimal promoter with little to no background expression that could potentially interfere with the functional characterization of enhancers. We first tried common vertebrate minimal promoters - the mouse *c-Fos* and zebrafish *hsp70* promoters have proven to be effective for enhancer studies in lamprey and zebrafish; however, both were unsuccessful in driving GFP expression in *S. kowalevskii* (data not shown; (Parker et al., 2014a; Parker et al., 2011; Parker et al., 2014b). We then reasoned that a native *S. kowalevskii* minimal promoter would likely be more effective at driving expression than an exogenous vertebrate promoter.

Utilizing the ChIP-seq dataset, we identified putative minimal promoters from several *S. kowalevskii* genes based on enriched H3K27ac signal proximal to the transcriptional start site, including Gbx, Pax6, and Engrailed (Figure 4A). We also identified a 1.5 kb region enriched for H3K27ac located 2.7 kb upstream of hox11/13c, and within the 7kb region in our validated Hox11/13c:GFP construct, as a potential enhancer of interest for this study (Figure 4B). In order to test this region for enhancer activity, we used three constructs: (1) a negative control containing only the Gbx minimal promoter cloned into the I-SceI plasmid directly upstream of GFP; (2) the complete Hox11/13c:GFP construct with 7kb of upstream sequence as a positive control; and (3) the 1.5kb putative enhancer region cloned upstream of the Gbx minimal promoter (-2.7Hox11/13c-Gbx:GFP) (Figure 4C). The expression patterns of Gbx and Hox11/13c have been well-characterized throughout *S. kowalevskii* development, and are expressed in the anterior and far posterior trunk, respectively, in spatially distinct domains, which we reasoned would allow for unambiguous interpretation of GFP expression resulting from the enhancer-promoter fusion (Figure 2D; (Aronowicz and Lowe, 2006; Freeman et al., 2012; Lowe et al., 2003). When injected into fertilized eggs, the minimal promoter control construct, Gbx:GFP, produces no expression (Figure 4E). As previously observed, the Hox11/13c:GFP construct produced far-posterior expression beginning in the gastrula stage and persisting past hatching in a pattern recapitulating the expression observed in previous *in situ* studies (Figure 4F). The enhancer construct, -2.7Hox11/13c-Gbx:GFP, produced GFP expression closely matching the pattern observed from the full 7-kb fragment containing the endogenous Hox11/13c promoter (Figure 4G, Supplemental Movie 7). With the success of testing the hox11/13c and hedgehog enhancer elements from *S. kowalevskii* as well as sonic hedgehog

enhancers from the mouse (Yao et al., 2016) with the same Gbx promoter, it is likely that this minimal promoter is general enough to be used in future enhancer tests in *S. kowalevskii*. An initial screen utilizing similar enhancer:promoter fusions between the -2.7Hox11/13c element and the minimal promoters identified upstream of Pax6 or Engrailed fusion constructs failed to produce any GFP expression, and therefore these promoters were not characterized further (data not shown).

Discussion

Here we report an I-SceI meganuclease approach to create transgenic *S. kowalevskii* embryos. The high efficiency of this protocol enables, for the first time in hemichordates, the functional characterization of putative non-coding regulatory elements in the genome. We demonstrate the feasibility of following gene expression in live embryos by cloning large upstream regions of genomic DNA containing regulatory information and the native promoter into a I-SceI plasmid. This protocol also enables the labeling of individual cell types, such as neurons, which will allow for a better understanding of the organization and diversity of neuronal subtypes that make up the *S. kowalevskii* nervous system. As additional cell-specific expression and single cell sequencing data becomes available, this technique will greatly improve our ability to explore cell type evolution within deuterostomes. Furthermore, the Gbx minimal promoter can be used to test enhancer activity in *S. kowalevskii* using both endogenous elements and enhancers from other species, making it a powerful tool for studying enhancer evolution.

The observation that GFP expression is seen predominantly to be either ubiquitous, or to be restricted to halves or quadrants of embryos developing from EF1α:GFP-injected zygotes provides insight as to when the I-SceI construct is integrating into the genome. This range of expression is likely explained by how the first two cell divisions establish the body plan of *S. kowalevskii*: the first cell division defines the left-right axis of the embryo, whereas the second defines the dorsoventral axis (Figure 1C, after (Colwin and Colwin, 1953)). Therefore, plasmid integration at the one-cell stage would lead to expression throughout the embryo, two-cell stage in half of the embryo, and four-cell stage in one quadrant of the embryo. By screening EF1α:GFP transgenic embryos, we found that the majority of I-SceI-mediated transgene integration likely occurs before first cleavage (Figure 1D).

Despite the apparent early integration, all transgenic vectors showed variable signs of mosaicism, which has also been seen in I-SceI-mediated transgenesis in other systems (Ogino et al., 2006; Parker et al., 2014b; Renfer and Technau, 2017). In certain scenarios, such as labeling individual neurons, mosaicism may be advantageous: observing the complete morphology of a subset of cells is more informative than observing an entire population of labeled cells in which substructures and features of individual cells are obscured by one another. However, mosaicism can also give an incomplete picture of the expression domain of a gene, or of the full activity of an enhancer. Mosaic transgene expression likely results from integration at a later developmental stage. Further work will be necessary to more accurately determine the timing and genomic location of transgene integration using the I-SceI system. However, this work demonstrates that integration occurs early enough in development for this technique to serve as an effective means of transgenesis in *S. kowalevskii*.

