FUNCTION AND REGULATION OF THE *STRONGYLOCENTROTUS PURPURATUS GATAE* GENE

Thesis by

Pei Yun Lee

In Partial Fulfillment of the Requirements

for the Degree of

Doctor of Philosophy

California Institute of Technology

Pasadena, California

2007

(Defended January 31, 2007)

ACKNOWLEDGEMENTS

First and foremost, I would like to thank my advisor, Eric Davidson, for his vast knowledge on just about everything, for teaching me how to do biology quantitatively, and for his vision. My committee members Marianne Bronner-Fraser, Ellen Rothenberg, and David Anderson, for all their comments and questions regarding my project. Each and every one of you contributed greatly to my learning in graduate school.

Members of the Davidson lab, who provided a fun and intellectually stimulating environment. It has been a real pleasure working with such a group of talented individuals. A special thank you goes to Paola Oliveri for taking me under her wing when I first started in the lab and teaching me how to work with the sea urchin embryo. Veronica Hinman for discussions on being a woman scientist. Jongmin Nam for his worldview and the hours discussing my project and the latest adventures of Jack Bauer. Deanna Thomas and Jane Rigg for ensuring that the lab functions on a daily basis. Past and present residents of 017K, Jane Wyllie, Cesar Arenas-Mena for his "tell it like it is" attitude, Takuya Minokawa for the daily Rathskeller lunches and for teaching me the "new" *in situ* method, Qiang Tu for his graciousness in offering to let me use whatever equipment I needed as I was finishing up my degree, and especially to Roger Revilla, whom I have spend my entire graduate student life sharing an office and bay with, and who is not just a wonderful scientist, but also a great friend.

Bob Goldberg at UCLA, who first awakened my love for science with his unorthodox teaching methods, who first taught me how to do great science in his

laboratory, and for being a mentor on all things science and nonscience for the past ten years.

The men and women of the Angeles Chorale, who have kept my connection to the non-science world on Tuesday evenings, and especially to Don Neuen, for demanding precision, artistry and passion in every note and challenging me to be a better musician.

Debbie Goh and Jillian Baker, for being the best girlfriends anyone could ask for.

My parents, for their unconditional love and support in the last 30 years. My brother Min, for keeping me grounded in 27 of those years. And my fiancé Eric, for having traveled this entire journey with me, and with whom I look forward to even greater things.

ABSTRACT

The *Strongylocentrotus purpuratus gatae* is orthologous to vertebrate *gata4/5/6* genes. *gatae* is expressed throughout embryogenesis, beginning in the 15 h blastula in presumptive mesoderm cells, and at mesenchyme blastula, in endoderm and mesoderm cells of the veg2 lineage. During gastrulation, *gatae* is expressed in the midgut, hindgut and mesoderm, while in the pluteus expression it is limited to the midgut and coelomic pouches. Perturbation of *gatae* expression resulted in the lowered RNA levels for many endomesoderm transcription factors, including *foxA*, *brachyury*, and *β1/2-otx*, highlighting Gatae's role as a regulator of transcription factors. *gatae* occupies an important node in the endomesoderm gene regulatory network, using its cross-regulatory interactions with *otx* to stabilize the endomesoderm gene expression program. *Cis*-regulatory analysis of *gatae* identified two modules responsible for its embryonic expression. Module 10 drives endomesoderm expression in the blastula, while module 24 activates gut expression in the gastrula and pluteus. Deletion of module 10 from a *gatae* GFP BAC resulted in a complete loss of blastula stage expression, demonstrating its necessity and sufficiency for early activity. Global *cis*-regulatory analysis of the *gatae* locus suggests that module usage is exclusionary; only one module can associate with the basal transcriptional apparatus and affect gene transcription at any given time. The endomesoderm gene regulatory network predicts that *gatae* is downstream of Otx and Notch signaling. Analysis of the sequence of module 10 identified Otx and Suppressor-of-Hairless (Su(H)) binding sites. Injection of Otx-engrailed RNA repressed the expression of module 10:GFP reporter; the effect is abolished when Otx binding sites

were mutated. Gel shifts demonstrated that the Otx protein binds to module 10. Module 10 expression was reduced under perturbation of Notch signaling. Mutations of either Otx or Su(H) binding sites resulted in lowered GFP RNA levels with no effect on spatial expression. Mutations of both Otx and Su(H) binding sites led to a further reduction but not elimination of reporter expression, suggesting that another input is involved. This unknown input was determined to be also downstream of Notch signaling and that *gatae* regulation functions via OR logic.

TABLE OF CONTENTS

INTRODUCTION

Function of GATA Transcription Factors in Endoderm Specification

GATA factors are a class of zinc finger transcription factors named for binding to a GATA motif (Evans and Felsenfeld, 1989). Binding site studies have determined that GATA factors bind a WGATAR consensus sequence (Evans et al., 1988), although different family members have subtle differences in binding site preference (Ko and Engel, 1993; Merika and Orkin, 1993). The first GATA factor cloned was the chicken *gata1* gene (Evans and Felsenfeld, 1989). Subsequently five more genes encoding GATA factors were identified in the vertebrates, named *gata1-6*, all of which play important roles during development. GATA factors are divided into two main classes: members of the *gata1/2/3* family function in hematopoiesis (Orkin and Zon, 1997), while the *gata4/5/6* genes are widely expressed and utilized in the specification of endoderm and mesoderm specification as well as associated organ development.

Network based approach to GATA factor function

Traditionally, the study of a developmental process involves the generation of mutants through chemical or insertional mutagenesis followed by a screen for phenotypes pertaining to the process of interest. While this approach has identified many important genes and contributed greatly to our understanding of animal development, it is not without its limitations. Mutations manifested by dramatic phenotypes have turned out to

encode differentiation proteins, while mutations in genes encoding transcription factors are often lethal and not identified through mutant screens. Furthermore, phenotypic observations alone do not provide any information on the epistatic relationships between genes.

Transcription factors which are widely expressed during development, such as the GATA4/5/6 factors, are particularly difficult to study. Mouse knockouts in GATA genes result in early embryonic lethality, which precludes the analysis of their roles in development. A more informative approach is to study the function of GATA genes in the context of a gene regulatory network (GRN). Developmentally expressed genes do not function in parallel linear processes, rather they are integrated in complex networks containing activating and repressive interactions, signaling toggle switches, feed-back and feed-forward loops, to name a few. Presenting our current state of knowledge of GATA factors in the context of gene networks will further clarify their functions at the molecular level, and also provide broader views of their roles during endoderm specification. Comparison of similar gene networks involving GATA factors in different organisms may also lead to insights on regulatory circuitry conservation during evolution. On a more general note, an understanding of developmental GRNs will also enable the definition of classical terms used by developmental biologists such as specification and commitment at the molecular level

In this review I will discuss the current state of knowledge pertaining to GATA transcription factor function and usage in endoderm specification and gut development. I have constructed a series of GRNs from six model organisms, the mouse, zebrafish, frog, worm, fly and sea urchin, using BioTapestry Editor (Longabaugh et al., 2005), focusing

on connections into and out of the GATA factors. All GRNs generated in this review are "views from the genome," meaning each includes interactions that take place over time and are not representative of events occurring in a single cell. In each GRN, connections originate from the upstream gene into a downstream one. Activating and repressive interactions are represented by arrows and bars respectively. Double arrows flanked by two circles denote intercellular signaling events. Any interaction that has been proven to be direct, either through transcription factor binding site mutation or otherwise, is indicated with a polygon of the same color as the upstream input. All of the GRNs are GATA-centric, and their main purposes are to highlight the role of GATA factors in each organism. Therefore while the connections into and out of the GATA factors are complete with respect to the current literature, none of the GRNs include all the gene interactions during endoderm specification.

While the extent and state of understanding of endoderm specification and gut development is different in each organism, a common theme emerges: the main function of GATA factors is to establish transcriptional domains through the regulation of endodermal transcription factors, in some cases by activating other GATA factors in a sequential fashion, and the direct activation of differentiation genes in later development. Interestingly, engagement of GATA factors in cross- and autoregulatory loops suggest that one of their main roles is in the stabilization and lockdown of the transcriptional program for endoderm specification and organogenesis.

Biochemical properties of GATA factors

The GATA factors contain either one or two class IV zinc fingers, characterized by the $CX_2CX_{17}CX_2C$ motif (Evans et al., 1988). Biochemical studies have shown that the two zinc fingers have slightly different roles: the C terminal finger is required for binding site recognition (Morrisey et al., 1997b; Omichinski et al., 1993b; Visvader et al., 1995; Yang and Evans, 1992), while the N terminal finger contributes to binding specificity and stability (Yang and Evans, 1992) and increasing the spectrum of binding sites recognized (Merika and Orkin, 1993). In the GATA proteins that only possess a single zinc finger, it is the C terminal finger that is present (Lowry and Atchley, 2000), consistent with its necessity for DNA binding. NMR structural studies have shown that the C-terminal zinc finger interacts with the major groove of DNA, while the successive basic domain interacts with the minor groove of target sequences (Omichinski et al., 1993a). Studies of the GATA protein has implicated the C-terminal basic domain of the protein in transactivation (Nemer et al., 1999; Yang and Evans, 1992), while the N terminal domain plays an important role in protein-protein interactions with cofactors such as FOG (Svensson et al., 1999; Tevosian et al., 1999; Tsang et al., 1997).

Evolution of GATA transcription factors

A comprehensive phylogenetic analysis of GATA factors was detailed in Lowry and Atchley (2000), which determined that all GATA factors evolved from an ancestral GATA factor contained a single zinc finger (Fig. 0.1). Zinc finger duplication occurred

before the split between metazoa and fungi, a conclusion supported by the fact that the N and C terminal zinc fingers in *Drosophila* are more closely related to its corresponding finger in the sea urchin and the vertebrates than to each other. Subsequent to zinc finger duplication, an insertion occurred between the two zinc fingers in the fungi, increasing the interfinger distance to 120 - 140 aa rather than the 30 aa observed in vertebrates. The ancestral deuterostome likely had two *gata* genes, supported by the fact that the echinoderms *Asterina miniata* (Hinman, pers. comm.) and *Strongylocentrotus purpuratus* (Pancer et al., 1999), and the urochordate *Ciona intestinalis* all have two GATA factors

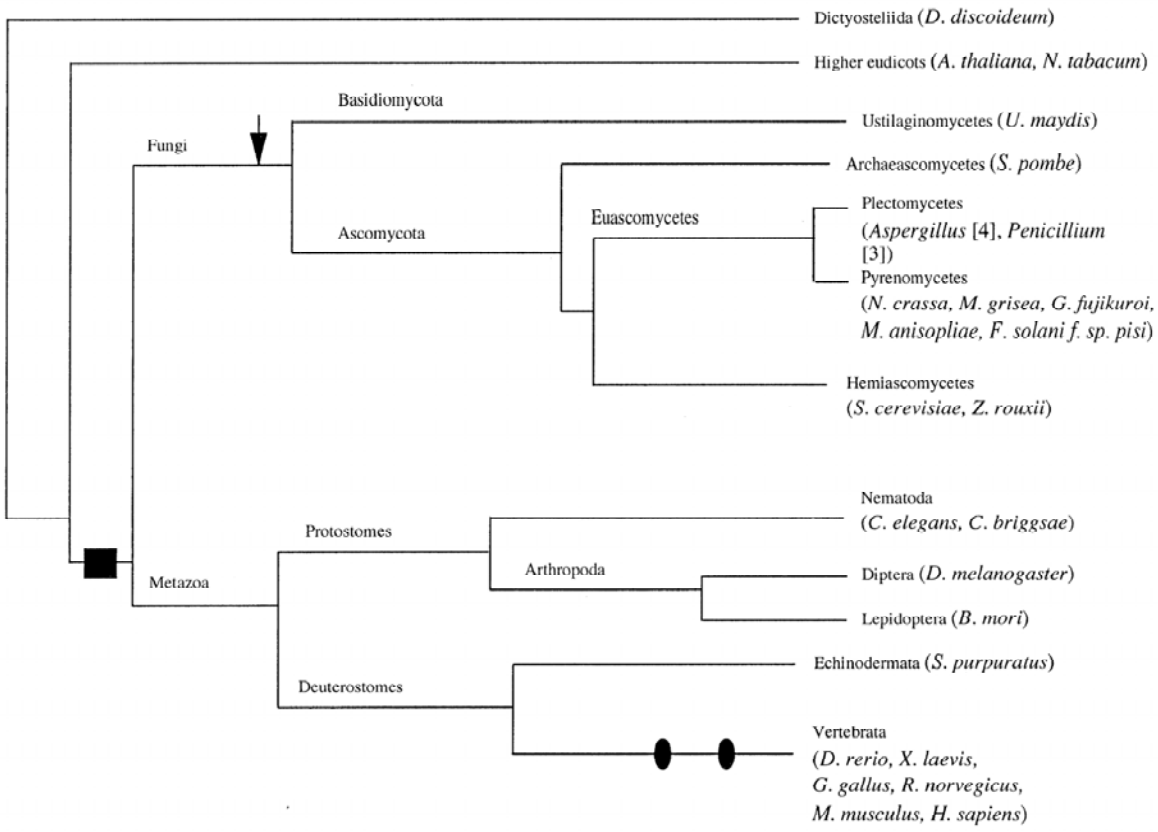


Figure 0.1. Summary diagram for the evolution of GATA factors. The black box represents the duplication of zinc fingers before the divergence of fungi and metazoa. The arrow indicates the occurrence of a 100 aa insertion between the two zinc fingers; black ovals highlight genome duplication events in the vertebrates. From Lowry and Atchley, 2000.

(Yamada et al., 2003), while all jawed vertebrates studied thus far contain six *gata* genes. *C. elegans* and *Drosophila* have eleven and five *gata* genes respectively, which could not be grouped with any certainty to *gata1/2/3* or *gata4/5/6*.

Mouse

gata4/5/6 genes are widely expressed in the mouse during embryonic development and adulthood. In the adult mouse *gata4/5/6* are expressed in the heart, lung, liver, gut, bladder, kidney and gonads, while in the embryo *gata4/6* expression have been detected in extraembryonic visceral and parietal endoderm, definitive endoderm at the foregut and midgut junction, and lateral plate mesoderm. All three genes are also expressed in the intestine epithelium and developing endodermal and mesodermal organs (Arceci et al., 1993; Jacobsen et al., 2002; Koutsourakis et al., 1999; Morrisey et al., 1996; Morrisey et al., 1997a; Narita et al., 1996; Soudais et al., 1995) (Table 0.1).

Functional studies of the *gata4/5/6* genes in the mouse have proven to be very difficult because of their importance in the development of extraembryonic endoderm. Knockouts in *gata4* resulted in embryos with no visceral endoderm and death at E8 with severe heart and gut defects (Kuo et al., 1997). *gata6* knockouts die shortly after implantation, are much smaller in size, lack part of visceral endoderm and displayed embryonic ectoderm defects (Koutsourakis et al., 1999; Morrisey et al., 1998). Analyses using chimeric embryos demonstrated that the lethalities from *gata4/6* knockouts were due to a lack of visceral endoderm differentiation (Koutsourakis et al., 1999; Morrisey et

al., 1998; Soudais et al., 1995), though *gata4* mutants also possessed intrinsic defects in gut epithelium development (Jacobsen et al., 2002). Unlike *gata4/6* mutants, *gata5* knockouts were viable and displayed only defects in female genitouninary tract development (Molkentin et al., 2000) (Table 0.1).

gata4/6 are among the earliest genes expressed in the endoderm. The HMG transcription factor Sox7 activates *gata4* transcription (Futaki et al., 2004) (Fig. 0.2). Once activated, Gata4 feeds back and activates *sox7* (Murakami et al., 2004), setting up the first cross-regulatory loop. Gata4, together with Sox7, have also been shown to directly activate *fgf3* (Murakami et al., 2004). In addition, Gata4 also turns on *gata6* (Fujikura et al., 2002) which is also regulated by retinoic acid (RA) signaling (Capo-Chichi et al., 2005). Once activated, Gata6 serves to reinforce *gata4* expression in the second cross-regulatory loop in this network (Futaki et al., 2004; Morrisey et al., 1998). Next, Gata4/6 activate a number of endoderm specific transcription factors, including *hnf1β*, *hnf3β*, *hnf4*, *sox17* and *gata4/5/6* (Fujikura et al., 2002; Futaki et al., 2004; Morrisey et al., 1998; Murakami et al., 2004; Soudais et al., 1995). *Cis*-regulatory analysis on gut differentiation genes such as *lactase-phlorizin hydrolase* (*LPH*), and *sucrase-isomaltase* (*SI*) have shown that they are under the direct control of GATA factors and Hnf1α (Boudreau et al., 2002; van Wering et al., 2004).

Table 0.1. Expression patterns and knockout phenotypes for GATA4/5/6 factors.

Gene	Expression Pattern[a]	Endodermal Phenotype
Mouse		
Gata4	Adult: heart, ovary, testis, lung, liver, stomach, small intestine Embryo: extraembryonic visceral and parietal endoderm, definitive endoderm at foregut and midgut junction, lateral plate mesoderm, heart, testis, ovary, liver, stomach and small intestine epithelium	No visceral endoderm, disorganized foregut, gastric epithelium defects, death at E8
Gata5	Adult: stomach, small intestine, bladder, lung Embryo: heart, lung bud, bladder, urogenital ridge, gut epithelium	Defects in genitourinary tract development
Gata6	Adult: heart, stomach, small intestine, ovaries, bladder, liver, lung, kidney Embryo: visceral and parietal endoderm, primitive streak mesoderm, lateral plate mesoderm, heart, lung buds, urogenital bridge, foregut and midgut, vascular smooth muscle	No visceral endoderm, death at E6.5 - E7.5, defects in hepatic differentiation, lung branching morphogenesis
Zebrafish		
Gata4	Lateral plate mesoderm, gut and liver primordia, heart tube, gut	Long thin gut tube with no epithelial folds, lack of severely reduced liver and pancreas, early endoderm specification appears normal
Gata5	Endodermal precursors in late blastula, yolk syncitial layer, gastrulating endoderm, gut tube epithelium, lateral plate mesoderm, heart tube	Reduced endoderm, lack of gut looping, endodermal organ defects
Gata6	Developing gut, heart tube, gut tube	Enlarged gut, uncoiled intestine with no lumen
Frog		
Gata4	Vegetal cells of blastula, supra blastoporal endoderm, ventral cardiac mesoderm, cardiac tube, heart, stomach	Loss of gut coiling and reduced gut tissue
Gata5	Vegetal cells of blastula, sub blastoporal endoderm, stomach	Loss or defect in gut coiling
Gata6	Vegetal cells of blastula, supra blastoporal endoderm	Reduced gut tissue with no coiling
Worm		
Med1/2	EMS at 4 cell stage, then MS and E, adult gonads	E, MS to C transformation, no MS derived pharynx
End1	E cell in 8 cell embryo	Premature division and defective gastrulation
End3	E cell activated slightly before end1	Some mutants lack endoderm, lack of end1/3 most embryos no endoderm
Elt2	Intermediate descendents of E, expressed throughout life	Death at L1 with malformed gut
Elt4	Intestine, posterior bulb of pharyns	No detectable phenotype, no downstream genes identified
Elt7	Similar to elt2	No obvious phenotype
Fly		
Srp	Vitellophages, hemocyte primordium, amnioserose, fat body precursors, midgut primordium	Transformation of endodermal midgut into ectodermal foregut and hindgut
Gatae	Anterior and posterior endoderm, malpighian tubules, larval and adult midgut	No obvious morphological defects, most embryos do not hatch
Grn	Developing endoderm, heard posterior spiracles, CNS	No obvious endodermal phenotype
Sea Urchin		
Gatae	Mesoderm precursors, endoderm and mesoderm at mesenchyme blastula, midgut and hindgut, mesoderm at tip of archenteron, coelomic pouches	No gastrulation, disorganized endoderm and mesoderm cells

[a] Gene expression patterns from the mouse, zebrafish, Xenopus, worm and fly and sea urchin are summarized together with any endodermal phenotypes from the perturbation of the genes. Phenotypes were obtained either from genetic knockouts, genetic mutations, RNAi and MASO injections.

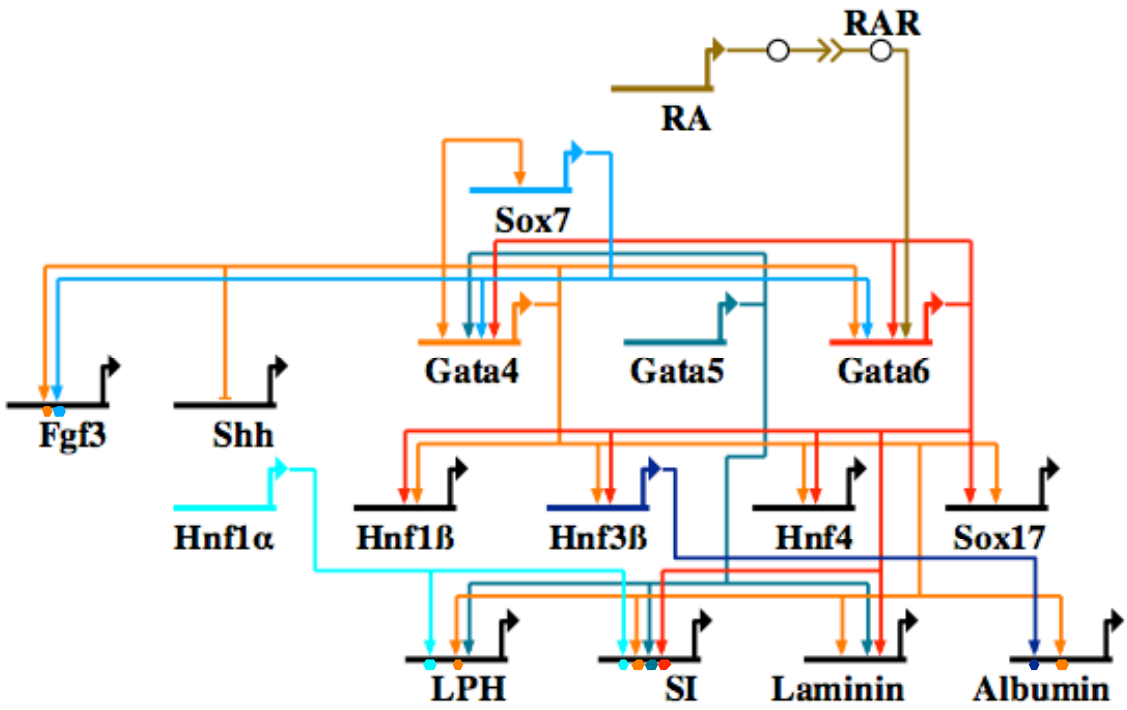


Figure. 0.2. Network model outlining gene interactions underlying endoderm specification in the mouse, *Mus musculus*. Abbreviations of gene names are as follows: LPH, *lactase-phlorizin hydrolase*; SI, *sucrase-isomaltase.*

Zebrafish

As a model organism, the zebrafish *Danio rerio* is well suited for the study of animal development due to the ability to generate mutants, the ease of gene transfers and perturbation with mRNA and morpholino antisense oligonucleotides (MASO). As in the mouse, GATA factors in the zebrafish also display overlapping expression patterns. *Drgata4* is expressed in the lateral plate mesoderm, gut and liver primordia, heart tube

and gut (Table 0.1). *gata5* is the earliest expressing GATA factor in the zebrafish, first observed in the endodermal precursors in the late blastula and yolk syncitial layer (Reiter et al., 2001). It is also expressed in the gastrulating endoderm, gut tube epithelium, lateral plate mesoderm and heart tube (Reiter et al., 1999). *gata6* is also expressed in the developing gut, heart and gut tubes (Holtzinger and Evans, 2005). Perturbation of the expression of any of the *gata* genes by genetic mutation or MASO injections led to defects in gut looping and morphogenesis (Holtzinger and Evans, 2005; Peterkin et al., 2003). Genetic mutation of the *gata5* (also known as *faust*) gene also led to reduced endoderm (Reiter et al., 1999). In addition to the gut, defects in organs such as heart, liver and pancreas were also observed (Holtzinger and Evans, 2005; Reiter et al., 2001).

