ANALYSIS OF CRES OF ETS1/2 TARGET GENES IN THE BASAL CHORDATE

CIONA INTESTINALIS

By

YOUNAN LI

____

A Thesis Submitted to the Honors College

In Partial Fulfillment of the Bachelor's degree

With Honors in

Molecular and Cellular Biology

THE UNIVERSITY OF ARIZONA

MAY 2012

Approved by:

____

Dr. Brad Davidson

Department of Molecular and Cellular Biology

Abstract:

Fibroblast growth factor (FGF) signaling and downstream activation of the transcription factor ETS1/2 specify the heart cell lineage in the basal chordate *Ciona intestinalis*. Based on our previous work, 144 candidate FGF/ETS1/2 dependent heart genes have been identified from the *Ciona* genome. The expression of a large subset of these genes has been verified in the heart precursor lineage using in-situ techniques. Using an bio-informatic approach, we identified candidate enhancers for some of these target genes. We then tested these predicted enhancers through reporter analysis. Our results indicate that the Ets1/2 binding site GGAW and the co-motif ATTA are required for heart lineage enhancer activity. This work can help us understand the conserved, primary role of FGF/Ets in chordate heart lineage specification.

Introduction:

Ciona intestinalis is a sea tunicate, which is a simple, invertebrate sea squirt. Despite appearances, *Ciona* is a member of our own phylum, *Chordata*, making it the closest invertebrate relative to the vertebrates (Dehal et al, 2002). In comparison to the complex and multi-layered gene regulatory network of vertebrate heart development, *Ciona* provides a more simplified model to study signal transduction and regulation in early heart development. One key feature of tunicates is that they diverged before the genome duplications that convoluted vertebrate genomes (Davidson, 2007), therefore providing a simpler system to study the function of genes that are redundant in vertebrates. In addition, earlier studies indicate that *Ciona* heart development is morphologically and functionally similar to vertebrate heart development (Davidson, 2007), making this system relevant for translational research. What's more, the *Ciona* genome is condensed, so it has condensed non-coding regions (Di Gregorio and Levine, 2002). Additionally, the *Ciona intestinalis* and *Ciona savignyi* genomes are fully sequenced, which allow us to identify conserved, candidate cis-regulatory elements through phylogenetic foot printing (Johnson et al., 2004).

Previous research indicates that the *Ciona* heart lineage is delineated when founder cells in the B7.5 cell lineage divide asymmetrically, giving rise to heart progenitor cells, also known as Trunk Ventral Cells (TVCs) and tail muscle precursors or Anterior Tail Muscle cells (ATMs) (Davidson, 2007). Previous work showed that FGF signaling is responsible for directing heart cell fate, and this is mediated through the activation of MAPK signaling and the downstream transcription factor Ets1/2. After asymmetric division, the heart precursor cells migrate towards the trunk where they undergo further proliferation and differentiation to form the heart tube (Davidson, 2007).

Fibroblast growth factors (FGFs) are important for development in many organisms ranging from nematodes to humans. Large numbers of studies demonstrate that FGFs play a significant role in cell proliferation, migration and differentiation during embryonic development. In adult organisms, FGF is a homeostatic factor, and it plays an important role in tissue repair and injury response (Ornitz and Itoh, 2001). There are 22 identified FGFs in human, but in *Ciona intestinalis*, only six FGF genes appear to be encoded by the genome (Satou et al., 2003). Previous work showed that the *Ciona FGF9/16/20* homolog is responsible for heart specification (Davidson, 2006). During early heart development, FGF9 is expressed by neighboring mesenchymal cells and binds to FGFR in the B7.5 lineage. According to established models, binding of FGF9 causes receptor trans-phosphorylation and recruitment of GRB2. GRB2 activates RAS leading to activation of the mitogen-activated protein kinase (MAPK) cascade. Phosophorylated MapK (dp-ERK) translocates into the nucleus and phosphorylates the transcription factor Ets1/2. Ets1/2 then binds to and activates many downstream target genes that are important for heart cell fate.