Beyond mosaicism, some constructs also produced slight ectopic expression, or expression that did not completely mimic the endogenous gene expression pattern observed by *in situ* hybridization. A challenge with the I-SceI system is the inability to control the genomic location of transgene integration. The genomic environment surrounding the site of integration greatly influences transgene expression, possibly due to activity of peripheral endogenous regulatory elements (Liu, 2013). Furthermore, in cases where only a short upstream fragment of genomic DNA was used in a reporter construct, it is possible that not all regulatory elements necessary to drive the complete expression pattern were included. A potential solution to address this concern is to apply the I-SceI system to BACs: a BAC containing the entire genomic locus surrounding a gene of interest is more likely to possess all of the regulatory information necessary to accurately recapitulate the endogenous expression pattern. Also, due to their large size, BACs may shield the transgene from unwanted positional effects at the site of integration, such as adjacent transcriptional activators or repressors. Due to the low efficiency of transgenesis seen when injecting linearized plasmids in *S. kowalevskii* without I-SceI enzyme, it may be necessary to incorporate I-SceI sites into the BAC vector to increase the likelihood of integration into the genome.

Finally, the choice of minimal promoter used when screening putative enhancers may be an important consideration for achieving accurate and robust reporter expression. When testing the -2.7Hox11/13c element using the Gbx minimal promoter, we observed more mosaicism than

was seen with the complete 7 kb Hox11/13c:GFP construct, which may suggest that the Gbx minimal promoter is not as efficient as the native Hox11/13c promoter. However, we observed no ectopic expression outside of the expected posterior domain in these experiments, which suggests that this minimal promoter may not be a source of 'leaky' ectopic expression. Thus, the Gbx minimal promoter is an excellent candidate for future studies of enhancer activity.

Previous work has shown that *S. kowalevskii* is a promising model for understanding deuterostome evolution and chordate origins (Aronowicz and Lowe, 2006; Darras et al., 2018; Fritzenwanker et al., 2014; Gillis et al., 2012; Green et al., 2013; Lowe et al., 2006; Lowe et al., 2003; Pani et al., 2012). However, studies of gene expression in this species were previously limited to *in situ* hybridization, with no means of assessing the contribution of non-coding elements to gene regulation. Initial studies of *cis*-regulation in *S. kowalevskii* have shown that an enhancer that is necessary for expression of Sonic Hedgehog during vertebrate brain development predates the origin of chordates (Table 1; (Epstein et al., 1999; Kiecker and Lumsden, 2004; Vieira and Martinez, 2006; Yao et al., 2016), demonstrating that there is much that can be learned about enhancer evolution from cross-phyla comparisons between hemichordates and other deuterostomes. The continued development of I-SceI transgenesis will increase of the ease and feasibility of further studies of *cis*-regulatory evolution in *S. kowalevskii*, and will greatly expand the range of experimental possibilities using this emerging model system.

Acknowledgements

The authors would like to thank Robb Krumlauf and Hugo Parker for providing the original lamprey construct as well as providing countless helpful comments and suggestions on transgenesis. Joanna Wysocka and her lab for assistance in developing the ChIP seq protocol. We would also like to thank Jim McIlvain and Zeiss for generous help with microscopy equipment, the staff of the Marine Biological Laboratory and Kevin Uhlinger for support during the *Saccoglossus* field seasons, and the Waquoit Bay Reserve for their assistance. Research reported in this publication was supported by the National Institute of Neurological Disorders and Stroke of the National Institutes of Health under award number 1F32NS098658-01A1 awarded to PJM, the Whitman Center Early Career Investigator Award from the Marine Biological Laboratory awarded to PJM, the National Science Foundation Graduate Research

Fellowship (DGE-114747 to DNC), the National Institute of Child Health and Human Development of the National Institutes of Health under award number 5R01HD073104-07 awarded to Marc Kirschner and providing funding for JG, a National Science Foundation Grant IOS Neural Systems grant 1656628 to CJL.

References

Aronowicz, J. and Lowe, C. J. (2006). Hox gene expression in the hemichordate Saccoglossus kowalevskii and the evolution of deuterostome nervous systems. *Integr Comp Biol* **46**, 890-901.

Buecker, C., Srinivasan, R., Wu, Z., Calo, E., Acampora, D., Faial, T., Simeone, A., Tan, M., Swigut, T. and Wysocka, J. (2014). Reorganization of enhancer patterns in transition from naive to primed pluripotency. *Cell Stem Cell* **14**, 838-853.

Cannon, J. T., Kocot, K. M., Waits, D. S., Weese, D. A., Swalla, B. J., Santos, S. R. and Halanych, K. M. (2014). Phylogenomic resolution of the hemichordate and echinoderm clade. *Curr Biol* **24**, 2827-2832.

Casco-Robles, M. M., Yamada, S., Miura, T. and Chiba, C. (2010). Simple and efficient transgenesis with I-SceI meganuclease in the newt, Cynops pyrrhogaster. *Dev Dyn* **239**, 3275-3284.

Cesca, F., Baldelli, P., Valtorta, F. and Benfenati, F. (2010). The synapsins: key actors of synapse function and plasticity. *Prog Neurobiol* **91**, 313-348.

Colwin, A. L. and Colwin, L. H. (1953). The normal embryology of Saccoglossus kowalevskii (Enteropneusta). *Journal of Morphology* **92**, 401-453.

Curie, C., Axelos, M., Bardet, C., Atanassova, R., Chaubet, N. and Lescure, B. (1993). Modular organization and development activity of an Arabidopsis thaliana EF-1 alpha gene promoter. *Mol Gen Genet* **238**, 428-436.

Darras, S., Fritzenwanker, J. H., Uhlinger, K. R., Farrelly, E., Pani, A. M., Hurley, I. A., Norris, R. P., Osovitz, M., Terasaki, M., Wu, M., et al. (2018). Anteroposterior axis patterning by early canonical Wnt signaling during hemichordate development. *PLoS Biol* **16**, e2003698.