Nodal signaling plays an important role in zebrafish endoderm formation. Mutations in the two zebrafish nodal genes, *squint* (*sqt*) and *cyclops* (*cyc*), or the Nodal co-receptor *one-eyed pinhead* (*oep*) lead to a complete loss of endoderm and mesoderm (Feldman et al., 1998; Gritsman et al., 1999). Downstream of Nodal signaling are three genes, *mezzo* (*mez*), *bonnie and clyde* (*bon*) and *gata5* (Poulain and Lepage, 2002; Rodaway et al., 1999) (Fig. 0.3). These three transcription factors, together with a maternal T box factor Eomesoderm (Eomes), activate the transcription of *casanova* (*cas*), which encodes a Sox-like protein whose mutation leads to a total loss of endoderm (Aoki et al., 2002; Bjornson et al., 2005; Kikuchi et al., 2001). Cas then activates *gata5* in a cross-regulatory loop, and also activates the *gata4* and *gata6* genes (Alexander et al., 1999). Like Cas, Gata5 also feeds back and cross-regulates with *gata4*, and plays a role in the activation of *gata6*. In addition, Cas also activates endoderm transcription factors

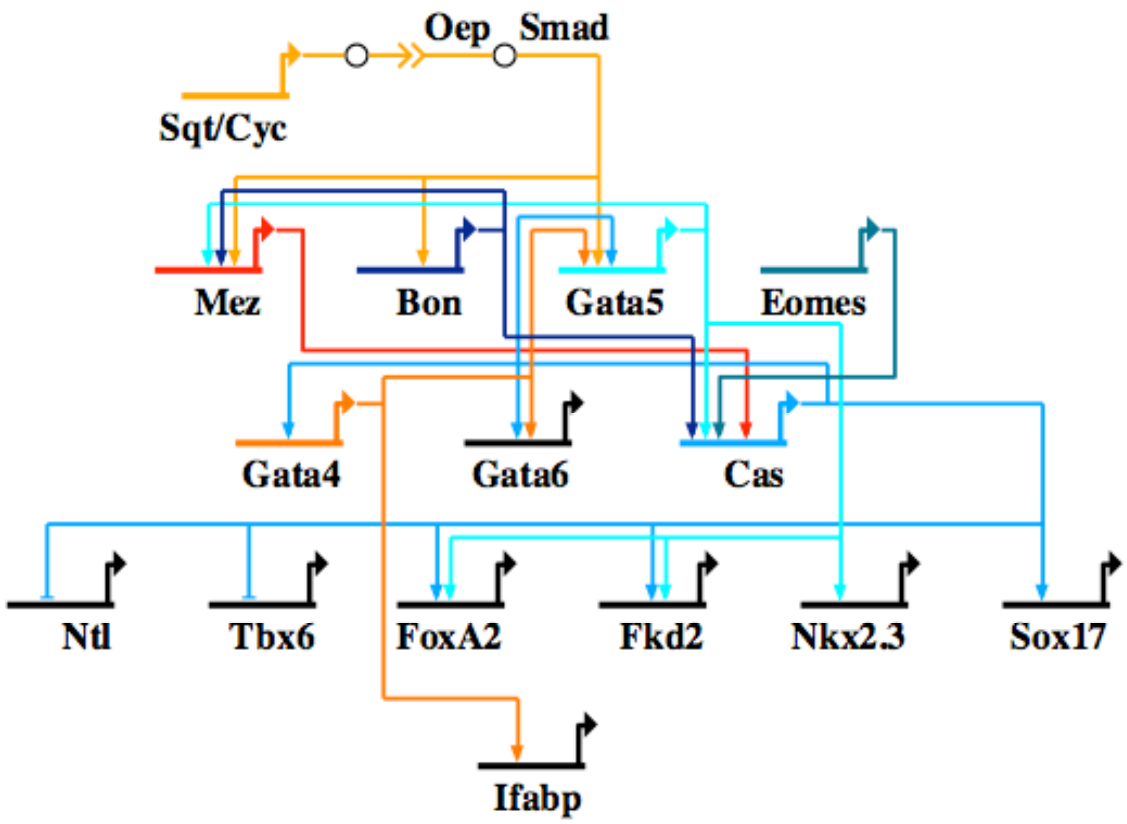


Figure 0.3. Network model outlining gene interactions underlying endoderm specification in the zebrafish *Danio rerio*. Gene abbreviations are as follows: Bon, *bonnie and clyde*; Cas, *casanova*; Cyc, *cyclops*; Eomes, *eomesodermin*; Mez, *mezzo*; Ntl, *no tail*; Oep, *one-eyed pinhead*; Ifabp, *intestinal fatty acid binding protein*; Sqt, *squint.*

axial/foxA2, *fkd2*, *nkx2.3* and *sox17* (Alexander et al., 1999; Kikuchi et al., 2001; Reiter et al., 2001) and functions in the repression of mesoderm genes *ntl* and *tbx6* (Aoki et al., 2002). It is unclear whether the *gata* genes play direct roles in activating any of these transcription factors, though Gata4 has been implicated in the activation of the intestinal fatty acid binding protein gene, *ifabp* (Holtzinger and Evans, 2005).

Frog

In the *Xenopus laevis* embryo, *gata4/5/6* are all expressed in the vegetal pole during blastula stages. *gata4/6* are expressed in the involuting supra blastoporal endoderm until gastrulation, while *gata5* is expressed in the non-involuting sub blastoporal endoderm in midgastrula (Afouda et al., 2005; Weber et al., 2000) (Table 0.1). Gata4 RNA has also been detected in the developing ventral cardiac mesoderm, cardiac tube, heart, stomach and other endodermally derived organs (Kelley et al., 1993). Translational inhibition of all three *gata* genes with MASO injections led to the same phenotype: a reduction in the amount of endodermal tissue and a loss or defect in gut coiling (Afouda et al., 2005).

The *Xenopus* endoderm specification program is initiated by the maternal T-box transcription factor VegT, which, when depleted, resulted in embryos that did not express many endodermal markers (Zhang et al., 1998). VegT is upstream of Nodal signaling, which plays a role in the activation of all three GATA factors (Hilton et al., 2003; Xanthos et al., 2001). VegT activates the *sox17* gene cell autonomously (Clements and Woodland, 2003) (Fig. 0.4). Once the *gata* genes are activated by Nodal signaling, they take over the regulation of *sox17* in the Nodal dependent phase of *sox17* expression (Afouda et al., 2005; Weber et al., 2000). In return, Sox17 engages the *gata5/6* genes in cross-regulatory loops (Sinner et al., 2006). A homeodomain transcription factor, Mixer, also regulates *gata6* and *sox17*. In turn, the three *gata* genes auto- and cross-regulate (Afouda et al., 2005). Once the Gata proteins are available, they activate endoderm

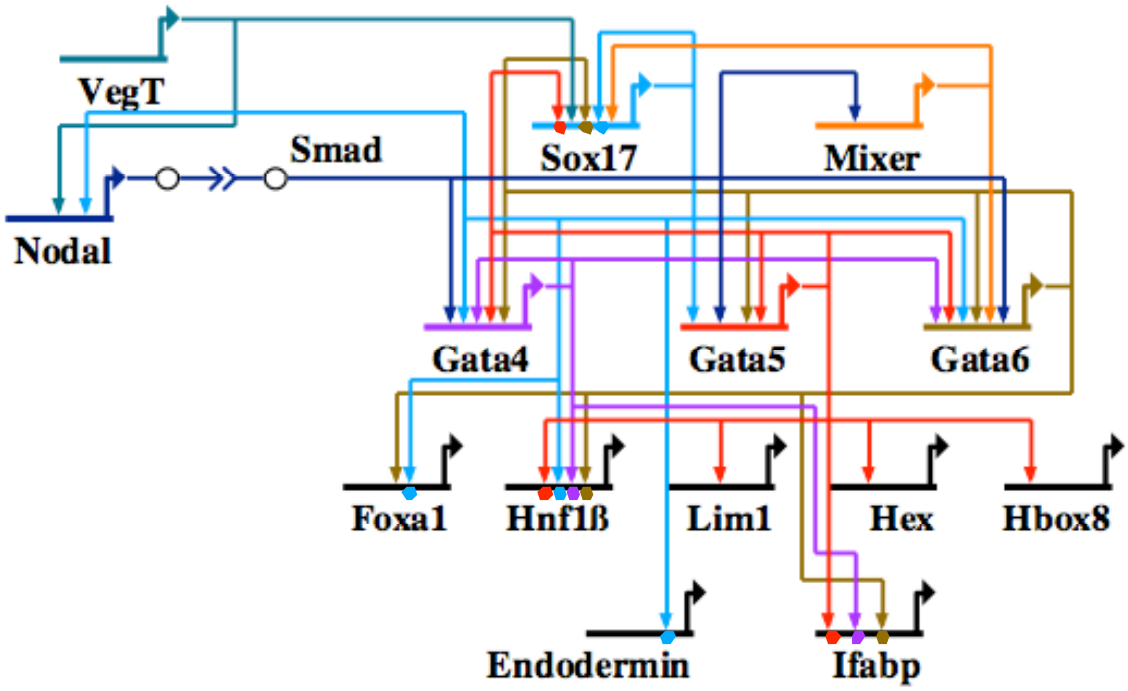


Figure 0.4. Network model outlining gene interactions underlying endoderm specification in the frog *Xenopus laevis*. Gene abbreviations are as follows: Ifabp, *intestinal fatty acid binding protein*.

transcription factors like *foxa1*, *hnf1β*, *lim1*, *hex* and *hbox8* (Afouda et al., 2005; Weber et al., 2000). At least one of these interactions, into *hnf1β*, is direct. In the gut differentiation gene battery, GATA factors have also been shown to be a direct activator for the intestinal fatty acid binding protein (*ifabp*) (Gao et al., 1998).

Worm

GATA factors have been studied extensively in the nematode *C. elegans*. The *C. elegans* genome encodes eleven GATA factors, seven of which are expressed in the endoderm. Usage of GATA factors appears to be at least partially redundant. The earliest expressing GATA factors are the *med-1/2* genes, which are activated in the EMS

cell at the four-cell stage. They are transiently expressed in the MS and E cells, which develop into the endodermal gut and body wall muscle (Table 0.1). In addition, they are also expressed in the adult gonads (Maduro et al., 2001). In *med-1/2* mutants, some of the embryos display E,MS to C (hypodermis) transformation. Recently, Maduro et al. (2006) demonstrated that the low penetrance of the zygotic *med-1/2* mutation was due to a maternal contribution to the Med-1/2 function.

The *end-1/3* genes are the earliest expressing genes in the E lineage, expressed in the E cell from the 8-cell embryo, with *end-3* activated slightly before *end-1* (Maduro et al., 2006; Maduro et al., 2005a; Maduro and Rothman, 2002; Zhu et al., 1997). Mutations in both *end-1* and *end-3* genes led to an E to C transformation, in which most embryos did not have any endoderm, whereas mutation of either one of the genes resulted in much weaker phenotypes (Maduro et al., 2005a; Zhu et al., 1997).

The final group of GATA factors transcribed are the *elt* genes. Three *elt* genes are expressed in the endoderm, *elt-2*, which is expressed from the 2E cell stages and whose expression lasts through the life of the animal (Fukushige et al., 1998). This includes *elt-4*, which is expressed in the intestine of the embryo and posterior bulb of the pharynx in late embryogenesis; and *elt-7*, which is expressed in a similar pattern as *elt-2* (Fukushige et al., 2003; Maduro and Rothman, 2002). Mutation of the *elt-2* gene led to death at the larval L1 stage with malformed guts, whereas mutations in *elt-4/7* did not result in a detectable phenotypes. To date, *elt-4* has not been demonstrated to be upstream of any gene and appears to be non-functional (Fukushige et al., 2003).

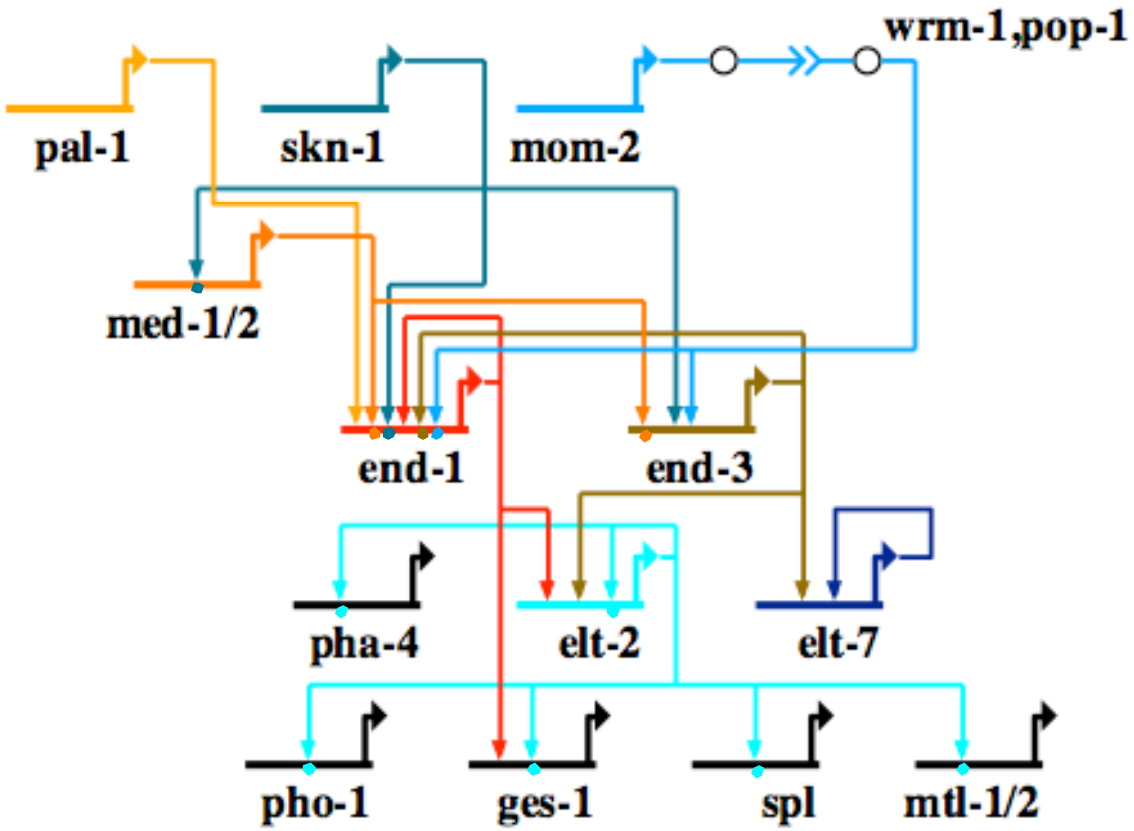


Figure 0.5. Network model outlining gene interactions underlying endoderm specification in the nematode *Caenorhabditis elegans*. Gene abbreviations are as follows: ges-1, *gut esterase;* mtl, *metallothionein*; pho-1, *essential acid phosphatase*; spl, *Sphingosine-1-phosphate Lyase.*

Notably, endoderm specification in *C. elegans* occurs through the sequential activation of the GATA factors. A bZIP maternal transcription factor, Skn-1, is responsible for turning on *med-1/2* in the EMS cell at the 4-cell stage (Maduro et al., 2001) (Fig. 0.5). Med1/2 have been demonstrated to directly bind the *cis*-regulatory region of *end-1/3* (Broitman-Maduro et al., 2005). In addition to activation by Med-1/2, *end-1/3* expression are also dependent on Wnt signaling. In the presence of Wnt signaling, Pop-1, a Lef homolog is switched from a repressor to an activator that activates *end-1/3* (Calvo et al., 2001; Maduro et al., 2005b). End-3, the earlier expressing of the

two *end* genes, also activates *end-1* (Maduro et al., 2006; Maduro et al., 2005b). A direct input from Skn-1 has also been identified in *end-1* (Maduro et al., 2005b). End-1/3 then activate the final pair of GATA factors involved in endoderm specification, *elt-2* and *elt-7*, which autoregulate and whose main roles are to activate differentiation gene batteries in the gut (Fukushige et al., 2005; Fukushige et al., 1998; Fukushige et al., 1999; Kalb et al., 1998; Maduro and Rothman, 2002; Moilanen et al., 1999; Oskouian et al., 2005; Peterkin et al., 2003; Zhu et al., 1998).

Fly

The *Drosophila* genome has five *gata* genes, *pannier*, *serpent* (*srp*), *grain* (*grn*), *gatad* and *gatae*. *pnr*, *grn* and *srp* are part of a cluster of genes on chromosome three (Okumura et al., 2005). Among that *Drosophila* GATA factors, *srp*, *grn* and *gatae* are expressed in the endoderm. In *Drosophila* only the midgut is derived from endoderm, whereas the fore- and hindguts are ectodermal derivatives. Of the three genes, *srp* is the earliest expressing GATA factor, first detected in the prospective endoderm and ceases to be expressed before obvious midgut differentiation. In addition to the endoderm, *srp* is also expressed in the vitellophages, hemocyte primordium, amnioserosa and fat body precursors, reflecting its numerous functions in *Drosophila* development (Rehorn et al., 1996; Sam et al., 1996) (Table 0.1). Two isoforms of Srp exist, one containing two zinc fingers and the other with one zinc finger. Both isoforms have the same expression pattern, though they appear to regulate genes differently (Waltzer et al., 2002). *srp* mutants exhibited a midgut to fore- and hindgut transformation (Reuter, 1994).

Drosophila gatae expression begins in the endoderm at stage 8 and continues in the midgut throughout the life of the animal. In addition, it is also expressed in the malpighian tubules. *gatae* mutants displayed no obvious morphological defects, but lacked expression of many midgut specific genes (Okumura et al., 2005). *grn* is expressed in the developing endoderm, head, posterior spiracles and central nervous system, though *grn* mutants had no discernable endodermal phenotype (Brown and Castelli-Gair Hombria, 2000; Lin et al., 1995).

Drosophila endodermal transcriptional program begins with the gap gene *hkb*, which itself is downstream of Torso RTK signaling. It activates *srp*, which in turn activates *gatae* in a sequential manner (Okumura et al., 2005) (Fig. 0.6). In addition, Srp activates zygotic *caudal* and *hnf4* expression in the endoderm (Reuter, 1994). Both *srp* and *gatae* have been demonstrated to be upstream of intestinal differentiation genes,

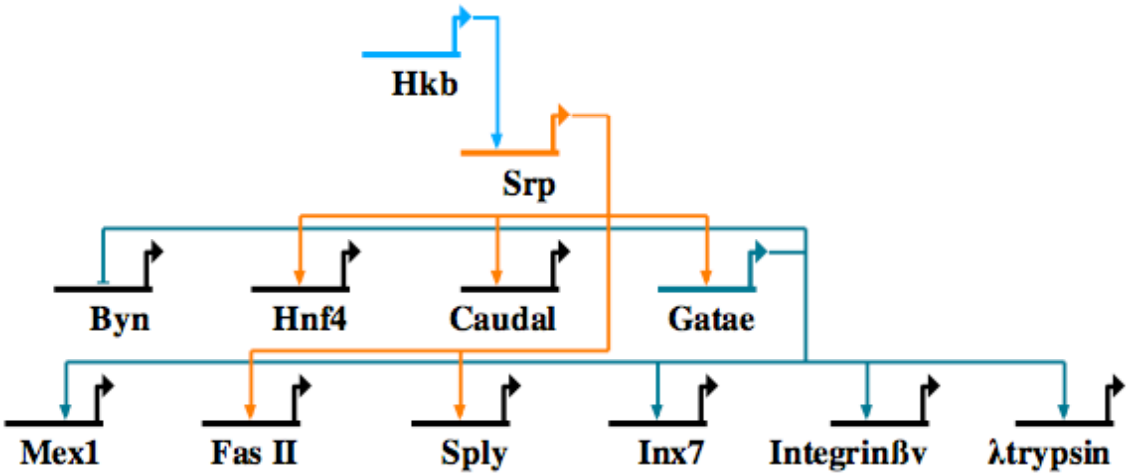


Figure 0.6. Network model outlining gene interactions underlying endoderm specification in the fruit fly *Drosophila melanogaster*. Gene abbreviations are as follows: Byn, *brachyenteron*; Cau, *caudal*; Fas II, *fasciclin*; Hkb, *huckebein*; Inx7, *innexin*; Sply, *Sphingosine-1-phosphate Lyase*; Srp, *serpent.*

though the precise epistatic relationships remain unclear. In case in the *sply* gene, *srp* regulation is not mediated by Gatae, suggesting that both *srp* and *gatae* directly activate differentiation genes (Okumura et al., 2005; Okumura et al., 2007). In addition to activation of endodermal differentiation genes, Gatae also plays a role in repressing the expression of *brachyenteron*, normally expressed in the hindgut, from the midgut (Okumura et al., 2005). Furthermore, Gatae has also been shown to directly activate immune genes in the intestine of the Drosophila larva (Senger et al., 2006).

Sea Urchin

The purple sea urchin *Strongylocentrotus purpuratus* has two GATA factors, *gatae*, orthologous to *gata4/5/6,* and *gatac*, orthologous to *gata1/2/3* (Pancer et al., 1999). *gatac* is expressed in a subset of mesoderm cells known as blastocoelar cells in the blastula and in the adult coelomyctes (Pancer et al., 1999 and J.P. Rast, unpublished data). *gatae* is expressed throughout embryogeneis, first detected in the prospective secondary mesenchyme cells in the 15 h blastula. By the mesenchyme blastula it is expressed in both endoderm and mesoderm, and in the gastrulating gut and mesoderm at the tip of the archenteron until the end of embryogenesis (Lee and Davidson, 2004). Perturbation of *gatae* translation by MASO injection resulted in embryos that failed to gastrulate and which exhibited severely compromised organization of the endomesoderm (Table 0.1). Even so, pigment expressing cells, presumably of mesodermal origin were still visible. *In situ* hybridizations on MASO injected embryos showed that the disorganized cells within the blastocoel expressed endoderm and mesoderm markers.

To determine the position of *gatae* in the endomesoderm gene network, fertilized eggs were injected with *gatae* MASO and the expression of genes were quantified in 18 h and 24 h embryos. Figure 0.7 is a graphical representation of the genes whose expression in response of *gatae* perturbation had been analyzed. A large number of endomesoderm expressing genes are downstream of *gatae*, including many transcription factors. In addition to Gatae's role as an activator of other endomesoderm transcription factors, it is also engaged in a cross-regulatory loop with *otx*, functioning to lock down the endomesoderm specification program (Davidson et al., 2002a; Davidson et al., 2002b) (Fig. 0.8). This interaction has been verified at the DNA level through binding site mutations (Lee and Davidson, 2007; Yuh et al., 2004). Differentiation genes, such as *apobec* and *kakapo*, have also been demonstrated to be downstream of *gatae*, though it is not clear if Gatae regulates them directly or via *brachyury*.