The E-twenty six (ETS) transcriptional factor family is one of the largest transcriptional families, and it's a classical model for combinatorial gene expression. This transcriptional factor family seems to be unique to the metazoan lineage (Sharrocks, 2001). There are currently 29 identified ETS proteins in human and six in *Ciona intestinalis* (Satou, 2003). ETS factors often act as nuclear targets of signal-transduction pathways (Sharrocks, 2001). The ETS DNA binding domain is highly conserved, and the ETS binding site, GGAW, is also highly conserved among species. However, ETS factors can function either as transcriptional repressors or activators. It is thought that the specific regulatory function of ETS is dependent on its interaction with multiple co-regulatory proteins (Sharrocks, 2001). In *Ciona*, ETS factors are expressed in various tissues of the developing embryo. Furthermore, Ets1/2 is expressed in both TVCs and ATMs, but because of the polarized FGF signal, it is only activated in the heart cells

(Davidson et al., 2006). We believe that Ets1/2 specifically regulates the expression of heart gene targets in heart precursor cells via its interaction with a specific co-regulatory protein, which remains to be identified.

Initial studies identified several downstream target genes for Ets1/2, using microarray analysis on sorted B7.5 lineage cells. In these experiments, the B7.5 cell lineage was fluorescently labeled with GFP using a lineage specific driver, Mesp, which targets GFP expression in the B7.5 lineage. Furthermore, surrounding cells were labeled with MyoD-YFP, which made it possible to separate them out. Expression profiles of B7.5 lineage cells isolated from wild type, transgenic Mesp-FGFR dominant-negative, and transgenic Ets dominant-negative embryos were analyzed via microarray analysis and compared. These 3 conditions allowed us to look at the specific contributions of FGF and Ets to gene expression. Expression was analyzed at 8.5 hours post fertilization, or approximately 0.5-1 hours after the onset of FGF signaling. Through this analysis, 144 candidate genes were found to be significantly down regulated in sorted cells from transgenic FGFRdn and EtsWRPW embryos in comparison to wild type embryos. These results suggest that these genes are potential targets of FGF and Ets during heart specification.

To validate our microarray data results, we applied in-situ hybridization for those identified gene targets. Overall, we observed staining for 47 candidate gene antisense probes, 80% of which displayed apparent expression in the TVCs. This result showed that the 8.5H microarray database accurately represented the initial FGF/Ets targets in the newly born TVC lineage.

To further explore how Ets regulates downstream targets during heart specification, we were interested in exploring potential co-factors. However, there are many potential binding partners for Ets factors (Sharrocks, 2001) . So, in lieu of identifying the co-factor directly, we decided to take a simpler approach by identifying potential co-factor binding sites. We predicted that these sites could be readily identified by analyzing cis-regulatory domains of the identified gene targets. We anticipated that the co-factor binding site would be conserved, adjacent to Ets binding sites and located in regions near primary TVC candidate genes.

We applied comparative sequence analysis and whole-genome motif ranking to cis-regulatory elements of *Hand-like* and *FoxF,* two early heart genes which are turned on by ETS, to identify candidate co-factor binding sites sequences. In order to identify potential co-factor binding sequences conserved in the CREs of both of these targets, we analyzed the CREs of *Hand-like* and *FoxF* with search parameters for all possible 4- and 5- nucleotide motifs with N and NN spacers added after the 2nd nucleotide (e.g. AAAA, AANNAA, AANAA, AANNAAA and all combinations of nucleotides). We enriched this search by comparing sequence motifs with a closely related species. *Ciona savinyi* and *Ciona intestinalis* are very close relatives, so we assumed that their heart developments are regulated in the same way. So, they use very similar CREs to regulated early heart genes activation. By comparing *Ciona intestinalis* CREs of both

Hand-like and *FoxF* with the analogous CREs of *Ciona savinyi*, we can more confidently assume that these regions are in fact regulatory regions, due to evolutionary constraint.