Darras, S., Gerhart, J., Terasaki, M., Kirschner, M. and Lowe, C. J. (2011). beta-catenin specifies the endomesoderm and defines the posterior organizer of the hemichordate Saccoglossus kowalevskii. *Development* **138**, 959-970.

De Camilli, P., Cameron, R. and Greengard, P. (1983a). Synapsin I (protein I), a nerve terminal-specific phosphoprotein. I. Its general distribution in synapses of the central and peripheral nervous system demonstrated by immunofluorescence in frozen and plastic sections. *The Journal of cell biology* **96**, 1337-1354.

De Camilli, P., Harris, S., Huttner, W. and Greengard, P. (1983b). Synapsin I (Protein I), a nerve terminal-specific phosphoprotein. II. Its specific association with synaptic vesicles demonstrated by immunocytochemistry in agarose-embedded synaptosomes. *The Journal of cell biology* **96**, 1355-1373.

De Camilli, P., Ueda, T., Bloom, F., Battenberg, E. and Greengard, P. (1979). Widespread distribution of protein I in the central and peripheral nervous systems. *Proceedings of the National Academy of Sciences* **76**, 5977-5981.

Epstein, D. J., McMahon, A. P. and Joyner, A. L. (1999). Regionalization of Sonic hedgehog transcription along the anteroposterior axis of the mouse central nervous system is regulated by Hnf3-dependent and -independent mechanisms. *Development* **126**, 281-292.

Freeman, R., Ikuta, T., Wu, M., Koyanagi, R., Kawashima, T., Tagawa, K., Humphreys, T., Fang, G. C., Fujiyama, A., Saiga, H., et al. (2012). Identical genomic organization of two hemichordate hox clusters. *Curr Biol* **22**, 2053-2058.

Fritzenwanker, J. H., Gerhart, J., Freeman, R. M., Jr. and Lowe, C. J. (2014). The Fox/Forkhead transcription factor family of the hemichordate Saccoglossus kowalevskii. *Evodevo* **5**, 17.

Gibson, D. G., Young, L., Chuang, R. Y., Venter, J. C., Hutchison, C. A., 3rd and Smith, H. O. (2009). Enzymatic assembly of DNA molecules up to several hundred kilobases. *Nat Methods* **6**, 343-345.

Gillis, J. A., Fritzenwanker, J. H. and Lowe, C. J. (2012). A stem-deuterostome origin of the vertebrate pharyngeal transcriptional network. *Proc Biol Sci* **279**, 237-246.

Goecks, J., Nekrutenko, A., Taylor, J. and Galaxy, T. (2010). Galaxy: a comprehensive approach for supporting accessible, reproducible, and transparent computational research in the life sciences. *Genome Biol* **11**, R86.

Gong, S., Yang, X. W., Li, C. and Heintz, N. (2002). Highly efficient modification of bacterial artificial chromosomes (BACs) using novel shuttle vectors containing the R6Kgamma origin of replication. *Genome Res* **12**, 1992-1998.

Grabher, C., Joly, J. S. and Wittbrodt, J. (2004). Highly efficient zebrafish transgenesis mediated by the meganuclease I-SceI. *Methods Cell Biol* **77**, 381-401.

Green, S. A., Norris, R. P., Terasaki, M. and Lowe, C. J. (2013). FGF signaling induces mesoderm in the hemichordate Saccoglossus kowalevskii. *Development* **140**, 1024-1033.

Jacquier, A. and Dujon, B. (1985). An intron-encoded protein is active in a gene conversion process that spreads an intron into a mitochondrial gene. *Cell* **41**, 383-394.

Jeong, Y., El-Jaick, K., Roessler, E., Muenke, M. and Epstein, D. J. (2006). A functional screen for sonic hedgehog regulatory elements across a 1 Mb interval identifies long-range ventral forebrain enhancers. *Development* **133**, 761-772.

Kaul, S. and Stach, T. (2010). Ontogeny of the collar cord: neurulation in the hemichordate Saccoglossus kowalevskii. *J Morphol* **271**, 1240-1259.

Kaul-Strehlow, S., Urata, M., Minokawa, T., Stach, T. and Wanninger, A. (2015). Neurogenesis in directly and indirectly developing enteropneusts: of nets and cords. *Organisms Diversity & Evolution* **15**, 405-422.

Kiecker, C. and Lumsden, A. (2004). Hedgehog signaling from the ZLI regulates diencephalic regional identity. *Nat Neurosci* **7**, 1242-1249.

Knight-Jones, E. (1952). On the nervous system of Saccoglossus cambrensis (Enteropneusta). *Phil. Trans. R. Soc. Lond. B* **236**, 315-354.

Krieg, P. A., Varnum, S. M., Wormington, W. M. and Melton, D. A. (1989). The mRNA encoding elongation factor 1-alpha (EF-1 alpha) is a major transcript at the midblastula transition in Xenopus. *Dev Biol* **133**, 93-100.

Langmead, B. and Salzberg, S. L. (2012). Fast gapped-read alignment with Bowtie 2. *Nat Methods* **9**, 357-359.

Liu, C. (2013). Strategies for designing transgenic DNA constructs. *Methods Mol Biol* **1027**, 183-201.

Lowe, C. J., Clarke, D. N., Medeiros, D. M., Rokhsar, D. S. and Gerhart, J. (2015). The deuterostome context of chordate origins. *Nature* **520**, 456-465.

Lowe, C. J., Tagawa, K., Humphreys, T., Kirschner, M. and Gerhart, J. (2004). Hemichordate embryos: procurement, culture, and basic methods. *Methods Cell Biol* **74**, 171-194.

Lowe, C. J., Terasaki, M., Wu, M., Freeman, R. M., Jr., Runft, L., Kwan, K., Haigo, S., Aronowicz, J., Lander, E., Gruber, C., et al. (2006). Dorsoventral patterning in hemichordates: insights into early chordate evolution. *PLoS Biol* **4**, e291.