GATA factors as regulators of endoderm transcription factors

In all organisms studied, GATA factors are widely expressed in both the mesoderm and endoderm. Attempts to determine function based on phenotype alone were challenging in many organisms due to the multitude and severity of defects in the mutants or knockdowns. Even though the states of understanding of endoderm specification in the different organisms surveyed are at different levels of completion, all the GRNs have one thing in common: GATA factors are activators of other endoderm specific transcription factors. Examples of genes encoding transcription factors downstream of the GATA factors include the *hnf3/forkhead* genes, Sox class of HMG

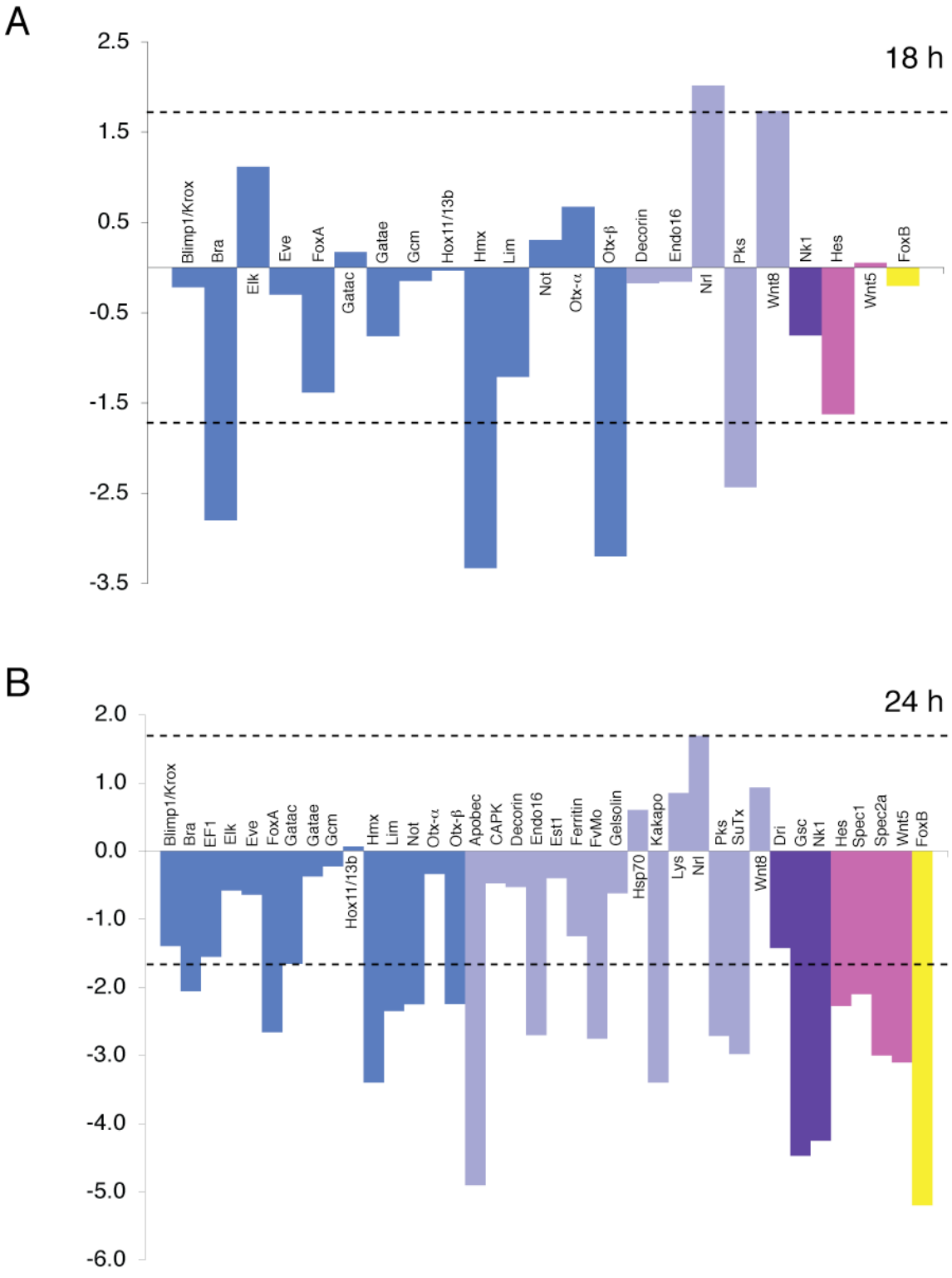

A
18 h
2.5
1.5
0.5
-0.5
-1.5
-2.5
-3.5
Blimp1/Krox
Bra
Elk
Eve
FoxA
Gatac
Gatae
Gcm
Hox11/13b
Hmx
Lim
Not
Otx-α
Otx-β
Decorin
Endo16
Nrl
Pks
Wnt8
Nk1
Hes
Wnt5
FoxB
B
24 h
2.0
1.0
0.0
-1.0
-2.0
-3.0
-4.0
-5.0
-6.0
Blimp1/Krox
Bra
EF1
Elk
Eve
FoxA
Gatac
Gatae
Gcm
Hox11/13b
Hmx
Lim
Not
Otx-α
Otx-β
Apobec
CAPK
Decorin
Endo16
Est1
Ferritin
FvMo
Gelsolin
Hsp70
Kakapo
Lys
Nrl
Pks
SuTx
Wnt8
Dri
Gsc
Nk1
Hes
Spec1
Spec2a
Wnt5
FoxB

Figure 0.7. Graphs depicting the expression pattern of genes in (A) 18 h, and (B) 24 h embryos after injection with an *Spgatae* MASO. RNA was extracted and QPCR performed on embryos injected with either *gatae* or control MASO. ΔC_t was calculated from *gatae* vs. control MASO injected embryos after normalization to ubiquitin to account for differing embryo numbers in each reaction. C_t is defined as the cycle number at a threshold when the PCR product is accumulating exponentially. The two dashed lines (positive and negative) indicate a ΔC_t of 1.7 between *gatae* and control MASO injected embryos. Due to variances between embryo batches, only perturbation effects at or above this level are considered to be significant. The genes are grouped into the following categories based on the color of the bars: blue, endomesoderm transcription factors; lavender, endomesoderm differentiation genes; purple, oral ectoderm genes, pink, ectoderm (other) genes; yellow, skeletogenic mesenchyme gene.

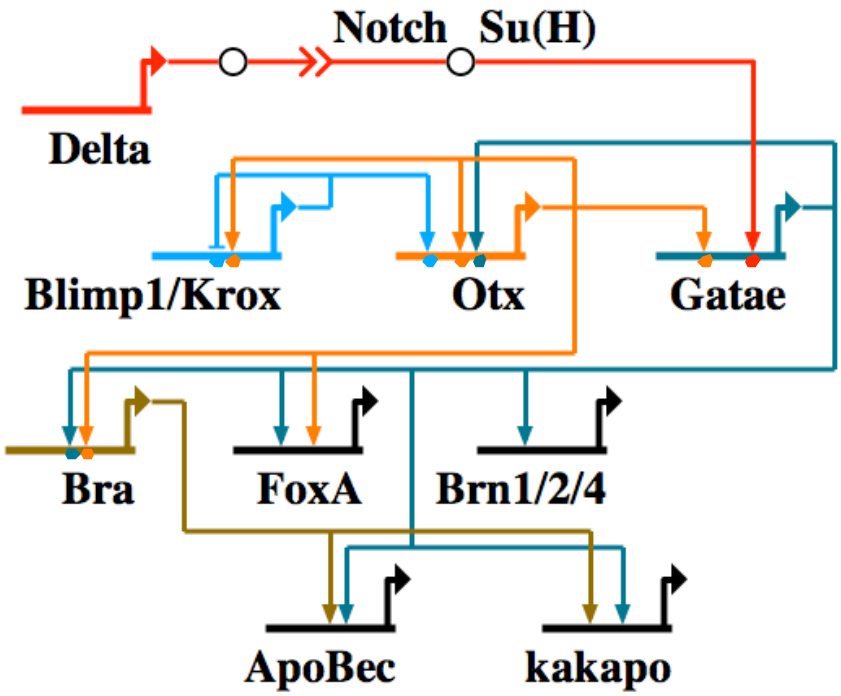


Figure 0.8. Network model outlining gene interactions underlying endoderm specification in the sea urchin *Strongylocentrotus purpuratus*. Abbreviations for genes are as follows: Bra, *brachyury*; Brn, *brain*.

transcription factors, homeodomain proteins such as Nkx, Hex, Otx, Lim and Caudal. In addition, GATA factors also play a role in the repression of non-endodermal states, such as the repression of *byn* in *Drosophila*.

During development transcriptional domains are initiated and progressively refined, and in every organism the GATA factors function near the top of this hierarchy, often the immediate downstream target of a maternal transcription factor or in direct response to signaling toggle switches. A few examples of signaling cascades that they are under the control of include Nodal in vertebrates (Feldman et al., 1998), Wnt in *C. elegans* (Maduro et al., 2005b), and Notch in the sea urchin (Lee and Davidson, 2007). GATA factors are not exclusive regulators of other transcription factors however, they have also been shown to directly activate the expression of differentiation genes, particularly in later development and the adult animal (Bossard and Zaret, 1998; Boudreau et al., 2002).

GATA factors as stabilizers of the endoderm transcriptional state

An interesting observation is that, at least in the deuterostome lineage, GATA factors are involved in cross-regulatory loops with other early acting transcription factors like to facilitate the "lockdown" of the endodermal transcriptional states. These feedback loops involve *gata4* and *sox7* in the mouse (Futaki et al., 2004; Murakami et al., 2004), *gata5* and *cas* in the zebrafish (Alexander et al., 1999; Kikuchi et al., 2001), *gata5/6* and *sox17* in *Xenopus* (Sinner et al., 2006) and *gatae* and *otx* in the sea urchin (Davidson et al., 2002a; Davidson et al., 2002b). While the partners for GATA factors in these cross-

regulatory loops are different for each organism, it is notable that in the vertebrates, they all involved members of the Sox family of transcription factors. Such feedback loops are not observed in the two ecdysozoans *Drosophila* and *C. elegans*, and may reflect a newly acquired function of the GATA factors in the deuterostome lineage, although it is also possible that those connections have simply not been identified in the ecdysozoa. However, *Drosophila* and *C. elegans* both have multiple GATA factors that are activated sequentially, and which could not be grouped into either the 1/2/3 or 4/5/6 GATA sub-families by phylogenetic analysis. This observation is also reflected in the function of these factors. For example, *Drosophila srp* is involved in both hematopoiesis and endoderm development, a feature not observed in deuterstome GATAs.

Summary

GATA factors play functionally conserved roles in setting up the endodermal transcriptional state through its activation of other transcription factors. Furthermore, in the deuterostomes they are also participants in cross-regulatory feedback loops that serve to "lockdown" the developmental program. These networks are by no means complete even in the best studied case of the sea urchin. However, even with the limited information available it is still possible to draw conclusions regarding the function of GATA factors in endoderm development.

Why do GATA factors function near the top of the developmental gene regulatory hierarchy in every organism studied? One possible explanation is that the protostome/deuterostome ancestor utilized GATA factors in such a fashion, and this

function has been conserved and fixed in the GRNs over time. Another possibility is that properties of the GATA protein, for example, its ability to interact with multiple co-factors to affect gene expression in subtle ways, are important for and lead to their active selection as pan endodermesodermal activators. As more details regarding endoderm specification become available, we will be able to better understand not just the developmental process, but also how the regulatory circuitries have evolved in the Bilateria.

List of transcription factors represented in GRNs

Transcription Factor	Family
Bon	Mix homeodomain
Bra/Byn/Ntl	T box
Blimp1/Krox	Zinc finger
Brn-1/2/4	POU homeodomain
Cas	Sox HMG
Caudal/Pal-1	Homeodomain
End-1/3	GATA zinc finger
Elt-2/7	GATA zinc finger
Eomes	T box
Fkd2	Forkhead
GATA4/5/6/Gatae (Sp)	GATA zinc finger
Gatae (Dm)	GATA zinc finger
Hbox8	Homeodomain
Hex	Homeodomain
Hkb	Zinc finger
Hnf1α/β	POU-homeodomain
Hnf3β/FoxA/Pha-4	Forkhead
Hnf4	Orphan nuclear receptor
Lim	Lim homeodomain
Med-1/2	GATA zinc finger
Mezzo	Mix homeodomain
Nkx2.3	Nk-2 homeodomain
Pop-1	HMG
Skn-1	bZIP
Sox 7	Sox HMG
Sox17	Sox HMG
Srp	GATA zinc finger
Tbx6	T box
VegT	T box

References

Afouda, B.A., Ciau-Uitz, A. and Patient, R., 2005. GATA4, 5 and 6 mediate TGFbeta maintenance of endodermal gene expression in Xenopus embryos. Development 132, 763-74.

Alexander, J., Rothenberg, M., Henry, G.L. and Stainier, D.Y., 1999. casanova plays an early and essential role in endoderm formation in zebrafish. Dev Biol 215, 343-57.

Aoki, T.O., David, N.B., Minchiotti, G., Saint-Etienne, L., Dickmeis, T., Persico, G.M., Strahle, U., Mourrain, P. and Rosa, F.M., 2002. Molecular integration of casanova in the Nodal signalling pathway controlling endoderm formation. Development 129, 275-86.

Arceci, R.J., King, A.A., Simon, M.C., Orkin, S.H. and Wilson, D.B., 1993. Mouse GATA-4: a retinoic acid-inducible GATA-binding transcription factor expressed in endodermally derived tissues and heart. Mol Cell Biol 13, 2235-46.

Bjornson, C.R., Griffin, K.J., Farr, G.H., 3rd, Terashima, A., Himeda, C., Kikuchi, Y. and Kimelman, D., 2005. Eomesodermin is a localized maternal determinant required for endoderm induction in zebrafish. Dev Cell 9, 523-33.

Bossard, P. and Zaret, K.S., 1998. GATA transcription factors as potentiators of gut endoderm differentiation. Development 125, 4909-17.

Boudreau, F., Rings, E.H., van Wering, H.M., Kim, R.K., Swain, G.P., Krasinski, S.D., Moffett, J., Grand, R.J., Suh, E.R. and Traber, P.G., 2002. Hepatocyte nuclear factor-1 alpha, GATA-4, and caudal related homeodomain protein Cdx2 interact functionally to modulate intestinal gene transcription. Implication for the developmental regulation of the sucrase-isomaltase gene. J Biol Chem 277, 31909-17.

Broitman-Maduro, G., Maduro, M.F. and Rothman, J.H., 2005. The noncanonical binding site of the MED-1 GATA factor defines differentially regulated target genes in the C. elegans mesendoderm. Dev Cell 8, 427-33.

Brown, S. and Castelli-Gair Hombria, J., 2000. Drosophila grain encodes a GATA transcription factor required for cell rearrangement during morphogenesis. Development 127, 4867-76.

Calvo, D., Victor, M., Gay, F., Sui, G., Luke, M.P., Dufourcq, P., Wen, G., Maduro, M., Rothman, J. and Shi, Y., 2001. A POP-1 repressor complex restricts inappropriate cell type-specific gene transcription during Caenorhabditis elegans embryogenesis. Embo J 20, 7197-208.

Capo-Chichi, C.D., Rula, M.E., Smedberg, J.L., Vanderveer, L., Parmacek, M.S., Morrisey, E.E., Godwin, A.K. and Xu, X.X., 2005. Perception of differentiation cues by GATA factors in primitive endoderm lineage determination of mouse embryonic stem cells. Dev Biol 286, 574-86.

Clements, D. and Woodland, H.R., 2003. VegT induces endoderm by a self-limiting mechanism and by changing the competence of cells to respond to TGF-beta signals. Dev Biol 258, 454-63.

Davidson, E.H., Rast, J.P., Oliveri, P., Ransick, A., Calestani, C., Yuh, C.H., Minokawa, T., Amore, G., Hinman, V., Arenas-Mena, C., Otim, O., Brown, C.T., Livi, C.B., Lee, P.Y., Revilla, R., Rust, A.G., Pan, Z., Schilstra, M.J., Clarke, P.J., Arnone, M.I., Rowen, L., Cameron, R.A., McClay, D.R., Hood, L. and Bolouri, H., 2002a. A genomic regulatory network for development. Science 295, 1669-78.

Davidson, E.H., Rast, J.P., Oliveri, P., Ransick, A., Calestani, C., Yuh, C.H., Minokawa, T., Amore, G., Hinman, V., Arenas-Mena, C., Otim, O., Brown, C.T., Livi, C.B., Lee, P.Y., Revilla, R., Schilstra, M.J., Clarke, P.J., Rust, A.G., Pan, Z., Arnone, M.I., Rowen, L., Cameron, R.A., McClay, D.R., Hood, L. and Bolouri, H., 2002b. A provisional regulatory gene network for specification of endomesoderm in the sea urchin embryo. Dev Biol 246, 162-90.

Evans, T. and Felsenfeld, G., 1989. The erythroid-specific transcription factor Eryf1: a new finger protein. Cell 58, 877-85.

Evans, T., Reitman, M. and Felsenfeld, G., 1988. An erythrocyte-specific DNA-binding factor recognizes a regulatory sequence common to all chicken globin genes. Proc Natl Acad Sci U S A 85, 5976-80.

Feldman, B., Gates, M.A., Egan, E.S., Dougan, S.T., Rennebeck, G., Sirotkin, H.I., Schier, A.F. and Talbot, W.S., 1998. Zebrafish organizer development and germ-layer formation require nodal-related signals. Nature 395, 181-5.

Fujikura, J., Yamato, E., Yonemura, S., Hosoda, K., Masui, S., Nakao, K., Miyazaki Ji, J. and Niwa, H., 2002. Differentiation of embryonic stem cells is induced by GATA factors. Genes Dev 16, 784-9.

Fukushige, T., Goszczynski, B., Tian, H. and McGhee, J.D., 2003. The evolutionary duplication and probable demise of an endodermal GATA factor in Caenorhabditis elegans. Genetics 165, 575-88.

Fukushige, T., Goszczynski, B., Yan, J. and McGhee, J.D., 2005. Transcriptional control and patterning of the pho-1 gene, an essential acid phosphatase expressed in the C. elegans intestine. Dev Biol 279, 446-61.

Fukushige, T., Hawkins, M.G. and McGhee, J.D., 1998. The GATA-factor elt-2 is essential for formation of the Caenorhabditis elegans intestine. Dev Biol 198, 286-302.

Fukushige, T., Hendzel, M.J., Bazett-Jones, D.P. and McGhee, J.D., 1999. Direct visualization of the elt-2 gut-specific GATA factor binding to a target promoter inside the living Caenorhabditis elegans embryo. Proc Natl Acad Sci U S A 96, 11883-8.

Futaki, S., Hayashi, Y., Emoto, T., Weber, C.N. and Sekiguchi, K., 2004. Sox7 plays crucial roles in parietal endoderm differentiation in F9 embryonal carcinoma cells through regulating Gata-4 and Gata-6 expression. Mol Cell Biol 24, 10492-503.

Gao, X., Sedgwick, T., Shi, Y.B. and Evans, T., 1998. Distinct functions are implicated for the GATA-4, -5, and -6 transcription factors in the regulation of intestine epithelial cell differentiation. Mol Cell Biol 18, 2901-11.

Gritsman, K., Zhang, J., Cheng, S., Heckscher, E., Talbot, W.S. and Schier, A.F., 1999. The EGF-CFC protein one-eyed pinhead is essential for nodal signaling. Cell 97, 121-32.

Hilton, E., Rex, M. and Old, R., 2003. VegT activation of the early zygotic gene Xnr5 requires lifting of Tcf-mediated repression in the Xenopus blastula. Mech Dev 120, 1127-38.

Hinman, V., pers. comm. personal communication.

Holtzinger, A. and Evans, T., 2005. Gata4 regulates the formation of multiple organs. Development 132, 4005-14.

Jacobsen, C.M., Narita, N., Bielinska, M., Syder, A.J., Gordon, J.I. and Wilson, D.B., 2002. Genetic mosaic analysis reveals that GATA-4 is required for proper differentiation of mouse gastric epithelium. Dev Biol 241, 34-46.

Kalb, J.M., Lau, K.K., Goszczynski, B., Fukushige, T., Moons, D., Okkema, P.G. and McGhee, J.D., 1998. pha-4 is Ce-fkh-1, a fork head/HNF-3alpha,beta,gamma homolog that functions in organogenesis of the C. elegans pharynx. Development 125, 2171-80.

Kelley, C., Blumberg, H., Zon, L.I. and Evans, T., 1993. GATA-4 is a novel transcription factor expressed in endocardium of the developing heart. Development 118, 817-27.

Kikuchi, Y., Agathon, A., Alexander, J., Thisse, C., Waldron, S., Yelon, D., Thisse, B. and Stainier, D.Y., 2001. casanova encodes a novel Sox-related protein necessary and sufficient for early endoderm formation in zebrafish. Genes Dev 15, 1493-505.

Ko, L.J. and Engel, J.D., 1993. DNA-binding specificities of the GATA transcription factor family. Mol Cell Biol 13, 4011-22.

Koutsourakis, M., Langeveld, A., Patient, R., Beddington, R. and Grosveld, F., 1999. The transcription factor GATA6 is essential for early extraembryonic development. Development 126, 723-32.

Kuo, C.T., Morrisey, E.E., Anandappa, R., Sigrist, K., Lu, M.M., Parmacek, M.S., Soudais, C. and Leiden, J.M., 1997. GATA4 transcription factor is required for ventral morphogenesis and heart tube formation. Genes Dev 11, 1048-60.

Lee, P.Y. and Davidson, E.H., 2004. Expression of Spgatae, the Strongylocentrotus purpuratus ortholog of vertebrate GATA4/5/6 factors. Gene Expr Patterns 5, 161-5.

Lee, P.Y. and Davidson, E.H., 2007. Use of OR logic in the regulation of a sea urchin GATA factor (in preparation).

Lin, W.H., Huang, L.H., Yeh, J.Y., Hoheisel, J., Lehrach, H., Sun, Y.H. and Tsai, S.F., 1995. Expression of a Drosophila GATA transcription factor in multiple tissues in the developing embryos. Identification of homozygous lethal mutants with P-element insertion at the promoter region. J Biol Chem 270, 25150-8.

Longabaugh, W.J., Davidson, E.H. and Bolouri, H., 2005. Computational representation of developmental genetic regulatory networks. Dev Biol 283, 1-16.

Lowry, J.A. and Atchley, W.R., 2000. Molecular evolution of the GATA family of transcription factors: conservation within the DNA-binding domain. J Mol Evol 50, 103-15.

Maduro, M.F., Broitman-Maduro, G., Mengarelli, I. and Rothman, J.H., 2006. Maternal deployment of the embryonic SKN-1-->MED-1,2 cell specification pathway in C. elegans. Dev Biol.

Maduro, M.F., Hill, R.J., Heid, P.J., Newman-Smith, E.D., Zhu, J., Priess, J.R. and Rothman, J.H., 2005a. Genetic redundancy in endoderm specification within the genus Caenorhabditis. Dev Biol 284, 509-22.

Maduro, M.F., Kasmir, J.J., Zhu, J. and Rothman, J.H., 2005b. The Wnt effector POP-1 and the PAL-1/Caudal homeoprotein collaborate with SKN-1 to activate C. elegans endoderm development. Dev Biol 285, 510-23.

Maduro, M.F., Meneghini, M.D., Bowerman, B., Broitman-Maduro, G. and Rothman, J.H., 2001. Restriction of mesendoderm to a single blastomere by the combined action of SKN-1 and a GSK-3beta homolog is mediated by MED-1 and -2 in C. elegans. Mol Cell 7, 475-85.

Maduro, M.F. and Rothman, J.H., 2002. Making worm guts: the gene regulatory network of the Caenorhabditis elegans endoderm. Dev Biol 246, 68-85.

Merika, M. and Orkin, S.H., 1993. DNA-binding specificity of GATA family transcription factors. Mol Cell Biol 13, 3999-4010.

Moilanen, L.H., Fukushige, T. and Freedman, J.H., 1999. Regulation of metallothionein gene transcription. Identification of upstream regulatory elements and transcription factors responsible for cell-specific expression of the metallothionein genes from Caenorhabditis elegans. J Biol Chem 274, 29655-65.