44 potential Ets co-factor binding motifs were identified using these search parameters. We quantified the enrichment of these 44 motifs in the predicted primary TVC CREs of the 144 candidate ETS early target genes. Predicted TVC CREs included conserved noncoding regions adjacent to those 144 target genes, with a low nucleosome score, and having Ets binding sites. Low-nucleosome occupancy score is strongly correlated with regulatory function in *Ciona* non-coding sequences (Khoueiry et al, 2010).

Next, based on the same method, we generated a set of predicted CREs for all 15569 *Ciona* gene models. We refer to this as the "background CREs". We compared the enrichment of Ets-motif + co- motif match in predicted TVC CREs and predicted background CREs. This analysis generated a list of "hits" within the "primary TVC" gene set and the background "all gene" set. We ranked those 44 motifs based on the level of their selective concentration around primary TVC candidate genes. An 'ATTA' motif occupied the top rank, indicating that the ATTA motif is selectively enriched in predicted TVC CREs and thus may be the binding site for Ets co-factor.

To test the predicted role of the ATTA motif in early heart gene regulation, we isolated CREs for *Hand-like* and *Foxf* genes, mutated the ATTA motifs, and fused the CREs onto the reporter gene *lacZ*. Reporter expression analysis of these transgenic plasmids showed significantly reduced expression in TVCs. This experiment indicated that the ATTA motif is essential for *Hand-like* and *FoxF* expression in the TVCs .

My project suggested a potent method for identifying CREs controlling early heart gene expression by looking for intergenic regions which involve Ets-binding site and ATTA motif. The Ets-binding site and potential co-factor binding motif ATTA are important to active characterized early heart genes. Therefore we predicted that the Ets-binding site and ATTA motif are also included in other early heart genes CREs. So, I tested whether paired GGAW (Ets binding site) + ATTA motifs designate possible CREs regions for those 144 candidate genes by using a bioinformatics assay.

Methods:

Embryological techniques:

Ciona intestinalis adults were purchased from M-REP. Fertilization, dechorionation and electroporation were performed as previously described (Beh et al., 2007). The embryos were incubated at 18 degrees until the desired stage of development.

Informatics software “chordator”:

We used the informatics software “chordator” (heep://chordator.vze.com/, Haeussler et al., 2010) to search for the conserved ETS/ATTA motif combination adjacent to the primary heart genes. The criteria is to search for two conserved ETS binding sites and one conserved ATTA motif within 50bp window, 10000bp adjacent to candidate genes.

Molecular cloning:

Based on the chordator search result, several possible enhancer regions were cloned into pCESA vector. They included predicted CREs for *RGS3*, *#1897*, *#11094*, and *UNC5*. Sequence specific primers carrying appropriate restriction sites were used to amplify those constructs. The construct included endogenous basal promoter and the start codon of the candidate gene fused in frame with *lacZ*. All the primers used are listed in Figure 3.

Beyond that, the predicted 3' *Hand-like* CREs was cloned into the *Hand-like* enhancer pCESA vector upstream of the strong 297bp basal promoter region of *Hand-like* that shows no independent staining. The *Hand-like* enhancer was replaced with downstream intergenic region of *Hand-like,* which included predicted 3' *Hand-like* CREs, by restriction using the Xba1/Bam1 sites.

Next, in order to boost reporter expression, we cloned possible enhancer regions of *RGS3* and *#1897* upstream of the *Hand-like* basal promoter using the same approach as for predicted 3' *Hand-like* potential enhancer region. *#1839* was cloned into this vector using 5'CCGCATGTCGCGCGAAGTCG3' forward primer and a 3'CGCCGCGGTAGTTAAATTGCTGTG5' reverse primer. A full length *RGS3* was cloned into this vector using the 5'CGATATACGATGCGCCTTTGGGTG3' forward primer and a 3'GACCATATACACTTCTCTAATGC5' reverse primer. To further test the function of the predicted CREs, two shorter regions containing only predicted CREs were cloned. One

125bp long was cloned by 5'CACGAGTACCGAATGTAG3' and the other 101bp long was using a 5'CGAGTGTGGCGTTCAATTAAC3'primer. Both of them were paired with the 3'CGTTTCGTCCGAGGCGCG5' reverse primer. The full length and the 125bp *RGS3* regions containing two predicted putative ATTA binding sites were mutated using the following mutagenesis primer 5'GTAGATAGATCGAGTGTGGCGTTCAATACACAC3' and introduced by PCR.