Lowe, C. J., Wu, M., Salic, A., Evans, L., Lander, E., Stange-Thomann, N., Gruber, C. E., Gerhart, J. and Kirschner, M. (2003). Anteroposterior patterning in hemichordates and the origins of the chordate nervous system. *Cell* **113**, 853-865.

Mundade, R., Ozer, H. G., Wei, H., Prabhu, L. and Lu, T. (2014). Role of ChIP-seq in the discovery of transcription factor binding sites, differential gene regulation mechanism, epigenetic marks and beyond. *Cell Cycle* **13**, 2847-2852.

Neph, S., Vierstra, J., Stergachis, A. B., Reynolds, A. P., Haugen, E., Vernot, B., Thurman, R. E., John, S., Sandstrom, R., Johnson, A. K., et al. (2012). An expansive human regulatory lexicon encoded in transcription factor footprints. *Nature* **489**, 83-90.

Ochiai, H., Sakamoto, N., Suzuki, K., Akasaka, K. and Yamamoto, T. (2008). The Ars insulator facilitates I-SceI meganuclease-mediated transgenesis in the sea urchin embryo. *Dev Dyn* **237**, 2475-2482.

Ogino, H., McConnell, W. B. and Grainger, R. M. (2006). High-throughput transgenesis in Xenopus using I-SceI meganuclease. *Nat Protoc* **1**, 1703-1710.

Pani, A. M., Mullarkey, E. E., Aronowicz, J., Assimacopoulos, S., Grove, E. A. and Lowe, C. J. (2012). Ancient deuterostome origins of vertebrate brain signalling centres. *Nature* **483**, 289-294.

Parker, H. J., Bronner, M. E. and Krumlauf, R. (2014a). A Hox regulatory network of hindbrain segmentation is conserved to the base of vertebrates. *Nature* **514**, 490-493.

Parker, H. J., Piccinelli, P., Sauka-Spengler, T., Bronner, M. and Elgar, G. (2011). Ancient Pbx-Hox signatures define hundreds of vertebrate developmental enhancers. *BMC Genomics* **12**, 637.

Parker, H. J., Sauka-Spengler, T., Bronner, M. and Elgar, G. (2014b). A reporter assay in lamprey embryos reveals both functional conservation and elaboration of vertebrate enhancers. *PLoS One* **9**, e85492.

Rada-Iglesias, A., Bajpai, R., Swigut, T., Brugmann, S. A., Flynn, R. A. and Wysocka, J. (2011). A unique chromatin signature uncovers early developmental enhancers in humans. *Nature* **470**, 279-283.

Renfer, E. and Technau, U. (2017). Meganuclease-assisted generation of stable transgenics in the sea anemone Nematostella vectensis. *Nat Protoc* **12**, 1844-1854.

Schwaiger, M., Schonauer, A., Rendeiro, A. F., Pribitzer, C., Schauer, A., Gilles, A. F., Schinko, J. B., Renfer, E., Fredman, D. and Technau, U. (2014). Evolutionary conservation of the eumetazoan gene regulatory landscape. *Genome Res* **24**, 639-650.

Simakov, O., Kawashima, T., Marletaz, F., Jenkins, J., Koyanagi, R., Mitros, T., Hisata, K., Bredeson, J., Shoguchi, E., Gyoja, F., et al. (2015). Hemichordate genomes and deuterostome origins. *Nature* **527**, 459-465.

Thermes, V., Grabher, C., Ristoratore, F., Bourrat, F., Choulika, A., Wittbrodt, J. and Joly, J. S. (2002). I-SceI meganuclease mediates highly efficient transgenesis in fish. *Mech Dev* **118**, 91-98.

Vieira, C. and Martinez, S. (2006). Sonic hedgehog from the basal plate and the zona limitans intrathalamica exhibits differential activity on diencephalic molecular regionalization and nuclear structure. *Neuroscience* **143**, 129-140.

Yao, Y., Minor, P. J., Zhao, Y. T., Jeong, Y., Pani, A. M., King, A. N., Symmons, O., Gan, L., Cardoso, W. V., Spitz, F., et al. (2016). Cis-regulatory architecture of a brain signaling center predates the origin of chordates. *Nat Genet* **48**, 575-580.

Figure legends

Figure 1. Expression of EF1α in *S. kowalevskii*

(A) Schematic of the EF1α transgenesis plasmid. Approximately 1200 bp upstream of the *S. kowalevskii* EF1α gene was amplified from genomic DNA and cloned into the I-SceI plasmid. (B) A group of mid-gastrulae screened for positive GFP expression. (C) Early cleavage patterns of *S. kowalevskii* can be used to determine the timing of plasmid integration. The first cleavage defines the left and right axis of the embryo and the second defines the dorsal and ventral axis. Integration of the EF1α plasmid at the one, two, or four-cell stage should drive GFP expression ubiquitously, in half, or in one quadrant, of the developing embryo, respectively. (D) Expression of EF1α in mid-gastrulae. The left images in D^{i}, D^{ii}, and D^{iii} show GFP in individual embryos expressed in one quadrant, one half, and throughout the embryo, respectively. Rhodamine was used as a tracer to follow successfully injected embryos. Merged GFP and rhodamine images are shown on the left along with the number of GFP expressing embryos that can be placed into each category. Integration of the EF1α construct occurs at the one-cell stage in the majority of GFP positive embryos. (E) Expression of EF1α in kink stage embryos. Strong GFP expression is seen throughout development of *S. kowalevskii*. (F) Schematic illustrating the tripartite body plan of the 3 gill slit stage *S. kowalevskii*. p, proboscis; c, collar; t, trunk; m, mouth; and g, gill slit. (G) GFP expression is maintained post hatching. (H) GFP expression shown in the proboscis and collar. (I) GFP expression in the trunk and post-anal tail of a hatched animal.