Molkentin, J.D., Tymitz, K.M., Richardson, J.A. and Olson, E.N., 2000. Abnormalities of the genitourinary tract in female mice lacking GATA5. Mol Cell Biol 20, 5256-60.

Morrisey, E.E., Ip, H.S., Lu, M.M. and Parmacek, M.S., 1996. GATA-6: a zinc finger transcription factor that is expressed in multiple cell lineages derived from lateral mesoderm. Dev Biol 177, 309-22.

Morrisey, E.E., Ip, H.S., Tang, Z., Lu, M.M. and Parmacek, M.S., 1997a. GATA-5: a transcriptional activator expressed in a novel temporally and spatially-restricted pattern during embryonic development. Dev Biol 183, 21-36.

Morrisey, E.E., Ip, H.S., Tang, Z. and Parmacek, M.S., 1997b. GATA-4 activates transcription via two novel domains that are conserved within the GATA-4/5/6 subfamily. J Biol Chem 272, 8515-24.

Morrisey, E.E., Tang, Z., Sigrist, K., Lu, M.M., Jiang, F., Ip, H.S. and Parmacek, M.S., 1998. GATA6 regulates HNF4 and is required for differentiation of visceral endoderm in the mouse embryo. Genes Dev 12, 3579-90.

Murakami, A., Shen, H., Ishida, S. and Dickson, C., 2004. SOX7 and GATA-4 are competitive activators of Fgf-3 transcription. J Biol Chem 279, 28564-73.

Narita, N., Heikinheimo, M., Bielinska, M., White, R.A. and Wilson, D.B., 1996. The gene for transcription factor GATA-6 resides on mouse chromosome 18 and is expressed in myocardium and vascular smooth muscle. Genomics 36, 345-8.

Nemer, G., Qureshi, S.T., Malo, D. and Nemer, M., 1999. Functional analysis and chromosomal mapping of Gata5, a gene encoding a zinc finger DNA-binding protein. Mamm Genome 10, 993-9.

Okumura, T., Matsumoto, A., Tanimura, T. and Murakami, R., 2005. An endoderm-specific GATA factor gene, dGATAe, is required for the terminal differentiation of the Drosophila endoderm. Dev Biol 278, 576-86.

Okumura, T., Tajiri, R., Kojima, T., Saigo, K. and Murakami, R., 2007. GATAe-dependent and -independent expressions of genes in the differentiated endodermal midgut of Drosophila. Gene Expr Patterns 7, 178-86.

Omichinski, J.G., Clore, G.M., Schaad, O., Felsenfeld, G., Trainor, C., Appella, E., Stahl, S.J. and Gronenborn, A.M., 1993a. NMR structure of a specific DNA complex of Zn-containing DNA binding domain of GATA-1. Science 261, 438-46.

Omichinski, J.G., Trainor, C., Evans, T., Gronenborn, A.M., Clore, G.M. and Felsenfeld, G., 1993b. A small single-"finger" peptide from the erythroid transcription factor GATA-1 binds specifically to DNA as a zinc or iron complex. Proc Natl Acad Sci U S A 90, 1676-80.

Orkin, S.H. and Zon, L.I., 1997. Genetics of erythropoiesis: induced mutations in mice and zebrafish. Annu Rev Genet 31, 33-60.

Oskouian, B., Mendel, J., Shocron, E., Lee, M.A., Jr., Fyrst, H. and Saba, J.D., 2005. Regulation of sphingosine-1-phosphate lyase gene expression by members of the GATA family of transcription factors. J Biol Chem 280, 18403-10.

Pancer, Z., Rast, J.P. and Davidson, E.H., 1999. Origins of immunity: transcription factors and homologues of effector genes of the vertebrate immune system expressed in sea urchin coelomocytes. Immunogenetics 49, 773-86.

Peterkin, T., Gibson, A. and Patient, R., 2003. GATA-6 maintains BMP-4 and Nkx2 expression during cardiomyocyte precursor maturation. Embo J 22, 4260-73.

Poulain, M. and Lepage, T., 2002. Mezzo, a paired-like homeobox protein is an immediate target of Nodal signalling and regulates endoderm specification in zebrafish. Development 129, 4901-14.

Rehorn, K.P., Thelen, H., Michelson, A.M. and Reuter, R., 1996. A molecular aspect of hematopoiesis and endoderm development common to vertebrates and Drosophila. Development 122, 4023-31.

Reiter, J.F., Alexander, J., Rodaway, A., Yelon, D., Patient, R., Holder, N. and Stainier, D.Y., 1999. Gata5 is required for the development of the heart and endoderm in zebrafish. Genes Dev 13, 2983-95.

Reiter, J.F., Kikuchi, Y. and Stainier, D.Y., 2001. Multiple roles for Gata5 in zebrafish endoderm formation. Development 128, 125-35.

Reuter, R., 1994. The gene serpent has homeotic properties and specifies endoderm versus ectoderm within the Drosophila gut. Development 120, 1123-35.

Rodaway, A., Takeda, H., Koshida, S., Broadbent, J., Price, B., Smith, J.C., Patient, R. and Holder, N., 1999. Induction of the mesendoderm in the zebrafish germ ring by yolk cell-derived TGF-beta family signals and discrimination of mesoderm and endoderm by FGF. Development 126, 3067-78.

Sam, S., Leise, W. and Hoshizaki, D.K., 1996. The serpent gene is necessary for progression through the early stages of fat-body development. Mech Dev 60, 197-205.

Senger, K., Harris, K. and Levine, M., 2006. GATA factors participate in tissue-specific immune responses in Drosophila larvae. Proc Natl Acad Sci U S A 103, 15957-62.

Sinner, D., Kirilenko, P., Rankin, S., Wei, E., Howard, L., Kofron, M., Heasman, J., Woodland, H.R. and Zorn, A.M., 2006. Global analysis of the transcriptional network controlling Xenopus endoderm formation. Development 133, 1955-66.

Soudais, C., Bielinska, M., Heikinheimo, M., MacArthur, C.A., Narita, N., Saffitz, J.E., Simon, M.C., Leiden, J.M. and Wilson, D.B., 1995. Targeted mutagenesis of the transcription factor GATA-4 gene in mouse embryonic stem cells disrupts visceral endoderm differentiation in vitro. Development 121, 3877-88.

Svensson, E.C., Tufts, R.L., Polk, C.E. and Leiden, J.M., 1999. Molecular cloning of FOG-2: a modulator of transcription factor GATA-4 in cardiomyocytes. Proc Natl Acad Sci U S A 96, 956-61.

Tevosian, S.G., Deconinck, A.E., Cantor, A.B., Rieff, H.I., Fujiwara, Y., Corfas, G. and Orkin, S.H., 1999. FOG-2: A novel GATA-family cofactor related to multitype zinc-finger proteins Friend of GATA-1 and U-shaped. Proc Natl Acad Sci U S A 96, 950-5.

Tsang, A.P., Visvader, J.E., Turner, C.A., Fujiwara, Y., Yu, C., Weiss, M.J., Crossley, M. and Orkin, S.H., 1997. FOG, a multitype zinc finger protein, acts as a cofactor for transcription factor GATA-1 in erythroid and megakaryocytic differentiation. Cell 90, 109-19.

van Wering, H.M., Bosse, T., Musters, A., de Jong, E., de Jong, N., Hogen Esch, C.E., Boudreau, F., Swain, G.P., Dowling, L.N., Montgomery, R.K., Grand, R.J. and Krasinski, S.D., 2004. Complex regulation of the lactase-phlorizin hydrolase promoter by GATA-4. Am J Physiol Gastrointest Liver Physiol 287, G899-909.

Visvader, J.E., Crossley, M., Hill, J., Orkin, S.H. and Adams, J.M., 1995. The C-terminal zinc finger of GATA-1 or GATA-2 is sufficient to induce megakaryocytic differentiation of an early myeloid cell line. Mol Cell Biol 15, 634-41.

Waltzer, L., Bataille, L., Peyrefitte, S. and Haenlin, M., 2002. Two isoforms of Serpent containing either one or two GATA zinc fingers have different roles in Drosophila haematopoiesis. Embo J 21, 5477-86.

Weber, H., Symes, C.E., Walmsley, M.E., Rodaway, A.R. and Patient, R.K., 2000. A role for GATA5 in Xenopus endoderm specification. Development 127, 4345-60.

Xanthos, J.B., Kofron, M., Wylie, C. and Heasman, J., 2001. Maternal VegT is the initiator of a molecular network specifying endoderm in Xenopus laevis. Development 128, 167-80.

Yamada, L., Kobayashi, K., Degnan, B., Satoh, N. and Satou, Y., 2003. A genomewide survey of developmentally relevant genes in Ciona intestinalis. IV. Genes for HMG transcriptional regulators, bZip and GATA/Gli/Zic/Snail. Dev Genes Evol 213, 245-53.

Yang, H.Y. and Evans, T., 1992. Distinct roles for the two cGATA-1 finger domains. Mol Cell Biol 12, 4562-70.

Yuh, C.H., Dorman, E.R., Howard, M.L. and Davidson, E.H., 2004. An otx cis-regulatory module: a key node in the sea urchin endomesoderm gene regulatory network. Dev Biol 269, 536-51.

Zhang, J., Houston, D.W., King, M.L., Payne, C., Wylie, C. and Heasman, J., 1998. The role of maternal VegT in establishing the primary germ layers in Xenopus embryos. Cell 94, 515-24.

Zhu, J., Fukushige, T., McGhee, J.D. and Rothman, J.H., 1998. Reprogramming of early embryonic blastomeres into endodermal progenitors by a Caenorhabditis elegans GATA factor. Genes Dev 12, 3809-14.

Zhu, J., Hill, R.J., Heid, P.J., Fukuyama, M., Sugimoto, A., Priess, J.R. and Rothman, J.H., 1997. end-1 encodes an apparent GATA factor that specifies the endoderm precursor in Caenorhabditis elegans embryos. Genes Dev 11, 2883-96.

CHAPTER 1

Expression of *Spgatae*, the *Strongylocentrotus purpuratus* Ortholog of Vertebrate GATA4/5/6 Factors

Pei Yun Lee and Eric H. Davidson

Published in *Gene Expression Patterns* **5**, 161-5 (2004).

Abstract

Spgatae is the sea urchin ortholog of the vertebrate *gata4/5/6* genes, as confirmed by phylogenetic analysis. The accumulation of *Spgatae* transcripts during embryonic development and the spatial pattern of expression are reported here. Expression was first detected in the 15 h blastula. The number of *Spgatae* RNA molecules increases steadily during blastula stages, with expression peaking during gastrulation. After gastrulation is complete, the level of expression decreases until the end of embryogenesis. Whole mount *in situ* hybridization showed that *Spgatae* transcripts were first detected in a ring of prospective mesoderm cells in the vegetal plate. *Spgatae* expression then expands to include the entire vegetal plate at the mesenchyme blastula stage. During gastrulation *Spgatae* is expressed at the blastopore, and at prism stage strongly in the hindgut and midgut but not foregut, and also in mesoderm cells at the tip of the archenteron. Toward

the end of embryogenesis, expression in the hindgut decreases. The terminal pattern of expression is in midgut plus coelomic pouches.

Keywords: Sea urchin, *Strongylocentrotus purpuratus*, Gata factors, Endomesoderm specification, Endomesoderm gene network, *Spgatae*

RESULTS AND DISCUSSION

GATA factors are a class of DNA binding zinc finger transcriptional regulators which are named for the WGATAR sequence its members recognize. The GATA factors contain class IV zinc fingers which are characterized by the sequence $CX_2CX_{17-18}CX_2C$ (Lowry and Atchley, 2000). There are two vertebrate families of GATA factors. Members of the *gata1/2/3* family play important roles in hematopoiesis (Bockamp et al., 1994), while members of the *gata4/5/6* family are indispensable for endoderm and mesoderm development. In the zebrafish, *gata5* is required for heart and endoderm development (Reiter et al., 1999). In *Caenorhabditis elegans*, pairs of GATA transcription factors, encoded by the *med*, *end*, and *elt* genes, operate sequentially in the specification of the EMS and then the endoderm lineages (reviewed by Maduro and Rothman, 2002). In *Drosophila* the *gata* gene *serpent* is involved in gut specification (Reuter, 1994). Genes of the *gata* family and their orthologs thus operate in the process of endoderm specification and development across the Bilateria.

Two GATA factors were identified in the sea urchin (Pancer et al., 1999). *Spgatac* is expressed in coelomycytes (Pancer et al., 1999), and this gene is an ortholog of the vertebrate *gata1/2/3* genes, while *Spgatae* is the ortholog of the vertebrate *gata4/5/6* genes. Functional analysis of *Spgatae* had shown that it plays an important role in endomesoderm specification during sea urchin embryogenesis. Perturbation of *Spgatae* expression using a morpholino antisense oligonucleotide revealed many genes which are positioned downstream of *Spgatae* in the endomesoderm gene regulatory network. The SpGatae transcription factor proves to be an important regulator of the

expression of many other endoderm and mesoderm regulatory genes; prominent examples include the *Spbrachyury* and *Spfoxa* genes. Most importantly, gene network analysis has shown that *Spgatae* participates in a cross regulatory loop that also involves *Spkrox* and *SpOtx*. The function of this gene regulatory loop is to lock down the expression of these three genes and drive the process of development forward (Davidson et al., 2002a; Davidson et al., 2002b).

Spgatae was initially isolated from a 48 h *S. purpuratus* cDNA library (Pancer et al., 1999). Identification and sequencing of a full length cDNA clone (Genbank accession No. AY623814) and mRNA blot hybridization showed that the length of the transcript is 4.3 kb (data not shown). Both the appearance of a single band on the mRNA blot (data not shown) and the search of an *S. purpuratus* EST catalog (Poustka et al., 2003) suggest that there are no alternative splice variants of *Spgatae*. A search of the traces from the *S. purpuratus* genome project did not identify any other paralogs. This, combined with the information that screens of *S. purpuratus* BAC and cDNA libraries identified no *gata4/5/6* sequences other than *Spgatae* clones, suggest that it is the only sea urchin ortholog of the vertebrate *gata4/5/6* genes.

A conceptual translation of the coding sequence revealed a protein that contains 567 amino acids (Fig. 1.1A) and two class IV zinc fingers. The zinc finger regions of *Spgatae* were aligned with those of various other GATA proteins, including *Asterina miniata gatae*, *Mus musculus gata4/5/6*, *Drosophila melanogaster pannier* and *Spgatac*, and a phylogenetic tree was constructed using neighbor joining analysis. The phylogenetic analysis confirmed that *Spgatae* is the *S. purpuratus* ortholog to vertebrate *gata4/5/6* genes, *Amgatae*, *Dmpannier* and is paralogous to the vertebrate *gata1/2/3*

A

MPHQVGVSHLQPEHNNSNILLPKEDVEVFFSNLDKNPATGLHQQYLYPQYHLTSESQMYQSASISLQSALQ
PTVSPGGMPPSQQPGYEASPASYIHSSANPVYVPTTRPTFSGMHHPAQFIQHIPAVSSPNHQNPSVIQANA
HAAAAVWSPQSDGSGGGGVVGGDGHHRGYSFPPSPALTTANSPLSGRHPGSTPNGLAGYSPYTDPWSGFDG
SMLHSSMGRAAAAGGNFAGRRPTAEAQMMKNMEGYTAVWPNEYGL**GRECVNCGAISTPLWRRDGTGHYLCN**
ACGLYHKMNGYNRPLIKNPRRLQSGSRREGITCANCHTSTTTLWRRNKDGEPVCNACGLYFKLHGVNRPLA
MKKDGIQTRKRKPKNPNKGNQQSNARNGGGQSSPNDVNIKASSPTGKQPSPLPTSSPYTSHPIKVEPQYRV
GLSPPPITNQVSYIPGLVHHSVPISSAQSLHHHHHHPGSYAAHLGAPPTHISHGINSEQTLHLTHQPGQTI
LLHNGPANSAINPLNLSANTNGGVSHHAGSPGEMTNTDSTSPHHVLFSGVNPSPPSAVAVPVSVSKVDSE*

B

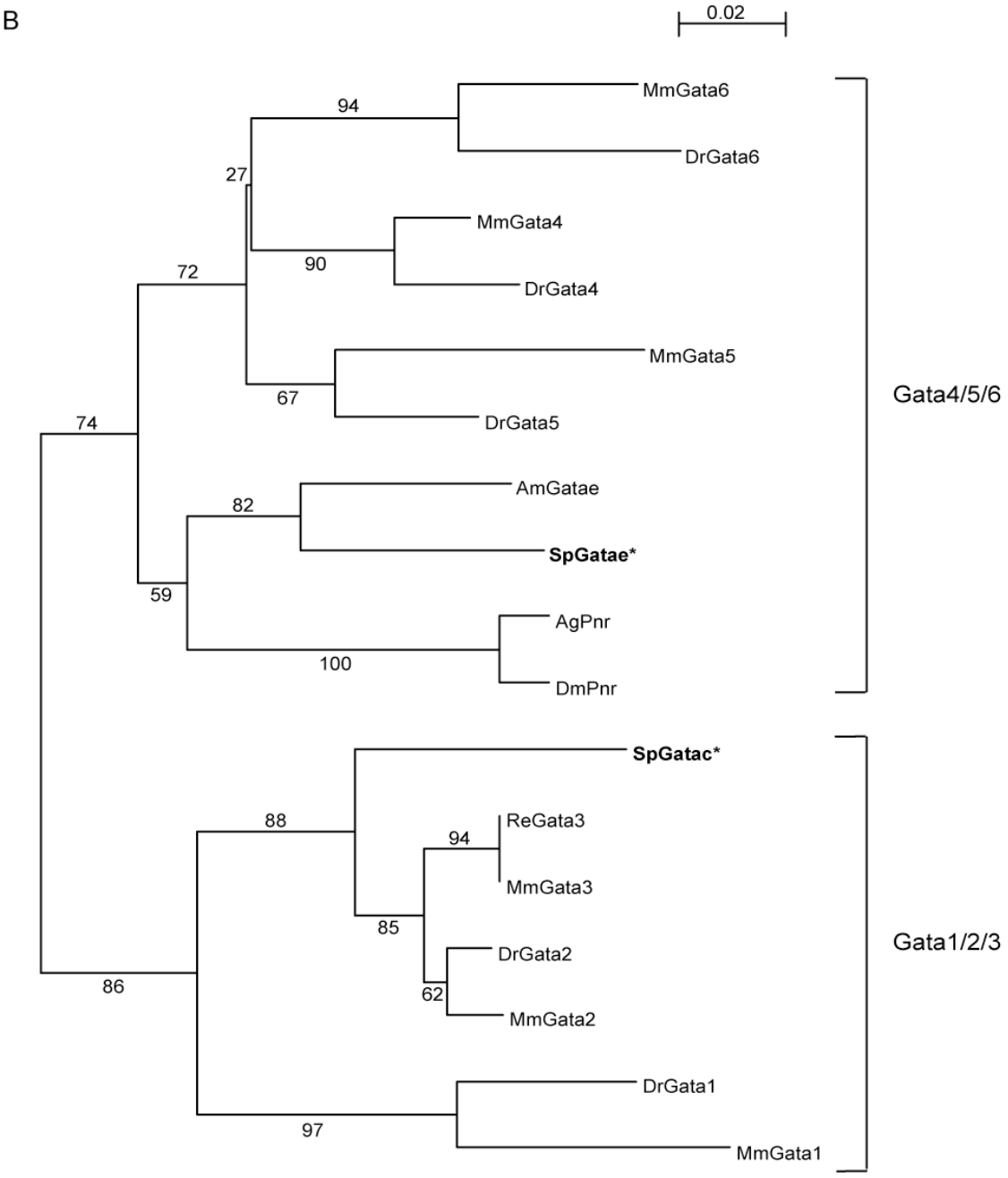


Figure 1.1. (A) Protein sequence of *Spgatae*, obtained by conceptual translation of the DNA coding sequence. The two zinc fingers are boxed. The sequences used for WMISH probes are underlined. Sequence used for phylogenetic analysis to generate the tree in (B) is highlighted in bold. **(B) Phylogenetic analysis showing that *Spgatae* belongs to the *gata4/5/6* family.** Amino acid sequences of the zinc finger regions of various Gata proteins (bold in A) were aligned using ClustalX. The neighbor joining tree was constructed using *Spkrl* as the outgroup, with all gaps excluded. Bootstrap values were calculated based on percentage of 1000 repeated iterations. Ag, *Anopheles gambiae*; Am, *Asterina miniata*; Dm, *Drosophila melanogaster*; Dr, *Danio rerio*; Mm, *Mus musculus*; Re, *Raja eglanteria*.

genes, of which *Spgatac* is the ortholog (Fig. 1.1B).

The temporal expression of *Spgatae* was determined using quantitative real-time PCR (QPCR). RNA was extracted from embryos at developmental stages ranging from the unfertilized egg to the pluteus larva (72 h). cDNA was generated by reverse transcription and QPCR performed using primers specific for *Spgatae*. The results show that *Spgatae* is a zygotic gene first expressed at the 15 h blastula stage (Fig. 1.2). The level of transcripts increases steadily until midgastrula, when it attains a plateau at 3500 copies per embryo. After gastrulation is completed, the expression of *Spgatae* decreases until the end of embryogenesis.

Spatial expression patterns of *Spgatae* were determined by whole mount *in situ* hybridization (WMISH) using the procedure described in Minokawa et al. (2004). To avoid any potential cross hybridization to *SpGatac*, two different digoxygenin labeled probes (N terminal to and C terminal to the zinc fingers) were made (Fig. 1.1A). At the 15 h blastula stage, expression of *Spgatae* was detected in a ring around the vegetal plate that represents the prospective secondary mesenchyme cells (SMCs) (Fig. 1.3, 15 h and

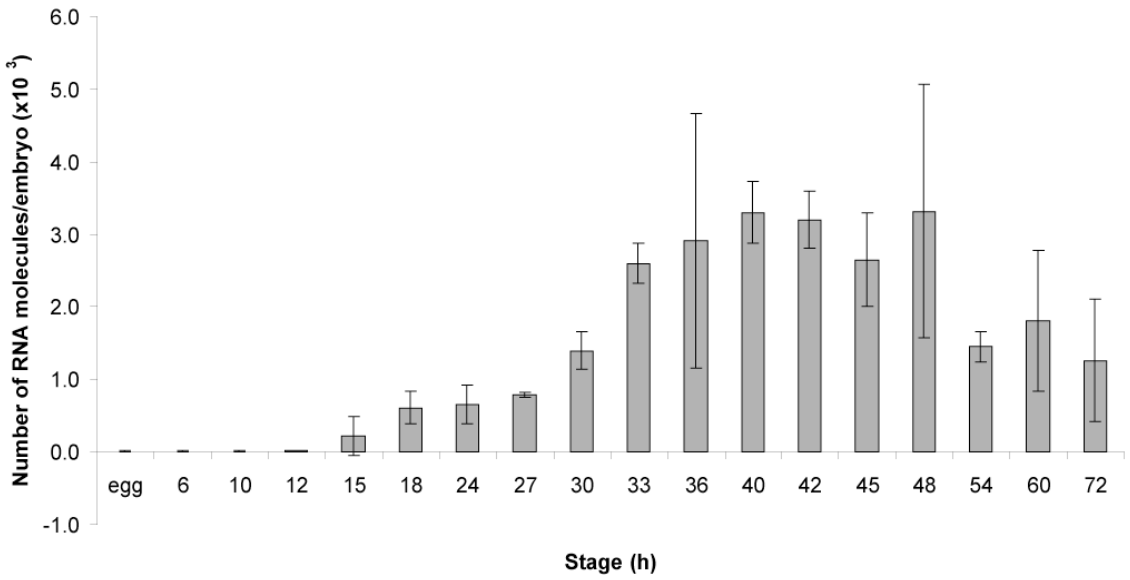


Figure 1.2. Time course of *Spgatae* expression. The time course was determined by QPCR measurements on staged embryo cDNAs. *SpZ12* was used as a standard from which transcript number was determined (Wang et al., 1995). Each time point was obtained from a minimum of five RNA samples from different embryo batches. The error bars (positive and negative) represent one standard deviation from the average value.