X-gal and secondary anti-body staining

Embryos for X-gal staining were fixed in MEM-GA (0.1 M MOPS, pH 7.4, 0.5 M NaCl, 2 mM $MgSO_4$, 1 mM EGTA, pH 8.0, 0.2% glutaraldehyde, 0.05% tween-20) for 25-30 minutes at room temperature, and then stained in X-gal as described in (Cooley et al., 2011) . Embryos for secondary anti-body staining were fixed in methanol for about one hour at -20 degrees, and then stained with primary and secondary antibody as described in (Cooley et al., 2011).

Results:

The bioinformatics software identified 64 clusters of predicted CREs adjacent to 33 candidate primary TVCs genes. Five predicted CREs were selected from this list for functional analysis based on their nucleosome occupancy scores and the ease of cloning the predicted CRE along with endogenous promoter and ATG start codon. This set included intergenic regions upstream of genes with predicted orthology to *RGS3* (aniseedV3_8034), two unidentified genes (aniseedV3_1839) and (aniseedV3_11094), and *UNC5*. The last predicted CRE was located in the 3'downstream intergenic region of

Hand-like. Predicted CREs for *RGS3* and *#1834* have relatively low nucleosome occupancy score, but others have mixed nucleosome occupancy score.

For each predicted CREs, a 1-3kb intergenic region containing the predicted clusters along with the endogenous promoter region and predicted ATG start codons were amplified and fused in frame to a *lacZ* reporter. Two of these constructs, *#1839* (-2005:*lacZ*) and *RGS3* (-1120:*lacZ*), displayed weak reporter expression in the TVC lineage along with strong expression in adjacent lineages. Other constructs, *#11094*(-904:*lacZ*), *UNC5*(*-2248*:*lacZ*) and *Hand-like*3'(7127-9220:*lacZ*), displayed no staining in TVCs (Figure 1).

To further examine the presumed role of the potential *RGS3* and *#1839* enhancers in TVC expression, we designed two strategies Initially, we built minimal intergenic fragments beginning just prior to the predicted CRE domain, *RGS3*(-274:*lacZ*), *#1839*(-1124:*lacZ*), to examine if they can drive reporter expression in TVCs, and found that both of them retained weak TVC lineage reporter activity (Fig. 1). We then attempted to test whether removal of the predicted CRE domain would abrogate TVC lineage reporter expression (*RGS3* -116:*lacZ*, *#1839* -849:*lacZ*). Although removal of the CRE region did seem to reduce the reporter expression in TVCs, the low penetrance of TVC reporter expression driven by the CRE containing constructs (< 5% of transgenic embryos) made it difficult to rigorously assess the impact of CRE removal (Fig.1). Subsequently, instead of using endogenous promoters, we amplified *RGS3*(-1120-166) and *#1839*(-2005-441), and fused them to a strong 297 bp *Hand-like* basal promoter to see if *#1839*(-2005-441:HL296:*lacZ*) and *RGS3*(-1120-166:HL296:*lacZ*) could boost TVC reporter expression. *#1839*(-2005-441:HL296:*lacZ*) had no reporter gene expression in TVCs, however, transgenic *RGS3*(-1120-116:HL296:*lacZ*) embryos had strong and consistent reporter staining in TVCs (Fig.1,2).