Figure 2. Reporter constructs for genes with spatiotemporally restricted expression patterns recapitulate *in situ* hybridization results

(A) *In situ* hybridizations for Engrailed at the kink stage, 72hpf. (B) GFP expression in *Engrailed*:GFP-injected embryos at the same stage as in (A) recapitulating the *in situ* hybridization expression pattern. (C) *In situ* hybridizations for Hox11/13c at the kink stage, 72hpf. (D) *Hox11/13c*:GFP expression at the same stage.

Figure 3. Expression of Synapsin:GFP in the juvenile illustrates the possibility of labeling specific cell types with I-SceI meganuclease in *S. kowalevskii*

(A) Synapsin:GFP expression in the juvenile worm. Expression is seen in neuronal cell bodies in the far posterior of the worm, shown by the black arrowhead, with axon trajectories stretching to

the proboscis, shown by the white arrowhead. This plasmid confirms that it is possible to label individual cell types in *S. kowalevskii*, in this case neurons. The mosaicism of the I-SceI method makes it easier to track individual neurons in the animal. (B) 40x magnification of neuronal cell bodies and axon projections found in the proboscis.

Figure 4. The Gbx minimal promoter can be used to test enhancer activity in *S. kowalevskii*

(A)The Gbx minimal promoter, as identified from ChIP-seq data, is a 222 bp fragment located directly 5' of the Gbx start site. (B) The ChIP-seq dataset for the 7 kb region directly upstream of the Hox11/13c coding sequence. A putative Hox11/13c enhancer is found approximately 1 kb upstream of the start site. (C) The cloning strategy to test the efficacy of the putative Gbx minimal promoter and Hox11/13c enhancer element. Three different plasmids were created: (1) the putative Gbx minimal promoter cloned immediately upstream of the GFP start site, (2) the entire 7 kb region directly upstream of Hox11/13c cloned immediately upstream of the GFP start site, and (3) the putative Hox11/13c enhancer and cloned upstream of the Gbx minimal promoter. (D) Schematics of Gbx and Hox11/13c expression from *in situ* data during *S. kowalevskii* development. (E) Injection of the Gbx:GFP plasmid yields no expression. (F) Three different embryos (shown at mid-gastula 24 hpf ($4F^{i}$) and kink stage 72 hpf ($4F^{ii-iii}$) that have been injected with the Hox11/13c:GFP plasmid. GFP expression recapitulates *in situ* data. Expression is only seen in the far posterior of the developing embryo. (G) Expression patterns of the -2.7Hox11/13c-Gbx:GFP plasmid in the far posterior of the embryo throughout development (shown at gumedo 36 hpf (G^{i}), double-groove 48 hpf (G^{ii}), and a posterior view of a double-groove embryo at 48 hpf (G^{iii}).

Table 1. I-SceI Plasmids mediate transgenesis in *S. kowalevskii*

Plasmids generated and microinjected into *S. kowalevskii*. The columns from left to right reflect the presence or absence of the I-SceI meganuclease enzyme, name of the construct generated, enhancer element present, native or ectopic promoter used, penetrance of GFP expression, the number of viable embryos successfully injected, and the corresponding figure number for each construct.

Supplemental Movie Legend

Supplemental Movie 1. EF1α expression in a pre-hatching embryo

Supplemental Movie 2. EF1α expression in a hatched embryo. The movie focuses on the proboscis and collar region of the animal.

Supplemental Movie 3. Embryos injected with the Engrailed:GFP construct. Kink stage embryos display a circumferential band of expression posterior to the collar as reported with *in situ* data (Lowe et al., 2003).

Supplemental Movie 4. Embryos injected with the Engrailed:GFP construct. Multiple groove stage embryos show that despite mosaicism little ectopic expression is found in the developing embryos.

Supplemental Movie 5. Hox11/13c-Gbx:GFP expression at gastrula stage. Expression is seen around the blastopore of the embryos.

Supplemental Movie 6. Hox11/13c-Gbx:GFP expression at gastrula stage of one embryo. Circumferential expression is seen around the blastopore of one embryo.

Supplemental Movie 7. 2.7Hox11/13c-Gbx:GFP expression in pre-hatchlings. Expression is seen in the posterior end of the embryos similar to *in situ* data (Lowe et al., 2003).

Supplemental Table 1. Primers used for the construction of all I-SceI plasmids generated. Red font corresponds to Gibson assembly overhands and black font is used to show base pairs corresponding to regions of *S. kowalevskii* genomic DNA.

Supplemental Figure 1. H3K27ac profiles of the genomic regions of additional genes examined in this study: (A-D) The promoter regions isolated from *S. kowalevskii* EF1α, Engrailed, Actin1, and Synapsin, respectively, for testing in the I-SceI construct are indicated in red. (Ci) Representative micrograph of transgenic embryos produced by microinjection with the Actin1 reporter construct; micrographs corresponding to the other constructs are in Figures 1 - 3.