15 h V). Expression persists in this region in the 18 h embryo (Fig. 1.3, 18 h and 18 h V). The expression appears to be radially symmetrical in these embryos. At the mesenchyme blastula stage (24 h), expression has expanded across the entire vegetal plate to include both the mesoderm and possibly all the endoderm cells, though *Spgatae* expression in the future SMCs is decreasing at this point (Fig. 1.3, 24 h and 24 h V).

As gastrulation ensues, *Spgatae* is expressed in the invaginating vegetal plate. While the gut elongates through convergent extension movements, *Spgatae* is expressed both at the mesodermal tip of the archenteron and around the blastopore, that is, the posterior invaginating endoderm of the embryo. The gene is not active in the anterior portions of the gut (Fig. 1.3, 33 h). As SMCs delaminate off the tip of the archenteron,

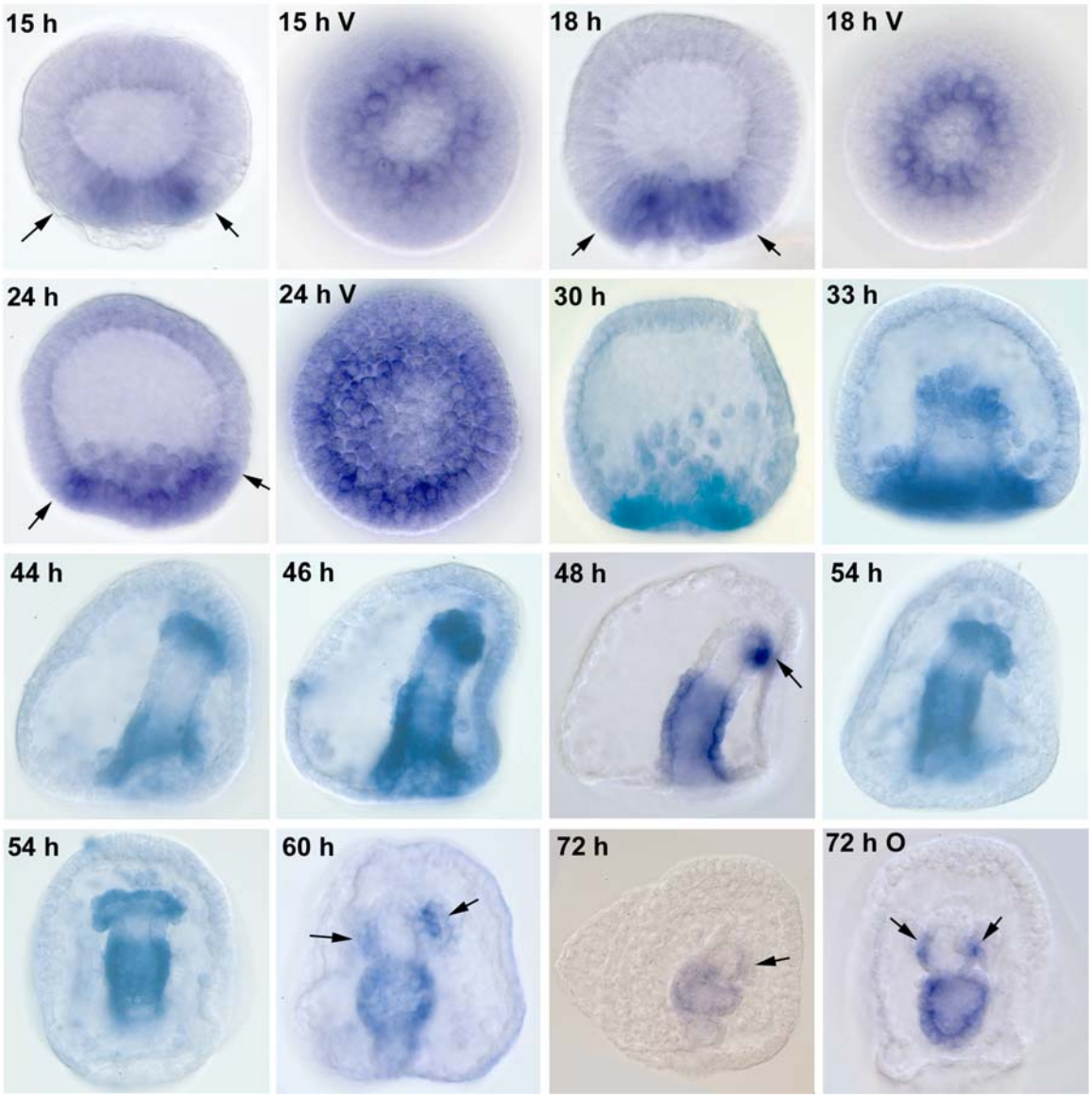


Figure 1.3. Spatial expression of *Spgatae*. WMISH was performed on embryos at various stages of development using a mixture of digoxygenin labeled probes corresponding to exons flanking the Zn fingers (underlined in Fig. 1A). Each panel is labeled with the stage of development. Note the absence of expression in descendant cells of the micromere lineage in 15 h and 18 h blastula embryos. In the 24 h mesenchyme blastula the PMCs have ingressed and *Spgatae* is expressed across the entire vegetal plate. Arrows in the 15 h, 18 h and 24 h embryos mark the boundaries of vegetal plate expression. *Spgatae* is expressed in the gut and SMCs during gastrulation. The arrow in the 48 h embryo points to the tip of the archenteron where the coelomic mesoderm is located. Note the downregulation of *Spgatae* expression in hindgut towards the end of embryogenesis. Arrows point to coelomic pouches in 60 h and 72 h embryos.

Side view of the 72 h embryo shows the expression of *Spgatae* in the midgut only. V, vegetal view; O, oral view.

they cease to express *Spgatae*. At the end of gastrulation, the prism stage (Fig. 1.3, 44-48 h), *Spgatae* is expressed in the mesoderm at the tip of the archenteron and in the midgut and hindgut. This expression persists in the 54 h embryo, where *Spgatae* can clearly be seen to be expressed in the developing coelomic pouches. In the 60 h embryo, hindgut expression begins to decrease. This becomes more evident in the pluteus larva, in which hindgut expression has almost completely disappeared, leaving expression in the coelomic pouches and the stomach (Fig. 1.3, 72 h and 72 h O).

Expression of *Spgatae* is very similar to that of *Amgatae*, with the notable exception that there is no mesoderm expression in the starfish blastula (Hinman and Davidson, 2003). This difference in expression pattern can be explained by differences in the endomesoderm gene regulatory networks of the two organisms. In the starfish, an input from the AmFoxa repressor into the *Amgatae cis*-regulatory element prevents expression in the mesoderm (Hinman et al., 2003). This particular input is absent in the sea urchin, the net result being that *Spgatae* continues to be expressed in the mesoderm (Davidson et al., 2002a; Davidson et al., 2002b). Studies on how the *gatae* genes in these two organisms are regulated will provide direct insight into the process of *cis*-regulatory evolution.

MATERIALS AND METHODS

Spgatae clones were isolated as described (Pancer et al., 1999). The resulting plaques were purified and the phagemid was excised using Stratagene Rapid Excision Kit. The excised clones were selected for size and restriction mapped. The clone containing the largest insert and displaying the correction restriction map based on partial *Spgatae* cDNA sequence was sequenced extensively to generate the sequence in Fig. 1.1.

Embryos were fertilized in filtered sea water and cultured at 15 °C. At various stages embryos were collected and RNA isolated. cDNA was made using ABI's "Taqman" kit. Two sets of QPCR primers for *Spgatae* were designed; one primer set corresponds to exons 3 and 4 and the other primer set corresponds to exons 4 and 5. Both sets of primers were used in the generation of the *Spgatae* time course. On three cDNA batches the primer set corresponding to exons 4 and 5 were used; on the remaining batches both primer sets were used on separate PCR reactions to generate the time course. The results obtained from two primer pairs were similar. *SpZ12* was used as a standard for determination of *Spgatae* transcript numbers (Wang et al., 1995).

ACKNOWLEDGEMENTS

We are grateful to Dr. Jonathan Rast for providing help with the identification of a full length cDNA clone for *Spgatae* and to Dr. Veronica Hinman for help with construction of the phylogenetic tree. We would also like to thank Titus Brown for help with searching the sea urchin genome traces and Pat Leahy at the Kerckhoff Marine lab

for care of the sea urchins. The research described in this paper was supported by grants to E.H.D. from the Office of Science (BER), U.S. Department of Energy, Grant No. DE-FG02-03ER63584 and from NIH, Grant No. HD-37105. P. Y. L. was supported by an NSF predoctoral fellowship.

REFERENCES

Bockamp, E.O., McLaughlin, F., Murrell, A., Green, A.R., 1994. Transcription factors and the regulation of haemopoiesis: lessons from GATA and SCL proteins. BioEssays 16, 481-8.

Davidson, E.H., Rast, J.P., Oliveri, P., Ransick, A., Calestani, C., Yuh, C.H., Minokawa, T., Amore, G., Hinman, V., Arenas-Mena, C., Otim, O., Brown, C.T., Livi, C.B., Lee, P.Y., Revilla, R., Rust, A.G., Pan, Z., Schilstra, M.J., Clarke, P.J., Arnone, M.I., Rowen, L., Cameron, R.A., McClay, D.R., Hood, L., Bolouri, H., 2002a. A genomic regulatory network for development. Science 295, 1669-78.

Davidson, E.H., Rast, J.P., Oliveri, P., Ransick, A., Calestani, C., Yuh, C.H., Minokawa, T., Amore, G., Hinman, V., Arenas-Mena, C., Otim, O., Brown, C.T., Livi, C.B., Lee, P.Y., Revilla, R., Schilstra, M.J., Clarke, P.J., Rust, A.G., Pan, Z., Arnone, M.I., Rowen, L., Cameron, R.A., McClay, D.R., Hood, L., Bolouri, H., 2002b. A provisional regulatory gene network for specification of endomesoderm in the sea urchin embryo. Dev. Biol. 246, 162-90.

Hinman, V.F. and Davidson, E.H., 2003. Expression of a gene encoding a Gata transcription factor during embryogenesis of the starfish Asterina miniata. Gene Expression Patterns 3, 419-22.

Hinman, V.F., Nguyen, A.T., Cameron, R.A., Davidson, E.H., 2003. Developmental gene regulatory network architecture across 500 million years of echinoderm evolution. Proc. Natl. Acad. Sci. USA 100, 13356-61.

Lowry, J.A. and Atchley, W.R., 2000. Molecular evolution of the GATA family of transcription factors: conservation within the DNA-binding domain. J. Mol. Evol. 50, 103-15.

Maduro, M.F. and Rothman, J.H., 2002. Making worm guts: the gene regulatory network of the Caenorhabditis elegans endoderm. Dev. Biol. 246, 68-85.

Minokawa, T., Rast, J.P., Arenas-Mena, C., Franco, C.B. and Davidson, E.H., 2004. Expression patterns of four different regulatory genes that function during sea urchin development. Gene Expression Patterns 4, 449-56.

Pancer, Z., Rast, J.P., Davidson, E.H., 1999. Origins of immunity: transcription factors and homologues of effector genes of the vertebrate immune system expressed in sea urchin coelomocytes. Immunogenetics 49, 773-86.

Poustka, A.J., Groth, D., Hennig, S., Thamm, S., Cameron, A., Beck, A., Reinhardt, R., Herwig, R., Panopoulou, G. and Lehrach, H., 2003. Generation, annotation, evolutionary analysis, and database integration of 20,000 unique sea urchin EST clusters. Genome Research 13, 2736-46.

Reiter, J.F., Alexander, J., Rodaway, A., Yelon, D., Patient, R., Holder, N., Stainier, D.Y., 1999. Gata5 is required for the development of the heart and endoderm in zebrafish. Genes. Dev. 13, 2983-95.

Reuter, R., 1994. The gene serpent has homeotic properties and specifies endoderm versus ectoderm within the Drosophila gut. Development 120, 1123-35.

Wang, D.G., Britten, R.J., Davidson, E.H., 1995. Maternal and embryonic provenance of a sea urchin embryo transcription factor, SpZ12-1. Mol. Mar. Biol. Biotechnol. 4, 148-53.

CHAPTER 2

Exclusive Developmental Functions of *gatae cis*-Regulatory Modules in the *Strongylocentrorus purpuratus* Embryo

Pei Yun Lee, Jongmin Nam and Eric H. Davidson

In press, *Developmental Biology*

Abstract

The *gatae* gene of *Strongylocentrotus purpuratus* is orthologous to vertebrate *gata-4,5,6* genes. This gene is expressed in the endomesoderm in the blastula and later the gut of the embryo, and is required for normal development. A *gatae* BAC containing a GFP reporter knocked into exon one of the gene was able to reproduce all aspects of endogenous *gatae* expression in the embryo. To identify putative *gatae cis*-regulatory modules we carried out an interspecific sequence conservation analysis with respect to a *Lytechinus variegatus gatae* BAC, which revealed 25 conserved non-coding sequence patches. These were individually tested in gene transfer experiments, and two modules capable of driving localized reporter expression in the embryo were identified. Module 10 produces early expression in mesoderm and endoderm cells up to the early gastrula stage, while module 24 generates late endodermal expression at gastrula and pluteus stages. Module 10 was then deleted from the *gatae* BAC by reciprocal recombination,

resulting in total loss of reporter expression in the time frame in which it is normally active. Similar deletion of module 24 led to ubiquitous GFP expression in the gastrula and pluteus. These results show that Module 10 is uniquely necessary and sufficient to account for the early phase of *gatae* expression during endomesoderm specification. In addition, they imply a functional *cis*-regulatory module exclusion, whereby only a single module can associate with the basal promoter and drive gene expression at any given time.

Keywords: sea urchin, gene regulation, GATA factors, *cis*-regulatory analysis, *gatae*

INTRODUCTION

GATA4,5,6 transcription factors and their orthologs are implicated in numerous aspects of endoderm and mesoderm development across the Bilateria (Maduro and Rothman, 2002; Murakami et al., 2005; Patient and McGhee, 2002). The sea urchin *Strongylocentrotus purpuratus* has two *gata* genes, of which *gatae* is orthologous to the vertebrate *gata4/5/6* genes (Pancer et al., 1999). The dynamic spatial expression of *gatae* in the sea urchin embryo was described by Lee and Davidson (2004). Expression is first detected in the 15 h blastula in cells of the presumptive mesoderm, and in the 24 h mesenchyme blastula the gene is expressed in endoderm and mesoderm cells of the veg2 lineage. At the onset of gastrulation the *gatae* gene is expressed in the invaginating vegetal plate and during gastrulation in the cells surrounding the blastopore as well as in mesoderm cells at the tip of the archenteron. In the later gastrula stages *gatae* is expressed in the midgut, hindgut and coelomic pouch regions of the archenteron. At the pluteus stage, hindgut expression is extinguished, leaving the definitive pattern of expression in the midgut and the coelomic pouches, which form the rudiment where the body plan of the adult sea urchin later develops.

The *gatae* gene occupies an important node in the sea urchin endomesoderm network. Perturbation analysis using morpholino antisense oligonucleotides (MASO), and many other observations, reveal that prior to gastrulation *gatae* is a direct activator of a number of genes encoding transcription factors, including the endodermal transcription factors *foxA*, *brachyury*, and *β1/2-otx* (Davidson et al., 2002a; Davidson et al., 2002b); see http://sugp.caltech.edu/endomes/ for current version of the endomesodermal gene

regulatory network). Of particular interest and importance is the interaction of Gatae factor with the *β1/2-otx* gene. These two genes cross-regulate, generating a positive feedback loop which serves to lock down the state of endoderm specification (Davidson et al., 2002a; Davidson et al., 2002b; Yuh et al., 2004).

Since the *gatae* gene is expressed in a complex spatial pattern which changes with developmental time, it seemed likely that more than one *cis*-regulatory module would be required to control its expression in the embryo. Here we show that a physically distinct "early module" is necessary and sufficient to account for expression up to the early gastrula stage, and that a separate "late module" takes over control of expression in the gut thereafter. Comparison of the expression patterns generated by deletion of either module from the genomic regulatory DNA with those generated by the individual modules in reporter constructs leads to the additional conclusion that *in situ* the function of one module excludes the function of the other.

MATERIALS AND METHODS

Identification of gatae *BACs and interspecific sequence comparison*

S. purpuratus and *L. variegatus* BAC libraries were screened with a mixture of two probes, one corresponding to exon 1 (5´ probe), and the other to exons 5 and 6 (3´ probe). Filters were hybridized in 5XSSPE, 5% SDS and 0.1% NaPPi at 65 °C and washed to a final concentration of 1XSSPE, 0.1% SDS. Positive clones were identified using the BioArray Software (Brown et al., 2002) and further confirmed by PCR and genomic DNA blots. Each clone was also mapped to determine the distance of the *gatae* gene from the vector. Mapping was done by digesting each BAC with Not I, which releases the insert, and either Bgl II, Xho I or Pst I. Genomic DNA blots were hybridized with combinations of probes corresponding to the vector (T7 and SP6), the 5´ and 3´ *gatae* exon probes. *Sp* and *Lv* BACs in which the *gatae* gene was furthest from the vector were sequenced at either the Joint Genome Institute or the Institute for Systems Biology (Seattle, Washington).

Interspecific sequence analysis was carried out using FamilyRelations (Brown et al., 2002). FamilyRelations software is available at http://family.caltech.edu.

Generation of cis*-regulatory reporter constructs*

Fusion PCR (Yon and Fried, 1989) was used in the generation of all reporter constructs. Each reporter construct consists of three separate PCR products stitched

together in a subsequent PCR reaction: the conserved sequence patch, the *gatae* basal promoter and the GFP coding region. PCR primers were designed for each conserved sequence patch identified by FamilyRelations analysis. The reverse primer also included the tail sequence 5´-GTGTTGAAGTAGCTGGCAGTGACGT-3´, which overlaps with the sequence of the *gatae* basal promoter. The *gatae* basal promoter was amplified using the forward primer 5´-ACGTCACTGCCAGCTACTTC-3´, and the reverse primer 5´-GTGAACAGTTCCTCGCCCTTGCTCATCTGATGTGGCATACCACGC-3´. The sequence underlined in this primer corresponds to the GFP coding region. The GFP reporter included the SV40 polyadenylation signal, and was amplified using as forward primer 5´-ATGAGCAAGGGCGAGGAACTG-3´, and as reverse primer 5´-TGACTGGGTTGAAGGCTCTC-3´. Each resulting PCR product was cloned into the pGEMTEZ vector (Promega Corporation, Madison, WI) and verified by sequencing. PCR reporter constructs were purified using a PCR Purification Kit (Qiagen Inc., Valencia, CA), and injected into fertilized eggs.

BAC homologous recombinations

BAC modifications involving homologous recombination utilized the method described by Lee et al. (2001). The targeting cassette consists of the GFP coding region and a kanamycin resistance gene flanked by frt sites. In this way the kanamycin resistance gene, used to select for recombinants, can be removed by arabinose induction after successful recombination. To generate the targeting cassette for creation of the *gatae* GFP BAC knockin, in which the GFP coding region was inserted into *gatae*'s first

exon, primers corresponding to exon 1 were designed as follows: forward primer, 5´CAGCAGTATCTTTATCCCCAGTATCATTTGACAAGCGAATCCCAAATGAGC AAGGGCGAGGAACT-3´; reverse primer, 5´ACTCCACACGGCTGCAGCAGCGTGAGCATTGGCCTGGATCACGCTTCGAAG AGCTATTCCAG-3´.

For deletion of *cis*-regulatory modules 10 and 24, primers were designed to flank the region marked for removal. The targeting cassettes used for module deletions did not include the GFP coding DNA, consisting only of the kanamycin resistance gene flanked by frt sites. Module 10del forward primer: 5´AAGTATTAATATATTGGAATTGTTACAATGTTAGATTTGTATTCATCATGTC TGGATCGAACACC-3´; module 10del reverse primer: 5´GCAAGATTATTAGTCACCGCTTGAAGAACATCGGGAAGAGAATGGGCTACC ATGGAGAAGTTCC-3´; module 24del forward primer: 5´AAAACTTGAATGATAACGACGCCTTGACTTACTGCCGTTTAAAGATCATGT CTGGATCGAACACC-3´; module 24del reverse primer: 5´TAAAGTTAGTCAAATAAGCTAATGTTTGGTGAGAAGGGTATGAGAGGCTAC CATGGAGAAGTTCC-3´.

Sequences corresponding to the targeting cassette are underlined. The targeting cassettes were electroporated into EL250 cells containing the Gatae BAC (GFP insertion) or Gatae GFP BAC (module deletion), and the λ recombination system activated by heat shock at 42 °C. After selection for recombinants and removal of the selectable marker, clones containing the targeted insertions or deletions were linearized with Not I and column purified before microinjection.

Quantitative PCR reporter analysis

Embryos injected with reporter constructs were collected at various stages of development and their RNA extracted using Qiagen's RNeasy Micro Kit (Qiagen Inc., Valencia, CA). RT-PCR was carried out using ABI's (Applied Biosystems, Foster City, CA) Taqman Reverse Transcription Reagents using random hexamer priming, while real-time QPCR reactions were performed in triplicate with ABI's SYBR Green PCR Master Mix. C_t is defined as the cycle number at which DNA in a PCR reaction reaches a particular threshold, set to a level where PCR products are increasing exponentially. C_ts for GFP were normalized to C_ts for SpZ12 as a control to account for differences in number of embryos in each preparation and converted to relative RNA levels using the formula $2^{\Delta Ct}$, where $\Delta C_t = C_{t(SpZ12)} - C_{t(GFP)}$.

Embryo culture, microinjection and whole mount* in situ *hybridizations

Fertilized eggs were injected with 10 pl of a solution containing 250 molecules of reporter construct/pl, following the microinjection and embryo culture procedures described by McMahon et al. (1985). Whole mount *in situ* hybridizations on injected embryos were performed as described (Minokawa et al., 2005).