Next, we built and tested the minimal enhancer region for *RGS3*. First we found that *RGS3*(-274-166:HL296:*lacZ*), which contained only the predicted CREs, retained strong TVC lineage expression. This construct included one nonconserved ATTA motif and one conserved ATTA motif. Next, we removed the nonconserved ATTA by deleting the 5' end of the predicted CREs (*RGS3*-274-166:HL296:*lacZ*). This construct still displayed robust TVC lineage expression (Figure 2). However, targeted mutagenesis of both ATTA motifs in the context of the minimal predicted element (-274-166A1/A2-mut) and the full length construct (-1120-116 A1/A2-mut) nearly eliminated TVC reporter expression.

Discussion:

Our result demonstrates a new approach to identify CREs. This approach is to look for intergenic regions including important paired motifs adjacent to the candidate target gene. For early heart genes, we looked for intergenic regions including Ets-binding site and ATTA motif. Our result also helped us to understand the initial gene regulatory network during early heart specification. Our results strongly support a central role for the Ets/ATTA motif combination in the primary TVC cis-regulatory code. CREs for *RGS3* and *#1839* had low nucleosome occupancy scores, but others had mixed nucleosome scores. For the following reporter expression analysis, only *RGS3* and *#1839* CREs could drive the reporter expression in TVCs. The result indicates that there is a strong correlation between low-nucleosome binding scores and regulatory function in *Ciona* non-coding sequences as previously observed in (Khoueiry, 2010). The mutation assay for *RGS3* CREs proved that the ATTA motif is important to drive early heart gene expression. It would be possible to link these findings to vertebrate heart development after the determination of the transcription factor binding to this ATTA motif.

Only two out of five of our predicted regions resulted to be the corrected CREs which can drive reporter expression in TVCs. We intended to modify the search criteria to optimize the bioinformatics software prediction. Firstly, a lower nucleosome occupancy score will be added to the criteria, since based on our results, both predicted low nucleosome scoring regions worked. Secondly, predicted regions for *RGS3* and *#1839* are both within three thousand base pairs of target genes. So, we can search for intergenic regions closer to candidate genes. Then, new criteria for the chordator search could be: to search for two conserved ETS binding sites and one conserved ATTA motif within 50bp window, 3000bp adjacent to candidate genes, and have low-nucleosome occupancy score. I would like to test predicted CREs from this search to see if accuracy is increased.

After achieving highly accurate predictions, we can use this method to look for unidentified primary heart genes. Instead of looking for clusters adjacent to the primary TVC gene set, we can use chordator to search for predicted CREs against all gene models in *Ciona intestinalis*. We can focus on those genes which are adjacent to the predicted region, but are not included in the primary heart gene set. Then, we can do the function analysis of the predicted CREs for those unidentified genes, to see if they can drive reporter expression in TVCs. If they can, then those genes are potential unidentified primary heart genes. Next, we can do in-situ to see if those potential unidentified primary heart genes are actually expressed in TVCs.

References:

Beh, J., Shi, W., Levine, M., Davidson, B., Christiaen, L., (2007). FoxF is essential for FGFinduced migration of heart progenitor cells in the ascidian Ciona intestinalis. Development, 134, 3297-305.

Cooley, J., Whitaker, S., Sweeney, S., Fraser, S., Davidson, B., (2011). Cytoskeletal polarity mediates localized induction of the heart progenitor lineage. Nat Cell Biol, 13, 952-7.

Davidson, B., Shi, W., Beh, J., Christiaen, L., Levine, M., (2006). FGF signaling delineates the cardiac progenitor field in the simple chordate,Ciona intestinalis. Genes and Development, 20: 2728-2738.

Davidson B. (2007). Ciona intestinalis as a model for cardiac development. Cell and Developmental Biology, 18,16-26.