Figure 1

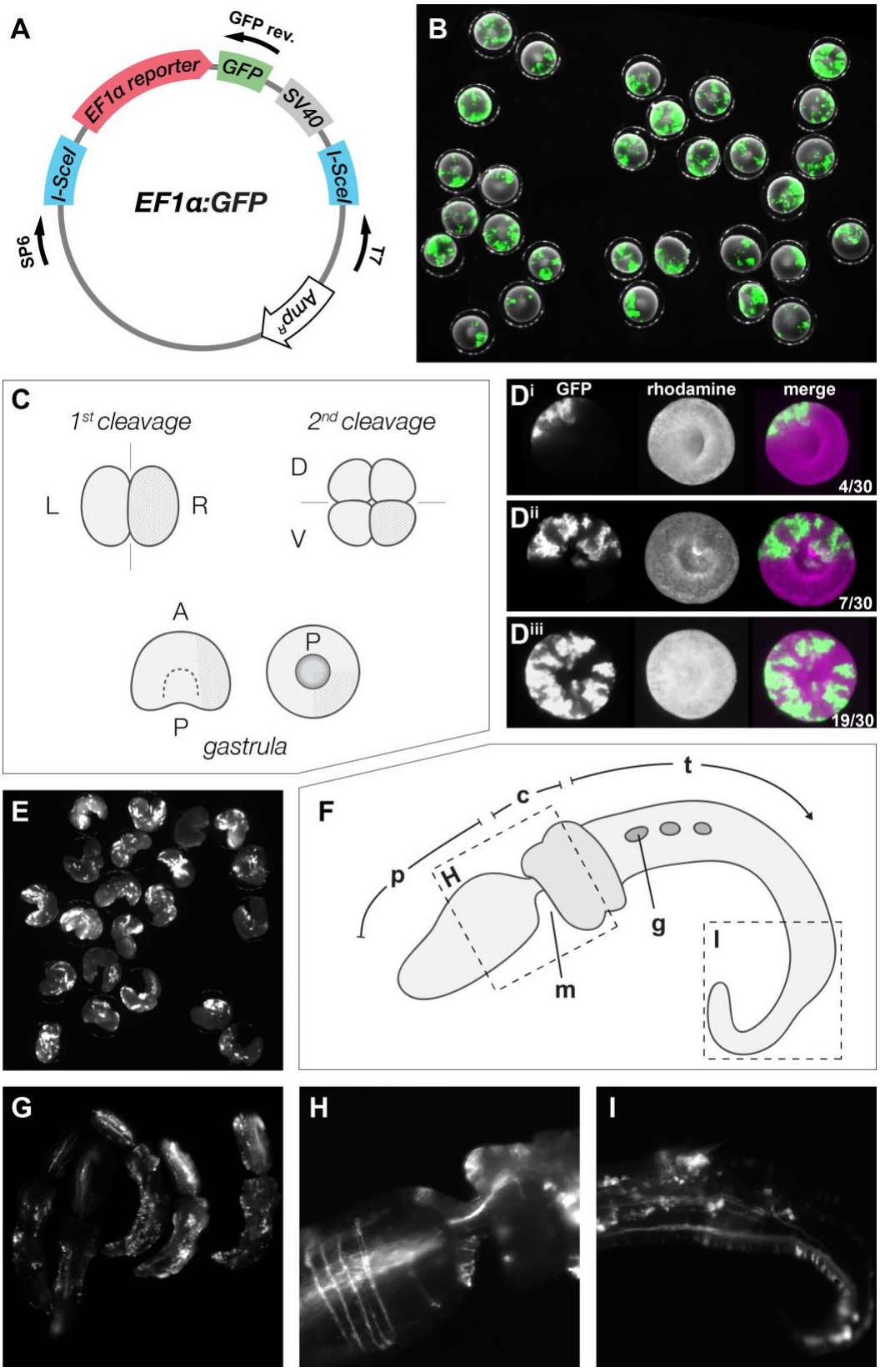

A
GFP rev.
EF1α reporter
GFP
SV40
I-SceI
I-SceI
EF1α:GFP
SP6
T7
AmpR
B
C
1st cleavage
2nd cleavage
L
R
D
V
A
P
P
gastrula
Di
GFP
rhodamine
merge
4/30
Dii
7/30
Diii
19/30
E
F
t
c
p
H
g
l
m
G
H
I