RESULTS

Structure of the gatae *genomic locus*

Comparison of the *gatae* BAC sequence to that of *gatae* cDNA (Genbank Accession No. AY623814) revealed that the *gatae* gene contains 6 exons extending over 29 kb of genomic DNA (Fig. 2.1A). The two class IV zinc fingers are encoded in exons 3 and 4. Sea Urchin Genome Annotation Resource software (Brown et al., 2002) was used to predict the locations of the two genes flanking *gatae*. The nearest predicted coding region was 19 kb upstream of exon 1, matching a predicted sea urchin *beta-2 lactamase* gene (Genbank Accession No. XM_001177319). The nearest downstream

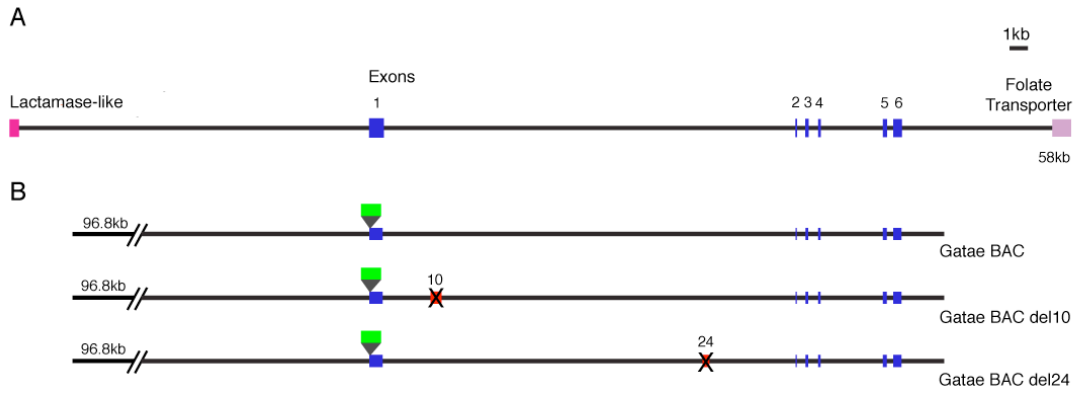


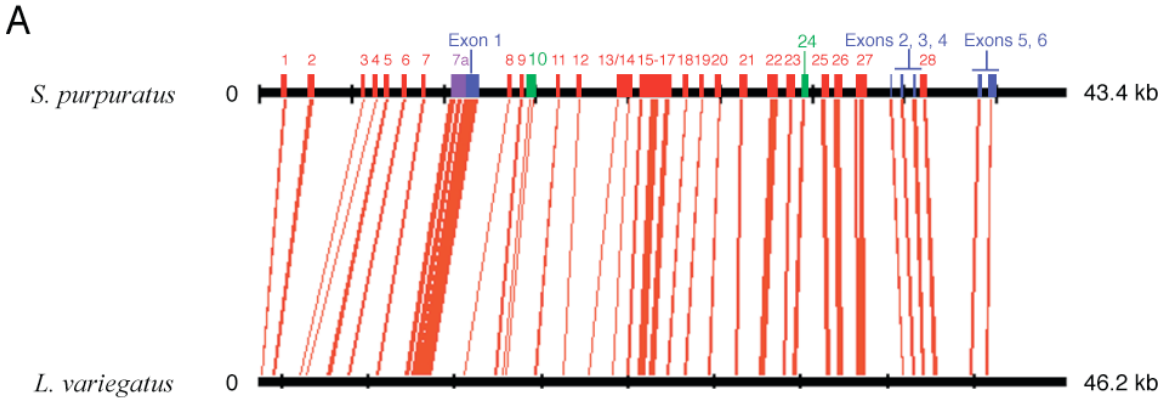
Figure 2.1. (A) Genomic locus of the *gatae* gene. The exons are represented by blue boxes and labeled by number. The *gatae* gene is flanked by a *beta2 lactamase-like* gene (pink box) upstream and a *folate transporter* gene (lavender box) downstream. (B) Maps of *gatae* GFP BACs, using the same scale as in (A). The green box represents the GFP coding region, inserted into the first exon of *gatae*, in an in-frame insertion replacing the ATG of *gatae* with that of GFP. The red boxes with crosses over them represent the positions of active *cis*-regulatory modules that were deleted from Gatae BAC using homologous recombination.

gene is a predicted *folate transporter* gene (Genbank Accession No. XM_001177178), located 8 kb 3′ of the *gatae* stop codon. These genes are both transcribed in the same direction as *gatae* (left to right in Fig. 2.1A). The assembled sequence of the *S. purpuratus* genome (Sodergren et al., 2006) confirmed that *gatae* is a single copy gene.

Conserved non-coding sequence patches in the vicinity of the gatae *gene*

Using FamilyRelations software (Brown et al., 2002), we compared the genomic sequence surrounding the *gatae* gene in *Strongylocentrotus purpuratus* and *Lytechinus variegatus gatae* BACs. The region scanned extended from the *lactamase* to the *folate transporter* gene (cf. Fig. 2.1A). Parameters were set to require 85% threshold identity within a 50 bp sliding window. This analysis (Fig. 2.2A) revealed the presence of 31 conserved sequence patches, five of which corresponded to *gatae* exons 2-6, and one patch which corresponded partly to exon 1. The conserved sequences range from 196 bp to 1.7 kb, with an average size of 440 bp.

In order to identify active *cis*-regulatory modules that drive *gatae* expression in the embryo, a series of reporter constructs were made (Fig. 2.2B). The individual



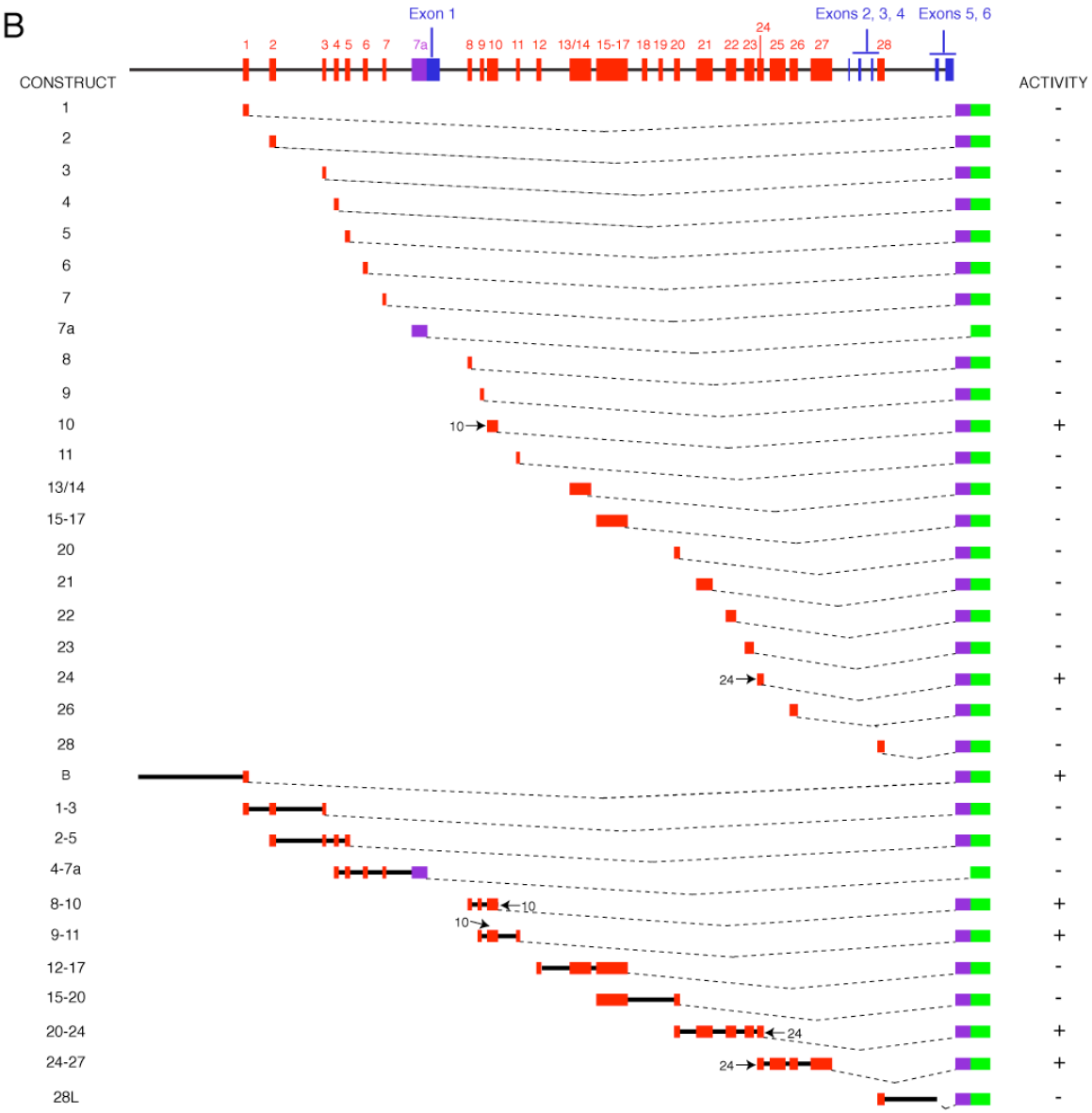


Figure 2.2. Conserved sequence patches in the vicinity of the *gatae* gene, and their *cis*-regulatory activities. (A), FamilyRelations analyses of *S. purpuratus* and *L. variegatus gatae* BACs. A 50 bp window was applied in the analysis, with an 85% identity threshold. Almost the same results were obtained using a 92% identity threshold, except for the disappearance of a few patches that were inactive in the experiments of part (B). Exons are represented by blue boxes, and the two active *cis*-regulatory modules are shown in green. (B), Diagrammatic representation of constructs created in *gatae cis*-regulatory analysis, and summary of activity. At the top is a map of the conserved sequence patches (red boxes) from the analysis in (A). The reporter constructs tested in this work are shown below. The name of each construct is listed in the left column, and an indication of its activity over background is at the right.

Here purple boxes denote the conserved sequence patch containing the *gatae* basal promoter, and green boxes represent the GFP coding region.

conserved sequence patches were amplified by PCR, and inserted into the expression vector, as described in Materials and Methods. Additional longer constructs were also prepared as indicated in the lower part of Fig. 2.2B (see Materials and Methods) to control for the possibility that functional sequence elements might be excluded from those inserts defined by conservation pattern, though this turned out not to be a concern. The conserved sequence immediately upstream of exon 1 (7a) appeared likely to include the *gatae* basal promoter, given its location, and indeed it includes the TATA box and initiator element sequences of the *gatae* gene. When cloned into a GFP reporter and introduced into eggs, fragment 7a generated no expression on its own, as characteristic of basal promoters in our expression vectors (Arnone et al., 1998; Sucov et al., 1988; Yuh and Davidson, 1996). Module 7a was included as the basal promoter in all of the *gatae cis*-regulatory constructs; experiments in which the *endo16* basal promoter was instead combined with active *gatae cis*-regulatory modules showed that the two basal promoters function in the same way (data not shown). Each reporter construct was injected into fertilized sea urchin eggs and observed at the mesenchyme blastula, gastrula and pluteus stages.

Two specifically active DNA fragments that generated specific endoderm and mesoderm expression in the embryo were identified in these preliminary experiments, viz. those included in conserved patches 10 and 24. The large distal fragment upstream of patch 1 was expressed ubiquitously, but was not studied further.

A GFP BAC that reproduces **gatae** *expression*

The sequence of the BAC containing the *gatae* gene begins 109 kb upstream of the first exon of *gatae*, and terminates 2.5 kb downstream of the last exon. Using *in vitro* recombination, we inserted the coding region of a GFP reporter into the first exon of *gatae* within the BAC (Fig. 2.1B, referred to as Gatae BAC). When injected into fertilized eggs, Gatae BAC was able to reproduce every aspect of endogenous *gatae* expression (Fig. 2.3 and Table 1). GFP fluorescence was detected in vegetal cells of the

Table 2.1. Expression of GFP in embryos injected with reporter constructs.

		24 h			
Construct	Number of embryos observed	Number of GFP+ embryos (%[a])	Endomesoderm (%[b])		Ectoderm (%[c])
10[b]	128	70 (55)	69(99)		0
24	163	80 (56)	30 (37)		65 (81.5)
Gatae BAC	303	198 (65)	198 (100)		0
Gatae BAC del10	320	5 (2)	5(100)		0
Gatae BAC del24	98	56 (57)	54 (97)		5 (10)
GataeBp	119	0 (0)	0		0
		48 h			
Construct	Number of embryos observed	Number of GFP+ embryos (%)	Endoderm (%[a])	Mesoderm (%[a])	Ectoderm (%[a])
10	272	76 (28)	42 (55)	14 (18)	34 (45)
24	166	98 (59)	93 (95)	9 (9)	2 (2)
Gatae BAC	313	161 (51)	146 (90)	53 (33)	1 (1)
Gatae BAC del10	269	125 (46)	115 (92)	47 (38)	0
Gatae BAC del24	179	111 (62)	72 (65)	32 (28)	53 (48)
GataeBp	100	2 (2)	0	2 (100)	0
		72 h			
Construct	Number of embryos observed	Number of GFP+ embryos (%)	Endoderm (%[a])	Mesoderm (%[a])	Ectoderm (%[a])
10	198	12 (6)	1 (8)	6 (50)	5 (42)
24	109	72 (66)	71 (99)	2 (3)	4 (7)
Gatae BAC	279	148 (53)	143 (97)	39 (26)	0
Gatae BAC del10	175	71 (41)	64 (90)	14 (20)	0
Gatae BAC del24	203	93 (46)	54 (58)	12 (13)	60 (65)
GataeBp	52	0 (0)	0	0	0

[a] Fertilized eggs were injected and analyzed by either *in situ* hybridization with a GFP probe or observation of GFP fluorescence. Because of the mosaic incorporation of DNA in sea urchin embryos, only a fraction of injected embryos will express the reporter gene in any given cell type.

[b] Percentages reflect embryos which expressed GFP in said cell type, including those that displayed GFP expression in two or more cell types.

[c] One PMC expressing embryo omitted for simplicity.

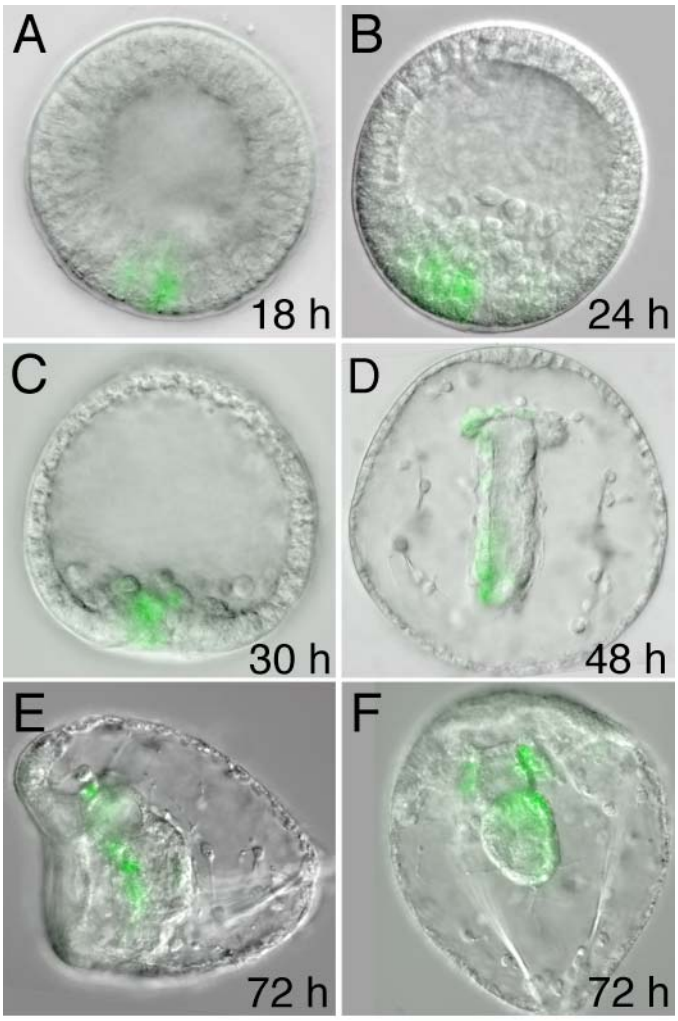


Figure 2.3. GFP fluorescence image overlays from embryos injected with the Gatae BAC. Each image is labeled with its developmental stage.

18 h blastula (Fig. 2.3A). In the mesenchyme blastula, GFP was observed in both endoderm and mesoderm cells of the veg2 lineage (Fig. 2.3B and Table 2.1). GFP reporter expression persisted in those cells until the onset of gastrulation (Fig. 2.3C). In the gastrula, GFP expression was restricted to endoderm cells of midgut and hindgut and mesoderm at the tip of the archenteron (Fig. 2.3D and Table 2.1). At 72 h, expression was limited to the midgut and coelomic pouches (Fig. 2.3E,F and Table 2.1). Thus, the

Gatae BAC must contain all the *cis*-regulatory information required to account for *gatae* expression in the embryo.

A cis*-regulatory module that reproduces early vegetal expression of* gatae

Region 10, a 585 bp conserved sequence located in the first intron (Fig. 2.2B), was capable of producing GFP reporter expression in the vegetal plate. In embryos injected with region 10 reporters, expression could be detected in a single localized region at 15 h (Fig. 2.4A). At 15 h it is not possible to determine the location of gene expression based on morphology alone, but by the time of vegetal plate thickening soon thereafter, it became obvious that expression driven by this DNA fragment is localized in

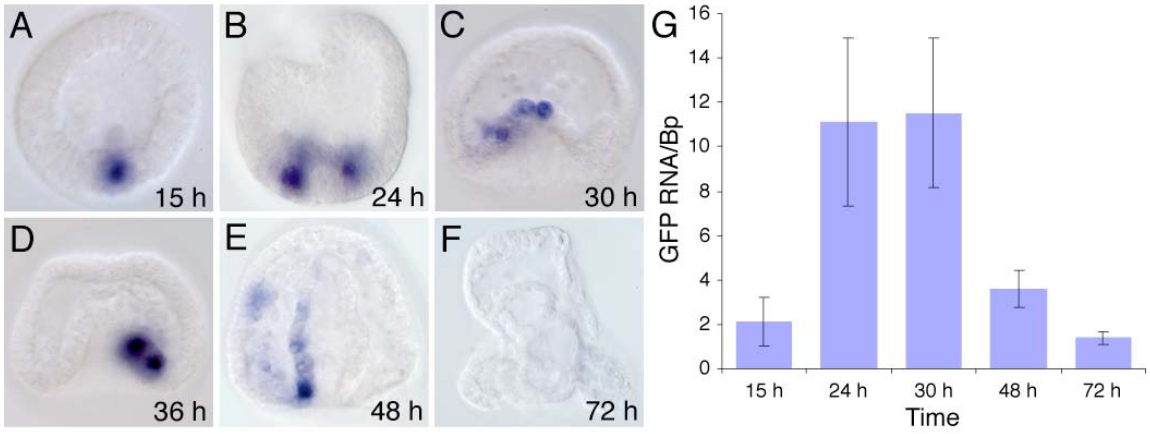


Figure 2.4. Expression of module 10 reporter construct. (A)-(F), Whole mount *in situ* hybridizations of embryos injected with module 10 GFP reporter constructs using a probe for GFP mRNA. Each image is labeled with the embryonic stage represented. (G), Activity of module 10 was normalized to that of embryos injected with a reporter construct containing only the *gatae* basal promoter (cf. Table1). Each time point is the average of seven trials; error bars represent two standard deviations from the mean.

the vegetal plate. In the mesenchyme blastula, this module generated GFP reporter expression in the endomesoderm specifically (Fig. 2.4B): 99% of GFP expressing embryos showed endomesoderm expression (Table 2.1). Expression persisted in the invaginating archenteron at the onset of gastrulation (Fig. 2.4C). However, in the 48 h gastrula, the module 10 construct produced ubiquitous expression (Fig. 2.4E and Table 2.1). This construct was completely inactive in the pluteus (Fig. 2.4F and Table 2.1). Consistent with these observations, constructs 10-12 and 9-11 produced the same patterns of expression as did the isolated module 10 (Fig. 2.2B and data not shown).

Expression of module 10 was studied in greater detail by quantifying the amount of GFP RNA generated by the construct over developmental time, using QPCR. As with the endogenous *gatae* gene (Lee and Davidson, 2004), reporter expression was first detected in the 15 h embryo. Expression then increased, peaking at 24 h and 30 h, before decreasing dramatically in the gastrula and pluteus (Fig. 2.4G). These data show that module 10 is a driver for *gatae* expression in the blastula. Since the turnover rate of GFP mRNA is not known in these cells, we cannot be sure when the transcriptional activity of module 10 constructs terminates, except that it is at or before the onset of gastrulation at 30h.

The late gatae *cis-regulatory module*

The second conserved patch in the first intron, the 334 bp region 24 (Fig. 2.2B), proved capable of driving endoderm-specific expression at gastrula and pluteus stages. However, both GFP fluorescence observation and *in situ* hybridizations revealed that

region 24 constructs are expressed ubiquitously up to 30 h (Fig. 2.5A-C and Table 2.1). By gastrula stage, expression has become highly specific and was confined to the midgut and hindgut (Fig. 2.5D and Table 2.1), while in the pluteus GFP reporter was only observed in the midgut (Fig. 2.5E,F and Table 2.1). It should be noted that module 24 was not expressed in the mesoderm cells at the tip of the archenteron in the gastrulating embryo or in the coelomic pouches at pluteus stage as is the endogenous gene and the Gatae BAC (Fig. 2.3). Regulatory functions required for coelomic pouch expression thus are missing from region 24, and from the extended constructs that include region 24, i.e., regions 15-20 or 20-24 (Fig. 2.2B). These extended fragments displayed the same endodermal activity in gastrula and pluteus stages as did the region 24 construct (data not shown).

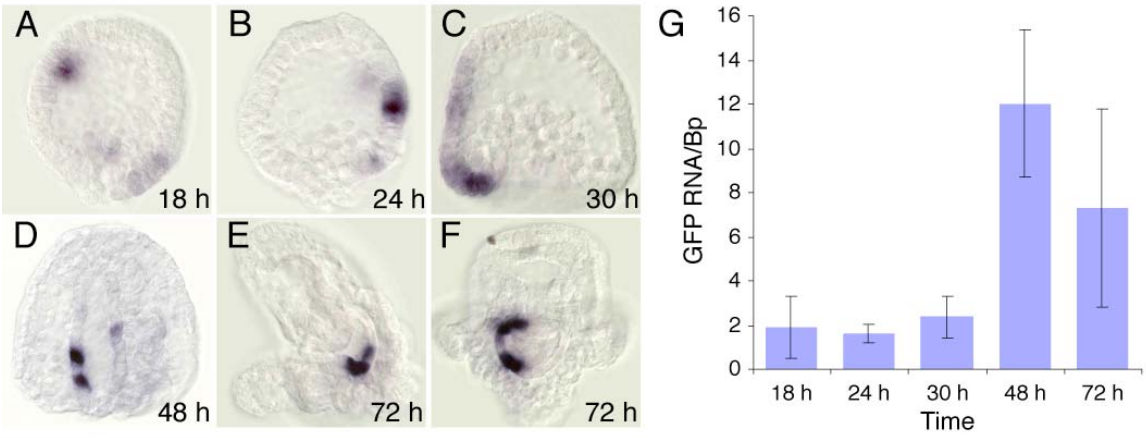


Figure 2.5. Expression of module 24 reporter construct. (A)-(F), Whole mount *in situ* hybridizations of embryos injected with module 24 GFP reporter constructs, using a probe for GFP mRNA. Each image is labeled with the embryonic stage represented. (G), QPCR of the module 24 reporter construct, at the indicated times. Activity of module 24 was normalized to that of embryos injected with the reporter construct containing only the *gatae* basal promoter. Each time point is the average from four trials; error bars represent two standard deviations from the mean.

QPCR time courses performed on embryos injected with the module 24 reporter construct revealed that reporter levels were relatively low up to 30 h, and the main activity was at the 48 h gastrula and the 72 h pluteus stages (Fig. 2.5G). Therefore the main function of module 24 is to drive *gatae* expression in the gastrula and pluteus. Considering the expression data for modules 10 and 24 together, it is clear that their expression patterns are complementary, both spatially and temporally. Together they account for the totality of embryonic *gatae* expression, except for the late expression in the mesodermal coelomic pouches. The control locus for this aspect of *gatae* expression remains undiscovered.

Necessity of module 10 for **gatae** *expression in the blastula*

To determine if module 10 is required for the early expression of *gatae*, it was deleted from Gatae BAC (Gatae BAC del10) by homologous recombination (see Materials and Methods). This enabled the study of the function of the module in the context of the complete *gatae* genomic locus and to identify any intermodular interactions. The result was clear: when Gatae BAC del10 was injected into embryos, no expression whatsoever was seen in 15 h, 24 h, or 30 h embryos (Fig. 2.6A-C and Table 2.1), but strong GFP expression was observed in the gastrula stage, in the midgut, hindgut and mesoderm (Fig. 2.6D and Table 2.1). In pluteus stage embryos bearing Gatae BAC del10, GFP was expressed in the midgut and the coelomic pouches (Fig. 2.6E, F and Table 2.1).