Dehal, P., Satou, Y., Campbell, R. K., Chapman, J., Degnan, B., De Tomaso, A., Davidson, B., Di Gregorio, A., Gelpke, M., Goodstein, D. M., Harafuji, N., Hastings, K. E., Ho, I., Hotta, K., Huang, W., Kawashima, T., Lemaire, P., Martinez, D., Meinertzhagen, I. A., Necula, S., Nonaka, M., Putnam, N., Rash, S., Saiga, H., Satake, M., Terry, A., Yamada, L., Wang, H. G., Awazu, S., Azumi, K., Boore, J., Branno, M., Chin-Bow, S., DeSantis, R., Doyle, S., Francino, P., Keys, D. N., Haga, S., Hayashi, H., Hino, K., Imai, K. S., Inaba, K., Kano, S., Kobayashi, K., Kobayashi, M., Lee, B. I., Makabe, K. W., Manohar, C., Matassi, G., Medina, M., Mochizuki, Y., Mount, S., Morishita, T., Miura, S., Nakayama, A., Nishizaka, S., Nomoto, H., Ohta, F., Oishi, K., Rigoutsos, I., Sano, M., Sasaki, A., Sasakura, Y., Shoguchi, E., Shin-i, T., Spagnuolo, A., Stainier, D., Suzuki, M. M., Tassy, O., Takatori, N., Tokuoka, M., Yagi, K., Yoshizaki, F., Wada, S., Zhang, C., Hyatt, P. D., Larimer, F., Detter, C., Doggett, N., Glavina, T., Hawkins, T., Richardson, P., Lucas, S., Kohara, Y., Levine, M., Satoh, N., Rokhsar, D. S., (2002). The draft genome of Ciona intestinalis: insights into chordate and vertebrate origins. Science. 298**,** 2157-67.

Di Gregorio, A., Levine, M., (2002). Analyzing gene regulation in ascidian embryos: new tools for new perspectives. Differentiation, 70,132-9.

Johnson, D. S., Davidson, B., Brown, C. D., Smith, W. C., Sidow, A., (2004). Noncoding regulatory sequences of Ciona exhibit strong correspondence between evolutionary constraint and functional importance. Genome Res, 14, 2448-56.

Khoueiry, P., Rothbacher, U., Ohtsuka, Y., Daian, F., Frangulian, E., Roure, A., Dubchak, I., Lemaire, P., (2010). A cis-regulatory signature in ascidians and flies, independent of transcription factor binding sites. Curr Biol, 20, 792-802.

Ornitz, D. M., Itoh, N., (2001). Fibroblast growth factors. Genome Biology, 3005.1-3005.12.

Satou, Y., Kawashima, T., Tokuoka, M., Awazu, S., Satoh, N., (2003). A genomewide survey of developmentally relevant genes in Ciona intestinalis. Dev Genes Evol, 213:254–263.

Sharrocks, A. D., (2001). The Ets-domain Transcription Factor Family. Molecular Cell Biology, 2,827-837.

Figure 1.