Figure 2

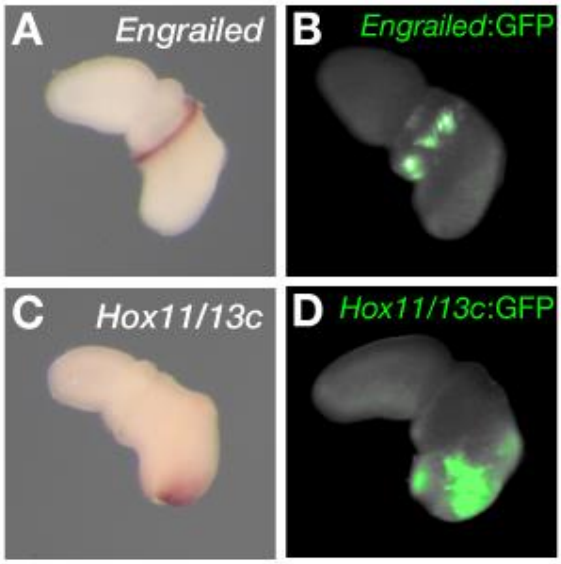


Figure 3

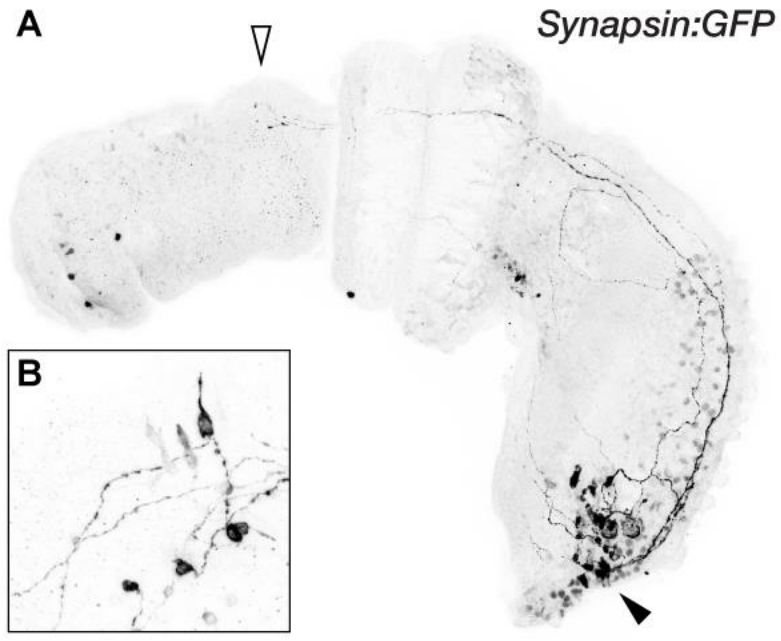


Figure 4

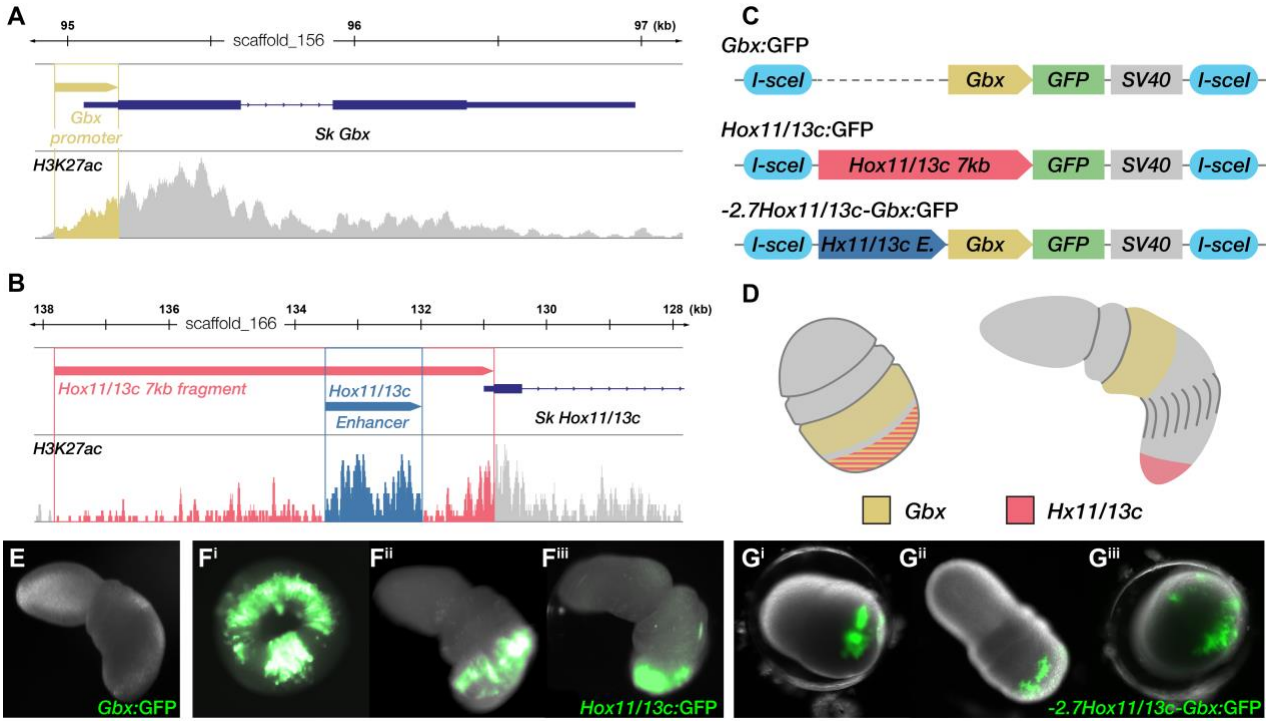
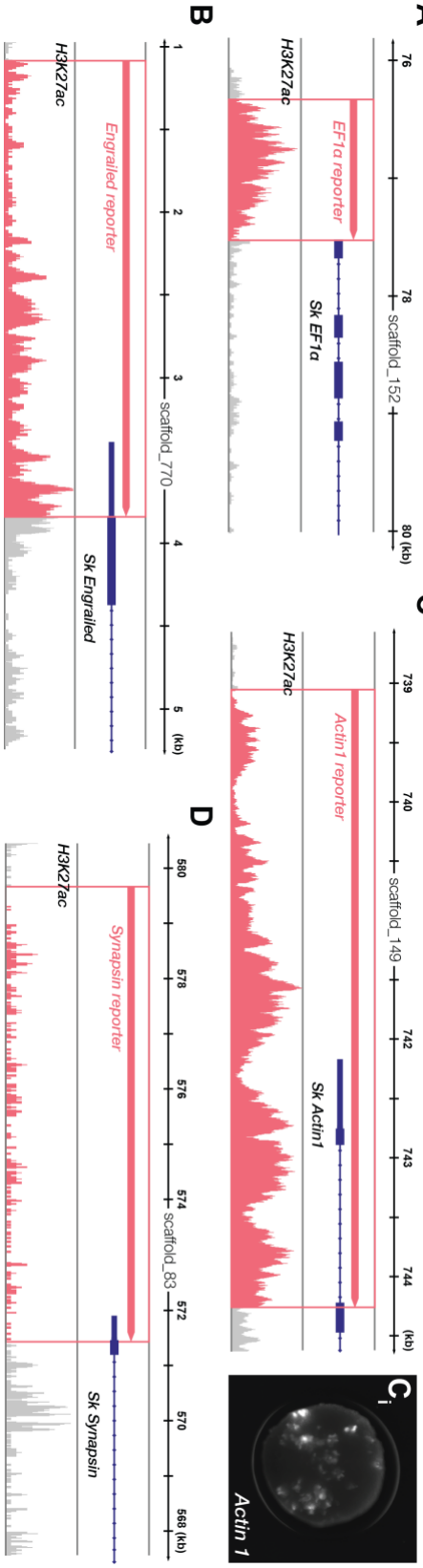