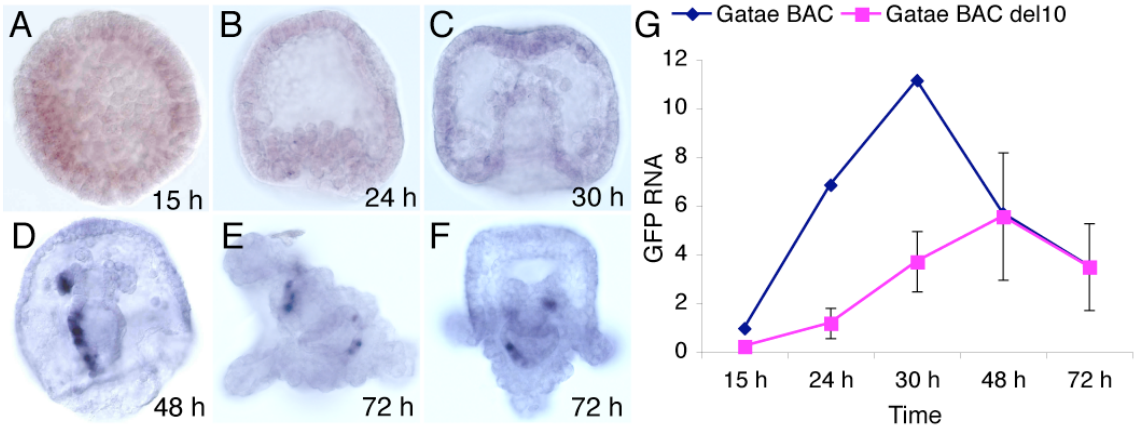


Figure 2.6. Expression of Gatae BAC del10. (A)-(F), Whole mount *in situ* hybridizations of embryos injected with Gatae BAC del10, using a probe for GFP mRNA. Each image is labeled with the embryonic stage represented. Curves were compiled from the average of seven trials, with error bars representing two standard deviations, normalized to Gatae BAC values at each point.

QPCR time courses were generated from embryos injected with Gatae BAC and Gatae BAC del10 (Fig. 2.6G), and the data were consistent with the spatial expression. GFP RNA levels in Gatae BAC del10 embryos remain low compared to the control until the gastrula stage, and by 48 h they revert to the levels produced by the wildtype Gatae BAC. The results demonstrate that module 10 is the only module utilized during blastula stages, and is necessary as well as sufficient for *gatae* expression in the vegetal pole.

Deletion of the late module

A construct lacking module 24 was similarly generated (Gatae BAC del24). Embryos injected with Gatae BAC del24 express GFP vegetally at 15 h and 24 h in the same spatial domain as the control Gatae BAC (Fig. 2.7A, B and Table 2.1). Furthermore the amount of early expression is exactly the same as recorded for the

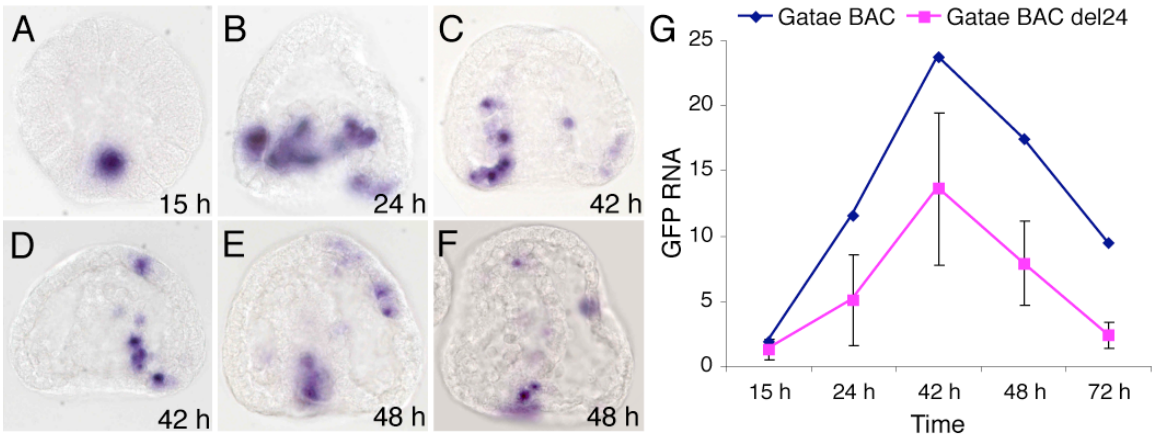


Figure 2.7. Expression of Gatae BAC del24. (A)-(F), Whole mount *in situ* hybridizations of embryos injected with Gatae BAC del24 using the GFP probe. Each image is labeled with the embryonic stage represented. (G), (G), QPCR measurements of GFP mRNA generated by Gatae BAC del24 and Gatae BAC. Curves were compiled from the average of five trials, with error bars representing two standard deviations, normalized to the Gatae BAC values at each point

isolated module 10 construct (55% vs. 57%). Surprisingly, however, we observed ubiquitous GFP expression in Gatae BAC del24 embryos after this (Fig. 2.7C-F). In the gastrula 52% of GFP expressing embryos showed expression in endoderm or mesoderm cells, but 48% displayed some level of expression in the ectoderm. In sharp contrast, in the parental Gatae BAC, 100% of GFP positive embryos expressed only in the endoderm. A similar observation was made in the pluteus, in which GFP was observed in endoderm and mesoderm in 35% of embryos and 65% displayed some level of ectodermal expression, while 100% of embryos bearing the Gatae BAC control expressed GFP in endoderm and mesoderm (Table 2.1).

QPCR analysis of levels GFP reporter RNA produced by Gatae BAC del24 support the spatial expression data. At no time was GFP RNA eliminated or drastically reduced. Instead, we observed reduced levels of GFP RNA in embryos injected with

Gatae BAC del24 compared to Gatae BAC. Even though we did not observe a loss of expression in the gastrula and pluteus stages, spatial expression at those time points had been completely disrupted by the removal of module 24. Therefore, as is module 10 at early stages, module 24 is necessary for the correct spatial regulation of *gatae* at late stages.

DISCUSSION

Cis-*regulation of* gatae

Here we show that two physically distinct *cis*-regulatory modules control different aspects of *gatae* expression in the sea urchin embryo. Module 10 is active early, from the onset of expression in the presumptive secondary mesenchyme cells to the early gastrula phase of expression in the vegetal plate endoderm and mesoderm. Sometime during early gastrulation module 24 takes over control from module 10, directing *gatae* expression in the gut endoderm of the gastrula and pluteus. This modular organization reflects the requirement for regulation by diverse sets of transcription factors at the respective stages, i.e., during specification of the endomesoderm, and during definitive regionalization and differentiation of the gut. The *gatae* gene itself plays different roles in these phases of its activity. The endomesodermal gene regulatory network shows explicitly how *gatae* functions to activate a number of other regulatory genes during the specification phase (Davidson, 2006; Davidson et al., 2002a; Davidson et al., 2002b). Given its regionalized pattern of expression in the gut of the late embryo, *gatae* may be involved in specification of first the hindgut and then the midgut, and in activation of gut differentiation gene batteries.

While the endogenous *gatae* gene and Gatae BAC express strongly in the mesoderm cells of the gastrula and the coelomic pouches of the pluteus embryo, neither module 10 nor module 24 directs expression to these cells. An additional control module is thus implied. This is likely to reside >10 kb upstream of conserved patch 1, the limit

of overlap of the *L. variegatus* BAC with the *S. purpuratus* sequence. This leaves roughly 9 kb to the *beta2 lactamase-like* gene which will be possible to explore by FamilyRelations only when the respective *L. variegatus* sequence becomes available. It is unlikely that the missing module is downstream of the region we have examined, since Gatae BAC expresses in coelomic pouches though it terminates only 2.5 kb beyond exon 6.

In the context of the endomesoderm gene regulatory network, an important result is that module 10 alone is necessary and sufficient to drive *gatae* expression throughout the phase of development to which the network analysis pertains. Therefore all interactions from upstream regulators into *gatae* will have to be mediated by and processed through this module. We have identified binding sites for such inputs as predicted by the endomesoderm gene network in module 10, and are in the process of mutating and analyzing these sites in detail, to be reported in a subsequent publication (Lee and Davidson, 2007).

Homologous BAC recombination as a tool for* cis*-regulatory analysis

Conventional *cis*-regulatory analysis on isolated modules, including site specific mutagenesis, provides our most powerful and direct tool for demonstrating functionally the roles of given *cis*-regulatory inputs. By this means proposed upstream linkages of a regulatory module into the gene regulatory network can be certified or rejected. The use of homologous BAC recombination further enhances the arsenal of functional *cis*-regulatory approaches, opening up several additional possibilities: (1) As have others

(Hadchouel et al., 2003; Teboul et al., 2002) we show here how deletion of a specific regulatory module can be used to establish its necessity as well as its sufficiency. This excludes the possibility of regulatory redundancy. (2) BAC reporter knockins which provide the complete and accurate spectrum of expression of a given gene are a useful starting place to narrow the genomic domain over which specific *cis*-regulatory modules are to be sought. (3) BAC reporter knockins provide built in components for single module expression constructs that include the endogenous basal promoter. (4) BAC reporter knockins enable the study of intermodular interactions in the natural context of the gene, and this has proved one of the most interesting aspects of the present work.

Exclusionary function of cis*-regulatory modules*

The expression of the module 10 and 24 reporter constructs differ in a revealing way from the expression of Gatae BAC. When individually cloned in front of the reporter, each module was capable of driving spatially specific expression for part of embryogenesis, but each produced ubiquitous, albeit weak expression at other stages. Yet in their natural context they work sequentially to produce highly specific patterns of expression with no ectopic expression of any kind, as seen from endogenous *gatae* expression and that of Gatae BAC. This difference devolves from the global structure of the locus: we see the whole has additional functions than do the sum of individual constructs. Individual constructs display outputs from *cis*-regulatory processing of their individual inputs while the overall regulatory function of the *gatae* locus includes mechanisms that determine which *cis*-regulatory modules are allowed to function; thus

far there has been little information regarding the experimental verification of such alternate use of *cis*-regulatory modules.

In Fig. 2.8 we present a model for how this might occur. The premise is that module function requires physical association with the basal transcription apparatus (BTA), and that a given association precludes all other modules from such association. This would be the consequence of association by looping, undoubtedly the general

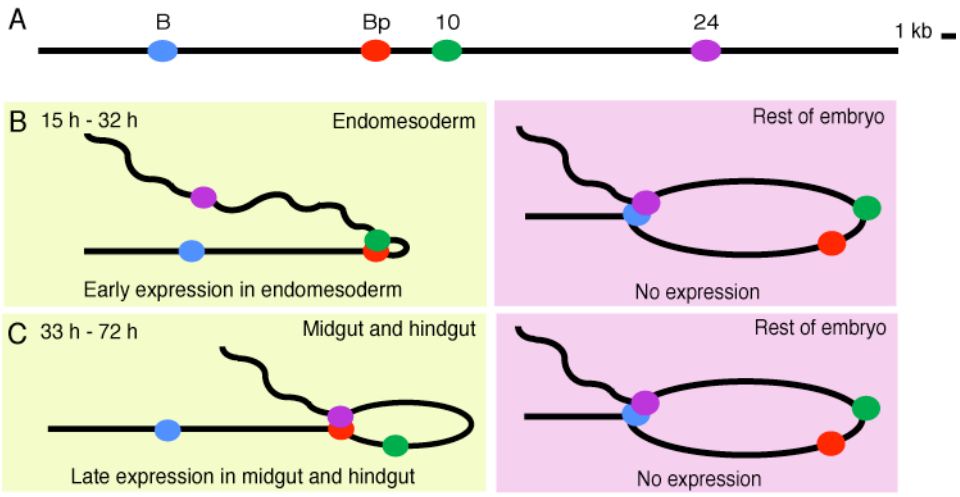


Figure 2.8. Cartoon representation of module exclusion by looping. (A), Map of the locus of the *gatae* gene with three active *cis*-regulatory modules; B, an upstream region which drives expression ubiquitously; the intron module 10, which drives expression in the endomesoderm during blastula; the intron module 24, which drives expression in endoderm at gastrula and pluteus stages. The basal promoter is denoted Bp. (B), In the endomesoderm of the blastula (green box), module 10 associates with Bp to drive expression when it is occupied by its endomesoderm transcription factors; in the rest of the embryo (pink box), modules B and 24 are associated by looping and the *gatae* gene is not transcribed. (C), In the midgut and hindgut of the gastrula (green box), module 24 associates with the Bp, when it is occupied with the gut factors for which it contains sites, to drive specific expression. The rest of the embryo (pink box) follows a similar scheme as in (B).

mechanism by which distant *cis*-regulatory modules are brought to the immediate vicinity of the BTA (reviewed by Davidson, 2006). With respect to choice of active *cis*-regulatory module, a looping mechanism confers a Boolean quality to the regulatory system (Istrail and Davidson, 2005). In our present case, the *gatae* gene contains two *cis*-regulatory modules active in the embryonic endomesoderm, viz., module 10 for early expression and module 24 for late expression. In the normal context, module 10 associates with the BTA up to the early gastrula, driving endoderm and mesoderm expression (Fig. 2.8B). This excludes module 24 from association with the BTA within the endomesoderm during this period, when module 10 is loaded with its transcription factors (Lee and Davidson, 2007). Outside the endomesoderm, module 24 is at this same early period capable of generating weak (ectopic) expression (Fig. 2.4) if it is cloned in juxtaposition to the BTA, but it does not do so in context. Therefore when this module is not loaded with its cognate transcription factors it cannot loop to the BTA. Sometime in early- to midgastrula, however, these factors become available in endoderm cells, and there module 24 is activated, loops to the BTA, and generates specific expression in the midgut and hindgut (Fig. 2.8C). At this time module 10 is essentially relieved of its duty and is excluded from association with the basal promoter. As for module 24 at early times, in cells outside of the endoderm at late stages module 10 cannot now cause expression unless it is artificially brought into the immediate context of the BTA. Thus, though each of these *cis*-regulatory modules in isolation displays weak ubiquitous expression at certain times, in context they function alternately to produce highly specific expression.

However there is an asymmetry in this system, as shown by the results of the BAC deletions. Deletion of module 10 results in complete loss of early expression, followed by normal late expression; as above the potential of module 24 for early expression outside the endomesoderm cannot be realized unless it is artificially positioned next to the BTA. However, deletion of module 24 results in ubiquitous late Gatae BAC expression. This could be driven by the action of the distal "B" element in the undefined region upstream of conserved patch 1. A prediction within the framework of the model in Fig. 8 is that the asymmetry in the consequences of these two deletions is due to a second kind of looping: in all embryo cells at all times module 24 is looped to the B region, preventing it from functioning with the BTA, except in late endoderm cells when it becomes loaded with endoderm transcription factors and occupies the BTA itself. Deletion of module 24 would release this constraint, resulting in B-driven ectopic expression.

An alternative, that ectopic expression is precluded by specific repressors target sites for which are located within modules 10, 24, and B, seems too baroque to consider seriously. This would require that the repressor that acts on module 10 is present in all cells except endomesoderm at early times and in all cells at late times, while that which acts on module 24 is present everywhere early and then in all cells except gut at late times, etc. Furthermore the ectopic expression seen in module 24 deletions from the Gatae BAC cannot easily be explained in this way.

In summary, we describe two levels of *cis*-regulatory control in the *gatae* gene. The first is the classic, module-specific *cis*-regulatory design that determines time and place of regulatory function for each module. This is clearly revealed in experiments

with single module expression constructs. The second is the level of exclusionary *cis*-regulatory module interactions on the scale of the gene as a whole. This can only be perceived in experiments carried out on that scale, for which recombinant BAC constructs provide a ready approach.

ACKNOWLEDGEMENTS

We would like to express our gratitude to Pat Leahy at Kerckhoff Marine Lab for his excellent care of the sea urchins. We would also like to thank Julie Hahn for generating the Gatae GFP BAC and for technical advice with BAC deletions. This work was supported by NIH grant HD-37105.

REFERENCES

Arnone, M.I., Martin, E.L. and Davidson, E.H., 1998. Cis-regulation downstream of cell type specification: a single compact element controls the complex expression of the CyIIa gene in sea urchin embryos. Development 125, 1381-95.

Brown, C.T., Rust, A.G., Clarke, P.J., Pan, Z., Schilstra, M.J., De Buysscher, T., Griffin, G., Wold, B.J., Cameron, R.A., Davidson, E.H. and Bolouri, H., 2002. New computational approaches for analysis of cis-regulatory networks. Dev Biol 246, 86-102.

Davidson, E.H., 2006. The Regulatory Genome. Academic Press, San Diego.

Davidson, E.H. and Erwin, D.H., 2006. Gene regulatory networks and the evolution of animal body plans. Science 311, 796-800.

Davidson, E.H., Rast, J.P., Oliveri, P., Ransick, A., Calestani, C., Yuh, C.H., Minokawa, T., Amore, G., Hinman, V., Arenas-Mena, C., Otim, O., Brown, C.T., Livi, C.B., Lee, P.Y., Revilla, R., Rust, A.G., Pan, Z., Schilstra, M.J., Clarke, P.J., Arnone, M.I., Rowen, L., Cameron, R.A., McClay, D.R., Hood, L. and Bolouri, H., 2002a. A genomic regulatory network for development. Science 295, 1669-78.

Davidson, E.H., Rast, J.P., Oliveri, P., Ransick, A., Calestani, C., Yuh, C.H., Minokawa, T., Amore, G., Hinman, V., Arenas-Mena, C., Otim, O., Brown, C.T., Livi, C.B., Lee, P.Y., Revilla, R., Schilstra, M.J., Clarke, P.J., Rust, A.G., Pan, Z., Arnone, M.I., Rowen, L., Cameron, R.A., McClay, D.R., Hood, L. and Bolouri, H., 2002b. A provisional regulatory gene network for specification of endomesoderm in the sea urchin embryo. Dev Biol 246, 162-90.

Hadchouel, J., Carvajal, J.J., Daubas, P., Bajard, L., Chang, T., Rocancourt, D., Cox, D., Summerbell, D., Tajbakhsh, S., Rigby, P.W. and Buckingham, M., 2003. Analysis of a key regulatory region upstream of the Myf5 gene reveals multiple phases of myogenesis, orchestrated at each site by a combination of elements dispersed throughout the locus. Development 130, 3415-26.

Istrail, S. and Davidson, E.H., 2005. Logic functions of the genomic cis-regulatory code. Proc Natl Acad Sci U S A 102, 4954-9.

Lee, E.C., Yu, D., Martinez de Velasco, J., Tessarollo, L., Swing, D.A., Court, D.L., Jenkins, N.A. and Copeland, N.G., 2001. A highly efficient Escherichia coli-based chromosome engineering system adapted for recombinogenic targeting and subcloning of BAC DNA. Genomics 73, 56-65.

Lee, P.Y. and Davidson, E.H., 2004. Expression of Spgatae, the Strongylocentrotus purpuratus ortholog of vertebrate GATA4/5/6 factors. Gene Expr Patterns 5, 161-5.

Lee, P.Y. and Davidson, E.H., 2007. Use of OR logic in the regulation of a sea urchin GATA factor (in preparation).

Maduro, M.F. and Rothman, J.H., 2002. Making worm guts: the gene regulatory network of the Caenorhabditis elegans endoderm. Dev Biol 246, 68-85.

McMahon, A.P., Flytzanis, C.N., Hough-Evans, B.R., Katula, K.S., Britten, R.J. and Davidson, E.H., 1985. Introduction of cloned DNA into sea urchin egg cytoplasm: replication and persistence during embryogenesis. Dev Biol 108, 420-30.

Minokawa, T., Wikramanayake, A.H. and Davidson, E.H., 2005. cis-Regulatory inputs of the wnt8 gene in the sea urchin endomesoderm network. Dev Biol 288, 545-58.

Murakami, R., Okumura, T. and Uchiyama, H., 2005. GATA factors as key regulatory molecules in the development of Drosophila endoderm. Dev Growth Differ 47, 581-9.

Pancer, Z., Rast, J.P. and Davidson, E.H., 1999. Origins of immunity: transcription factors and homologues of effector genes of the vertebrate immune system expressed in sea urchin coelomocytes. Immunogenetics 49, 773-86.

Patient, R.K. and McGhee, J.D., 2002. The GATA family (vertebrates and invertebrates). Curr Opin Genet Dev 12, 416-22.

Sodergren, E., Weinstock, G.M., Davidson, E.H., Cameron, R.A., Gibbs, R.A., Angerer, R.C., Angerer, L.M., Arnone, M.I., Burgess, D.R., Burke, R.D., Coffman, J.A., Dean, M., Elphick, M.R., Ettensohn, C.A., Foltz, K.R., Hamdoun, A., Hynes, R.O., Klein, W.H., Marzluff, W., McClay, D.R., Morris, R.L., Mushegian, A., Rast, J.P., Smith, L.C., Thorndyke, M.C., Vacquier, V.D., Wessel, G.M., Wray, G., Zhang, L., Elsik, C.G., Ermolaeva, O., Hlavina, W., Hofmann, G., Kitts, P., Landrum, M.J., Mackey, A.J., Maglott, D., Panopoulou, G., Poustka, A.J., Pruitt, K., Sapojnikov, V., Song, X., Souvorov, A., Solovyev, V., Wei, Z., Whittaker, C.A., Worley, K., Durbin, K.J., Shen, Y., Fedrigo, O., Garfield, D., Haygood, R., Primus, A., Satija, R., Severson, T., Gonzalez-Garay, M.L., Jackson, A.R., Milosavljevic, A., Tong, M., Killian, C.E., Livingston, B.T., Wilt, F.H., Adams, N., Belle, R., Carbonneau, S., Cheung, R., Cormier, P., Cosson, B., Croce, J., Fernandez-Guerra, A., Geneviere, A.M., Goel, M., Kelkar, H., Morales, J., Mulner-Lorillon, O., Robertson, A.J., Goldstone, J.V., Cole, B., Epel, D., Gold, B., Hahn, M.E., Howard-Ashby, M., Scally, M., Stegeman, J.J., Allgood, E.L., Cool, J., Judkins, K.M., McCafferty, S.S., Musante, A.M., Obar, R.A., Rawson, A.P., Rossetti, B.J., Gibbons, I.R., Hoffman, M.P., Leone, A., Istrail, S., Materna, S.C., Samanta, M.P., Stolc, V., Tongprasit, W., Tu, Q., Bergeron, K.F., Brandhorst, B.P., Whittle, J., Berney, K., Bottjer, D.J., Calestani, C., Peterson, K., Chow, E., Yuan, Q.A., Elhaik, E., Graur, D., Reese, J.T., Bosdet, I., Heesun, S., Marra, M.A., Schein, J., Anderson, M.K., Brockton, V., Buckley, K.M., Cohen, A.H., Fugmann, S.D., Hibino, T., Loza-Coll, M., Majeske, A.J., Messier, C., Nair, S.V., Pancer, Z., Terwilliger, D.P., Agca, C., Arboleda, E., Chen, N., Churcher, A.M., Hallbook, F., Humphrey, G.W., Idris, M.M., Kiyama, T., Liang, S., Mellott, D., Mu, X., Murray, G., Olinski, R.P., Raible, F., Rowe, M., Taylor, J.S., Tessmar-Raible, K., Wang, D., Wilson, K.H., Yaguchi, S., Gaasterland, T., Galindo, B.E., Gunaratne, H.J., Juliano, C., Kinukawa, M., Moy, G.W., Neill, A.T., Nomura, M., Raisch, M., Reade, A., Roux, M.M., Song, J.L., Su, Y.H., Townley, I.K., Voronina, E., Wong, J.L., Amore, G., Branno, M., Brown, E.R., Cavalieri, V., Duboc, V., Duloquin, L., Flytzanis, C., Gache, C., Lapraz, F., Lepage, T., Locascio, A., Martinez, P., Matassi, G., Matranga, V., Range, R., Rizzo, F., Rottinger, E., Beane, W., Bradham, C., Byrum, C., Glenn, T., Hussain, S., Manning, F.G., Miranda, E., Thomason, R., Walton, K., Wikramanayke, A., Wu, S.Y., Xu, R., Brown, C.T., Chen, L., Gray, R.F., Lee, P.Y., Nam, J., Oliveri, P., Smith, J., Muzny, D., Bell, S., Chacko, J., Cree, A., Curry, S., Davis, C., Dinh, H., Dugan-Rocha, S., Fowler, J., Gill, R., Hamilton, C., Hernandez, J., Hines, S.,

Hume, J., Jackson, L., Jolivet, A., Kovar, C., Lee, S., Lewis, L., Miner, G., Morgan, M., Nazareth, L.V., Okwuonu, G., Parker, D., Pu, L.L., Thorn, R. and Wright, R., 2006. The genome of the sea urchin Strongylocentrotus purpuratus. Science 314, 941-52.