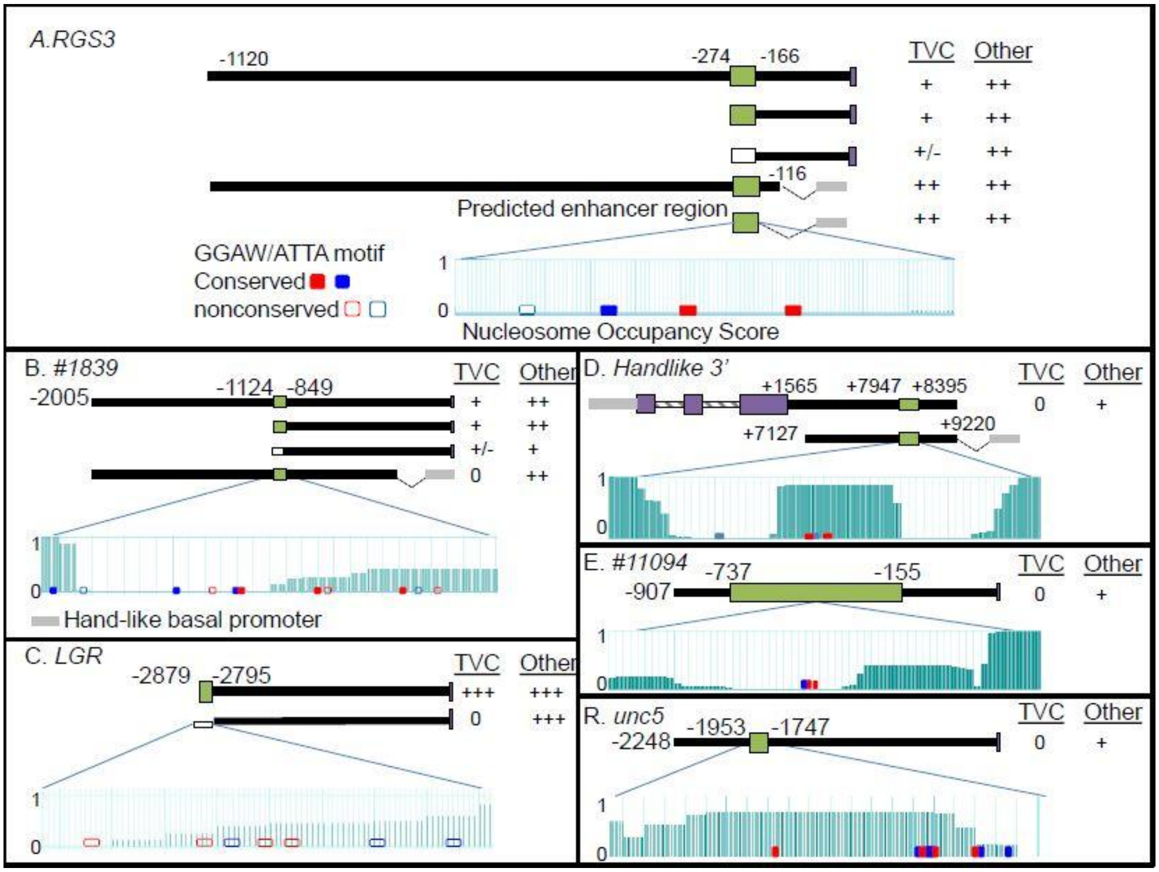

Figure 1. Reporter analysis of the predicted region.

Reporter testing of possible enhancer regions containing predicted CREs for primary TVC expression of *RGS3, aniseedV3_1839, Handlike, aniseedV3_11094* and *unc5*.

Nucleosome occupancy scores for predicted regions are shown.

Figure 2.