A
95
scaffold_156
96
97 (kb)
Gbx promoter
Sk Gbx
H3K27ac
B
138
136
scaffold_166
134
132
130
128 (kb)
Hox11/13c 7kb fragment
Hox11/13c Enhancer
Sk Hox11/13c
H3K27ac
C
Gbx:GFP
I-sceI
Gbx
GFP
SV40
I-sceI
Hox11/13c:GFP
I-sceI
Hox11/13c 7kb
GFP
SV40
I-sceI
-2.7Hox11/13c-Gbx:GFP
I-sceI
Hx11/13c E.
Gbx
GFP
SV40
I-sceI
D
Gbx
Hx11/13c
E
Gbx:GFP
Fi
Fii
Fiii
Hox11/13c:GFP
Gi
Gii
Giii
-2.7Hox11/13c-Gbx:GFP

Table 1

I-SceI	Construct	Enhancer Element	Promoter	GFP Expression	Total Injected	Figure
+	pGPS2 NotI-linearized Control Plasmid	-	-	0 (0%)	25	not shown
+	EF1α:GFP	*ef-1 alpha* endogenous upstream region	*ef-1 alpha*	23 (92%)	25	Fig. 1; Supp. Movies 1-2
-	linearized EF1α:GFP	*ef-1 alpha* endogenous upstream region	*ef-1 alpha*	5 (17%)	30	not shown
-	PCR-amplified EF1α:GFP	*ef-1 alpha* endogenous upstream region	*ef-1 alpha*	1 (4%)	27	not shown
+	EF1α:GFP without preincubation	*ef-1 alpha* endogenous upstream region	*ef-1 alpha*	1 (3%)	30	not shown
+	Actin1:GFP	*actin1*	*actin1*	10 (67%)	15	Supp. Fig. 2
+	Engrailed:GFP	*engrailed* endogenous upstream region	*engrailed*	16 (52%)	31	Fig. 2; Supp. Movies 3-4
+	Gbx:GFP	-	*gbx*	0 (0%)	60	Fig. 4E
+	Hox11/13c:GFP	*hox11/13c* endogenous upstream region	*hox11/13c*	18 (62%)	29	Fig. 2, 4F; Supp. Movies 5-6
-	Hox11/13c:GFP	*hox11/13c* endogenous upstream region	*hox11/13c*	0 (0%)	25	not shown
+	-2.7Hox11/13c-Pax6:GFP	*-2.7kb hox11/13c enhancer*	*pax6*	0 (0%)	20	not shown
+	-2.7Hox11/13c-Gbx:GFP	*-2.7kb hox11/13c enhancer*	*gbx*	19 (42%)	45	Fig. 4G; Supp Movie 7
+	-38.6skSBE1-Gbx:mNeonGreen	+38.6kb S.k. hh SBE1 enhancer	*gbx*	6 (10%)	60	Yao et. al 2016
+	-7.5mmSBE1-Gbx:mNeonGreen	+7.5 M.m. shh SBE1 enhancer	*gbx*	5 (9%)	55	
+	-1111mmSBE5-Gbx:mNeonGreen	-1111 M.m. ssh SBE5 enhancer	*gbx*	5 (33%)	15	

Supplemental Figure 1



Supplemental Table 1

Primer Name	Sequence
Sk_EF1a_F	AAGCTTTCTTGTACAAAGTGGTTCGggttgttcatcagaaagccgtg
Sk_EF1a_R	CTCGCCCTTGCTCACCATtctgatttgtttgttttttatcaacttggg
Sk_Actin1_F	AAGCTTTCTTGTACAAAGTGGTTCGcccttaacacagtccctcaagag
Sk_Actin1_R	CTCGCCCTTGCTCACCATgctgtctttctgtcccataccg
Synapsin F	AAGCTTTCTTGTACAAAGTGGTTCGttttgaggttgaattaaagcgggtga
Synapsin R	CTCGCCCTTGCTCACCATacagtacatgcagttaaggtgtgg
HLC F	ATGGTGAGCAAGGGCGAG
HLC R	CGAACCACTTTGTACAAGAAAGCTT
pGBX F	AGCTTTCTTGTACAAAGTGGTTCGataacttatacatgtttttataaacccctatgttgt
pGBX R	CTCGCCCTTGCTCACCATttccacgtttttttttaacaacaagagtact
PHC F	ATGGTGAGCAAGGGCGAG
PHC R	CGAGCTCGAATTCGTCGACC
Sk_SBE1 F	CGAGGTCGACGAATTCGAGCggtaattgctacggcaacta
Sk_SBE1 R	CGATACCGTCGACCTCGAGGagctgtaattgacccaactcggc
Mm_SBE1 F	CGAGGTCGACGAATTCGAGCctcgagatgtgtatttgaattttaaatggc
Mm_SBE1 R	CGATACCGTCGACCTCGAGGgatatcctcccaaatgagcagcgg
Mm_SBE5 F	CGAGGTCGACGAATTCGAGCgattaactgctctgattccaacctg
Mm_SBE5 R	CGATACCGTCGACCTCGAGGtattctagactgtgttcttggctaa
Hox en F	CGAGGTCGACGAATTCGAGCgtcccatatcgaacaggacgactg
Hox en R	CGATACCGTCGACCTCGAGGgacactggattggtttaatttcaactcgtg
Hox upstream F	GGTCGACGAATTCGAGCTCGacgctacttggaatctcaga
Hox upstream R	CTCGCCCTTGCTCACCATggtacgctctcacactgtcg
En upstream F	GGTCGACGAATTCGAGCTCGcaggtcacggaaatgtgtct
En upstream R	CTCGCCCTTGCTCACCATtgtcaagttggattggccag