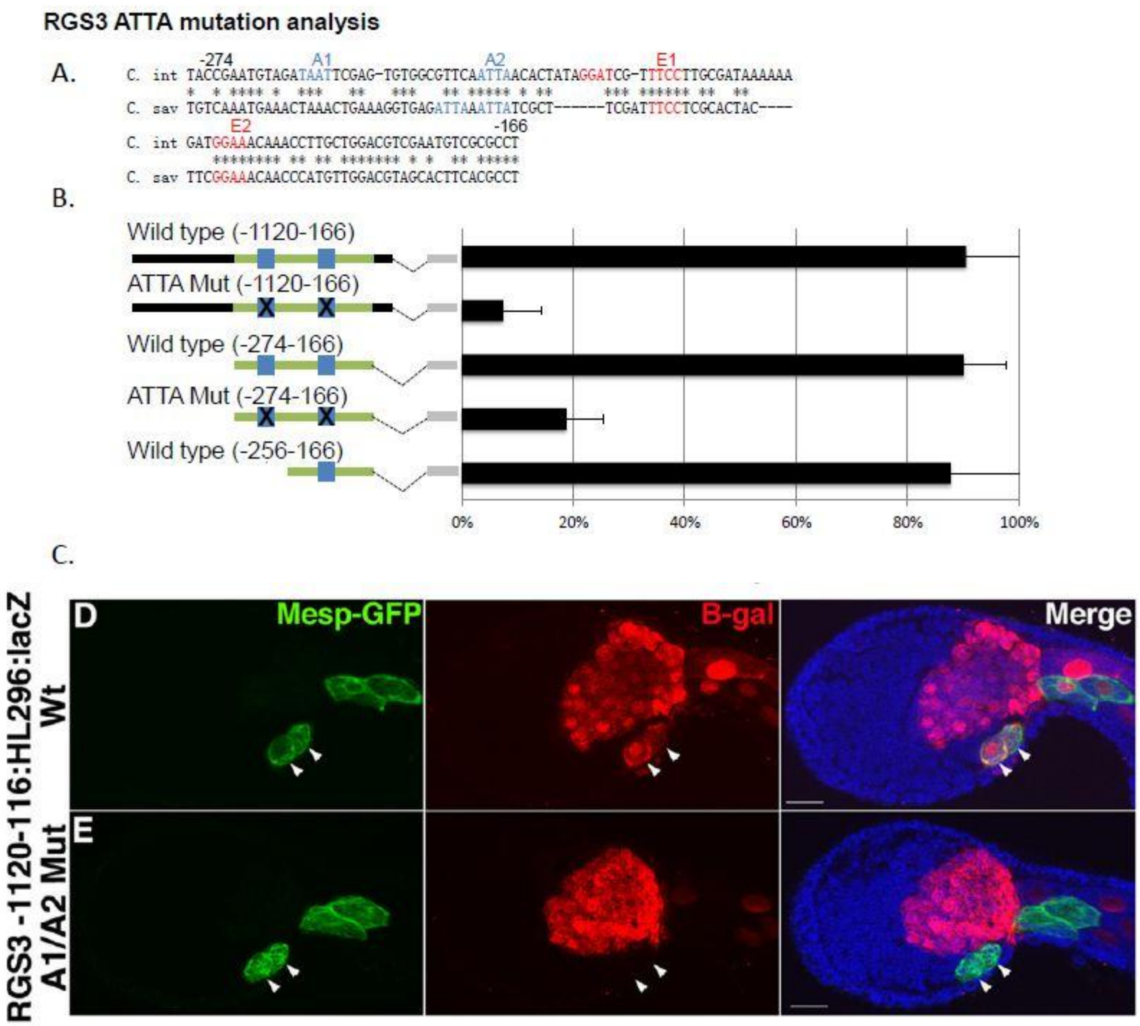

Figure 2. RGS3 mutation analysis. (A) Alignment of the predicted region for *RGS3* in *C. intestinalis* and *C. savignyi*. Red sequence highlights Ets1/2 binding site. Blue sequence highlights ATTA motif. (B) Pointed mutations were introduced into ATTA co-motif sequences. Histogram depicts percent of embryos showing TVC reporter expression (error bars indicate standard deviation, three trials for each sample). (D,E) representative embryos co-electroporated with wild-tye and mutated *RGS3* constructs and Mesp-GFP to visualize the B7.5 lineage.

Figure 3.

Primer	sequence
RGS3 EX1 forward	AAA GGATCC CGATATACGATGCGCCTTTGGGTG
RGS3 EX1 reverse	TTTGCGGCCGC CGG ACA CGC TTT CCA CTT CAT TG
#1897 EX1 forward	AAAGGATCCCCGCATGTCGCGCGAAGTCG
#1897 EX1 reverse	TTTGCGGCCGC GGT CTC CAT CTC GTC TGA GGA TGA C
#11094 forward	AAAGGATCC CATTGACTGGATAGAGGTTCACGCG
#11094 reverse	TTTGCGGCCGC CTC AAG CCG TGT CGC CAT
UNC5 forward	AAACTCGAG CCTGTGTCTGTTTATCTTCGGC
UNC5 reverse	TTTGCGGCCGC CGG CTC GTA TTC GTC AAC CAT GG
Hand-like 3' forward	AAA TCTAGA CCTGTTTAAAGTAGACATGTCCG
Hand-like 3' reverse	TTTGGATCC CATTGTACAGTAATTCCTAGGATC

Figure 3. Primer table. This primer table shows all the primers used to clone possible enhancer regions with endogenous promoter and ATG start codon.