Sucov, H.M., Hough-Evans, B.R., Franks, R.R., Britten, R.J. and Davidson, E.H., 1988. A regulatory domain that directs lineage-specific expression of a skeletal matrix protein gene in the sea urchin embryo. Genes Dev 2, 1238-50.

Teboul, L., Hadchouel, J., Daubas, P., Summerbell, D., Buckingham, M. and Rigby, P.W., 2002. The early epaxial enhancer is essential for the initial expression of the skeletal muscle determination gene Myf5 but not for subsequent, multiple phases of somitic myogenesis. Development 129, 4571-80.

Yon, J. and Fried, M., 1989. Precise gene fusion by PCR. Nucleic Acids Res 17, 4895.

Yuh, C.H. and Davidson, E.H., 1996. Modular cis-regulatory organization of Endo16, a gut-specific gene of the sea urchin embryo. Development 122, 1069-82.

Yuh, C.H., Dorman, E.R., Howard, M.L. and Davidson, E.H., 2004. An otx cis-regulatory module: a key node in the sea urchin endomesoderm gene regulatory network. Dev Biol 269, 536-51.

CHAPTER 3

Use of OR Logic in the Regulation of a Sea Urchin GATA Factor

Pei Yun Lee and Eric H. Davidson

In preparation, *Developmental Biology*

ABSTRACT

The sea urchin *gatae* gene occupies an important node in the endomesoderm gene network, playing important roles in transcription factor regulation and the lockdown of endomesoderm specification. Module 10 is a previously identified module that is necessary and sufficient for *gatae* expression in the blastula stages. Co-injection of module 10 and mRNA for an Otx-engrailed fusion greatly reduced GFP reporter expression. When the injection is repeated with module 10 with mutated Otx binding sites, the number of GFP expressing embryos was restored to WT levels. Mutations of Otx binding sites led to a decrease in GFP RNA levels, while spatial expression remained unaffected. Module 10 expression was also reduced in embryos injected with RNA for a dominant negative form of the Suppressor-of-Hairless (Su(H)) protein; mutation of Su(H) binding sites led to a decrease in reporter expression with no effect on spatial expression. Mutation of both Otx and Su(H) binding sites in module 10 led to a further reduction, but not elimination of GFP expression. Co-injection of module 10 containing Otx site

mutations and dNSu(H) RNA abolished all reporter expression, demonstrating that any remaining activity of module 10 was also mediated by notch signaling, and that the spatial regulation of *gatae* is mediated by OR logic.

Keywords: *gatae*, gene regulation, GATA factors, sea urchin

INTRODUCTION

The sea urchin endomesoderm gene regulatory network (GRN) provides a system based view into the process of development. The GRN was constructed by identifying genes expressed in the endomesoderm, then determining their relationships by perturbating gene expression using dominant negative, dominant repressive forms of a transcription factor or antisense morpholino oligonucleotides (MASO), followed by the assaying endomesodermal gene expression using quantitative PCR (QPCR) (Calestani et al., 2003; Davidson et al., 2002a; Davidson et al., 2002b; Rast et al., 2002). As the endomesoderm GRN is essentially a model, one of the goals is to determine its validity by testing individual connections, which can be done through *cis*-regulatory analysis by mutating the binding sites for predicted inputs (Yuh et al., 2004).

The *Strongylocentrotus purpuratus gatae* is orthologous to the vertebrate *gata4/5/6* genes (Pancer et al., 1999). *gatae* is expressed in the endoderm and mesoderm throughout embryogenesis, beginning in the mesoderm precursors in the 15 h blastula and expanding to the endoderm and mesoderm by the mesenchyme blastula stage. During gastrulation it is expressed in the gut and mesoderm cells at the tip of the archenteron (Lee and Davidson, 2004). MASO perturbation analysis has determined that *gatae* functions as a regulator of transcription factors in the endomesoderm and to stabilize the endomesoderm developmental program through its cross-regulation with *otx*. *gatae* has been determined to be downstream of *otx* and the Notch signaling pathway (Davidson et al., 2002a; Davidson et al., 2002b).

In an exhaustive analysis on the *cis*-regulatory region of *gatae*, we determined that *gatae* expression in the embryo is controlled by two modules in the first intron. The early module (referred to as module 10) is responsible for driving expression up to the early gastrula, while the late module controls endodermal expression in the gastrula and pluteus. Deletion of the module 10 from a *gatae* GFP reporter BAC led to a total loss of blastula expression while late expression was undisturbed, demonstrating that it is both necessary and sufficient for the expression of *gatae* during the time frame of the endomesoderm GRN (Lee et al., 2007).

We have undertaken a detailed study of the early module with the goal of determining the GRN's authenticity as it pertains to *gatae* regulation. We demonstrated that module 10 responds to the perturbation of both *otx* expression and Notch signaling, and through binding site mutations showed that both inputs act directly on module 10. In addition, we have determined that there is an unknown component of the Notch input not mediated through Su(H). Importantly, we have verified the existence of an important feature of the endomesoderm network, the *otx-gatae* cross-regulatory loop, at the DNA level. Furthermore, it is also determined that the spatial regulation of *gatae* by Otx and Notch signaling occurs via OR logic, whereby the presence of either input is able to recapitulate the full complement of *gatae* expression.

MATERIALS AND METHODS

Site-directed mutagenesis

Binding site mutations were generated with either the site-directed mutagenesis kit (Stratagene Corp., La Jolla, CA) or via fusion PCR (Yon and Fried, 1989). Mutations generated by fusion PCR were cloned in pGEMTEZ (Promega Corp., Madison, WI); all constructs were verified by sequencing. Fusion PCR was also used to generate the linear DNA molecules for microinjection, combining module 10 containing desired binding site mutations with the *gatae* basal promoter and GFP coding region. The lone Otx binding site (Otx3) was mutated by changing the TAATCY consensus to TGGTCY. The two Otx (Otx1/2) binding sites in proximity with each other were deleted from the module. Primer sequences are as follows: Otxmut3F:
5'-GGGTAGCTTGGGACCACAACCTTTTTGATTAGCGCC-3'; Otxmut3R:
5'-GCGCTAATCAAAAAGGTTGTGGTCCCAAGCTACCC-3'; Otxdel1/2F:
5'-GTACAGTTACAAGATGGGTGAACCTGGAC-3'; Otxdel1/2R:
5'-GTCCAGGTTCACCCATCTTGTAACTGTAC-3'.
Underlined sequences correspond to the mutated Otx binding site.

Su(H) binding sites were identified by how well they match the consensus binding site YRTGRGAD. Primer sequences used in the mutations of Su(H) binding sites are as follows:
SuHmut1/2F:

5’- CATTACTTTGATAAATTAGGGGGCACGCACTAAATCAATATTC-3’;

SuHmut3F:

5’- CATCCTTACATACACTCATAGTGCACCCTCCTTTTTTCTCTTTTG-3’;

SuHmut3R:

5’- CAAAAGAGAAAAAAGGAGGGTGCACTATGAGTGTATGTAAGGATG-3’;

SuHmut4F:

5’- GATTTTGAAGTTGGTTTGTGGAGAGCACCGTGTATCGTTGTTC-3’;

SuHmut4R:

5’- GAACAACGATACACGGTGCTCTCCACAAACCAACTTCAAAATC-3’;

SuHmut5F:

5’- GCACTTGAGCCGCGAAATCCGACACTCACTACAAAGAAAACACTC-3’;

SuHmut5R:

5’- GAGTGTTTTCTTTGTAGTGAGTGTCGGATTTCGCGGCTCAAGTGC-3’;

SuHmut6/7F:

5’- CTGCTGAAAACAAATGTTCTCTATAGTTCTCTACTGTGTTTTATG-3’;

SuHmut6/7R:

5’- CATAAAACACAGTAGAGAACATAGAGAACATTTGTTTTCAGCAG-3’;

SuHmut8F:

5’- CTGTGTTTTATGAATGCACGATACCGCAGACAATTCACTTTGC-3’;

SuHmut8R:

5’- GCAAAGTGAATTGTCTGCGGTATCGTGCATTCATAAAACACAG-3’;

SuHmut9F:

5’- GGGATTAGTAAGAGATTAAGTGCTCTACCTGGACAAATGCTAG-3’;

SuHmut9R:

5’- CTAGCATTTGTCCAGGTAGAGCACTTAATCTCTTACTAATCCC-3’;

SuHmut10R:

5’- GGCGCTAATCAAAAAGGTTGTCCGAAACCGCTACCCTTTTATTC-3’.

The underlined sequences correspond to putative Su(H) binding sites. SuHmutfusR 5’-GTGTTGAAGTAGCTGGCAGTGACGTGGCGCTAATCAAAAAGGTTG-3’ was used to generate a PCR product for fusion with *gatae* basal promoter and GFP reporter, with underlined sequences overlapping with the *gatae* basal promoter.

Embryo culture and microinjections

Embryo culture and microinjection procedures were performed as described in McMahon et al., 1985. Reporter constructs were diluted to a concentration of 250 DNA molecules/pl. 10 pl of solution was injected into each egg.

GFP reporter quantification

Embryos were harvested at various stages of development after microinjection for GFP RNA quantification. Total RNA was extracted using Qiagen’s RNeasy micro kit (Qiagen Inc., Valencia, CA). RT-PCR was performed using ABI’s (Applied Biosystems, Foster City, CA) Taqman Reverse Transcription Reagents using random hexamer priming or Bio-Rad’s (Bio-Rad Laboratories Inc., Hercules, CA) iScript cDNA synthesis

kit. QPCR was performed using ABI's SYBR Green reagent or Bio-Rad's iTaq SYBR Green Supermix with ROX. All reactions were performed in triplicates to account for any pipetting errors. GFP RNA levels were first normalized to SpZ12 to account for differences in embryo numbers in each reaction and then normalized to the amount of DNA incorporated into the sea urchin embryo following procedure described in Revilla-i-Domingo et al., 2004. Qiagen RNeasy micro kit was used to extract genomic DNA, and SpZ12 was used as the single copy gene used in quantifying the amount of incorporated DNA. All curves presented portray relative levels of GFP RNA compared to *gatae* basal promoter levels.

Gel shifts

Gel shift mobility assays were performed using procedure described in Yuh et al. (2004). Oligo sequences used are:

Otx1/2F 5'-CAGTTACAAGGGATTAGTAAGAGATTAATGGGTGAAC-3';

Otx1/2MF 5'- CAGTTACAAGGGACCAGTAAGAGACCAATGGGTGAAC-3';

Otx3F 5'-GGGTAGCTTGGGATTACAACCTTTTTG-3';

Otx3MF 5'-GGGTAGCTTGGGACCACAACCTTTTTG-3';

and their reverse complements. Underlined sequences correspond to Otx binding sites.

RESULTS

Module 10 responds to perturbation of otx

Using the Otx consensus binding site TAATCY (Gan et al., 1995), three putative Otx binding sites were identified in module 10 (Fig. 3.1). Two of the binding sites lie in close proximity separated by five nucleotides, while the third site occurs singly. Perturbation of *otx* expression was accomplished with an Otx-engrailed fusion (Otx-en), in which the Otx DNA binding domain was fused to the engrailed repression domain, thereby turning Otx into a dominant repressor (Li et al., 1999). Fertilized eggs were injected with either module 10:GFP reporter (10) alone or co-injected with 10 and Otx-en mRNA (10+Otx-en) and observed at the mesenchyme blastula stage. Injection of the 10 resulted in 57% of embryos that expressed GFP (Table 3.1). Of the expressing embryos, 99% (192) displayed expression in the vegetal plate (Fig. 3.2A). Embryos co-injected

CCACTCGCAGAAGCCATTACTTTGATAAATTAGTTTTACATCACTAAATCAATATTCAATATTCTGGAAC
TTTGAGCAAAATCCGTTTTTGTTTTAGCGGCAAACCGGCGGGCTAAAGCACACGAACTCAAACATAAATC
TTGGCACGTGTACCCGTAGATTGGCGGTAAGCTTTGTAGTCATTCTACTGGCACCATCCTTACATACACT
CATTCTCCCACCTCCTTTTTTCTCTTTTGTCGCTAAATTTGATTTTGAAGTTGGTTTGTTCTTCTACCCG
TGTATCGTTGTTCATCGAGCAAATACGTATGTTGTTTGTCGCCGCACTTGAGCCGCGAAATAATCACAGC
ACTACAAAGAAAACACTCCGGCGCAAACTGCTGAAAACAAATCGTATGAATCATGTGAACTGTGTTTTAT
GAATGCCATCGCAAGCAGACAATTCACTTTGCTTTGTAAAGCTTTGTACTTCTATGAAAGCGATCTTGTA
CAGTTACAAGGGATTAGTAAGAGATTAATGGGTGAACCTGGACAAATGCTAGAATAAAAGGGTAGCTTGG
GATTACAACCTTTTTGATTAGCGCC

Figure 3.1. Sequence of module 10. The Otx binding sites are highlighted in purple and Su(H) binding sites in green. Sequences for probes used in gel shift experiments are underlined in black (Otx1/2) and red (Otx3).

Table 3.1. Expression patterns of injected embryos.

Construct	Total	GFP +[a](%)	Endomesoderm[b] (%)	Ectoderm[a] (%)
10	336	193 (57)	192 (99)	3 (2)
10+Otx-en	303	35 (12)	33 (94)	3 (9)
Otxmut	206	114 (55)	111 (97)	3 (3)
Otxmut+Otx-en	68	38 (56)	37 (97)	1 (3)
10+dNSu(H)	152	71 (47)	69 (97)	4 (6)
Suhmut	278	247 (89)	246 (99)	2 (1)
Otxsuhmut	250	190 (76)	188 (99)	5 (3)
GataBp	261	13 (5)	3 (23)	10 (77)

[a] Embryos are scored at the mesenchyme blastula stage for GFP fluorescence. GataBp is a reporter construct containing only the *gatae* basal promoter.
[b] Numbers in parentheses represent percentage of GFP positive embryos displaying expression in said cell type.

with 10+Otx-en displayed a significant reduction in the percentage of GFP expressing embryos to 12% (Fig. 3.2B and Table 3.1). Consistent with the function of Otx-en, of the remaining GFP positive embryos, 94% (33) are expressed in the endomesoderm. The small fraction of GFP expressing embryos is likely a reflection of the incomplete shutdown of the basal transcriptional apparatus in some embryos. QPCR quantification of GFP RNA on injected embryos also supports this observation: embryos co-injected with 10+Otx-en RNA contained lower levels of GFP than embryos injected with 10 only (data not shown), demonstrating that module 10 is downstream of *otx*.

To determine if the effect of Otx-en on module 10 was mediated by its Otx binding sites, we generated a construct that lacked all Otx binding sites (Otxmut). Co-injection of Otxmut and Otx-en mRNA restores the percentage of GFP expressing embryos to 56% (Fig. 3.2C and Table 3.1), a level similar to that of the WT construct, thereby confirming that one of Otx's functions is to directly activate *gatae*.

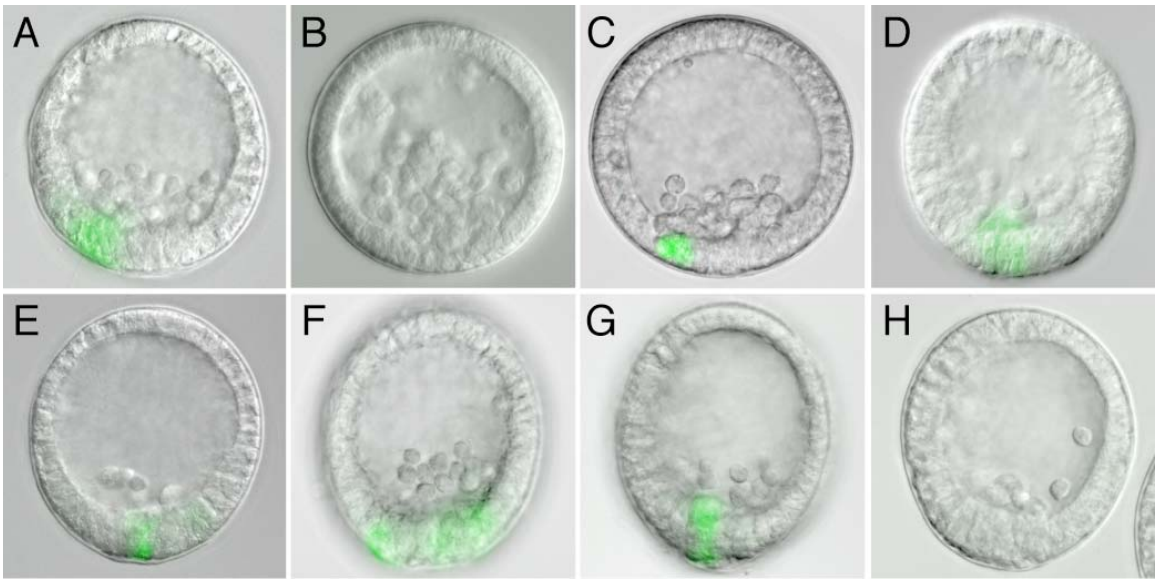


Figure 3.2. GFP fluorescence overlay images of mesenchyme blastula stage embryos. The embryos are injected with (A) 10; (B) 10+Otx-en; (C) Otxmut+Otx-en; (D) Otxmut; (E) 10+dNSu(H); (F) Su(H)mut; (G) OtxSu(H)mut, (H) Otxmut+dNSu(H).

Otx protein directly binds to module 10

To confirm that the Otx protein binds to module 10, two probes corresponding to the Otx binding sites in module 10 were designed for gel shifts (Fig. 3.1). The Otx1/2 probe corresponded to the sequence including and surrounding the two closely spaced Otx binding sites, while Otx3 probe spanned the single Otx site. Incubation of Otx1/2 and Otx3 oligonucleotides with 22 h sea urchin nuclear extract led to the formation of two protein complexes (Fig. 3.3), which could be competed away with the addition of increasing amounts of unlabeled Otx1/2 or Otx3 probe. Unlabeled Otx1/2M and Otx3M probes, which contain mutations in the Otx binding sites, were unable to compete away the complexes, demonstrating that the interaction is specific. Furthermore, addition of an

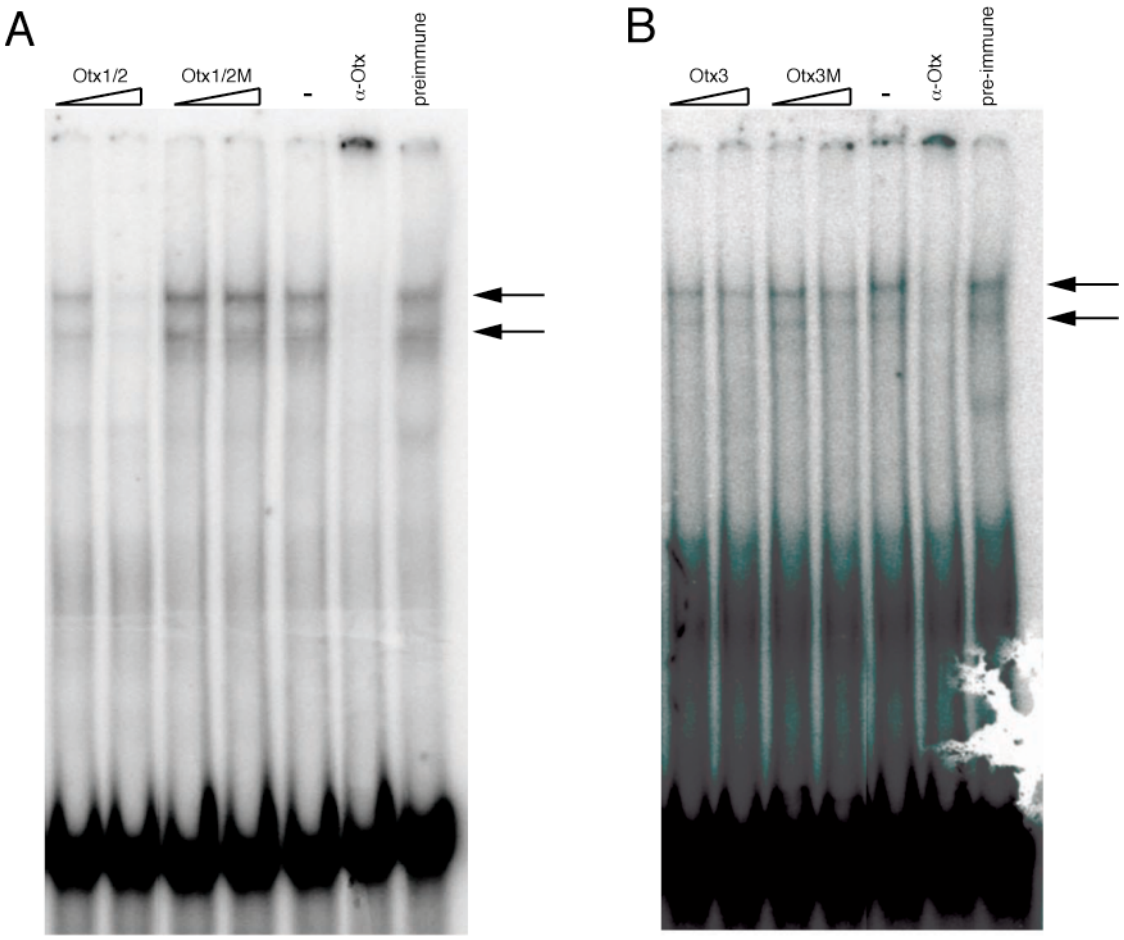


Figure 3.3. Gel shifts of Otx binding sites in module 10. Samples in each lane contained end labeled probes for (A), Otx1/2 and (B), Otx3 incubated with nuclear extract from 22 h sea urchin embryos. In lanes labeled Otx1/2 and Otx3, 1μM or 5μM of unlabeled competitor was added. Otx1/2M and Otx3M lanes included the addition of 1μM or 5μM of competitor in which the Otx binding sites were mutated. They symbol "-" refers to no addition of competitor. In the α-Otx lane, an antibody to the Otx protein was added. The arrows point to protein complexes that contained Otx.

antibody for Otx supershifts the protein complexes, verifying that the DNA binding complexes included the Otx protein.

Mutation of Otx binding sites reduces GFP reporter expression

As the mutation of Otx sites in module 10 did not abolish reporter expression, we decided to study Otxmut more closely. Embryos were injected with Otxmut and scored for GFP fluorescence at the mesenchyme blastula stage. Observation of Otxmut injected embryos did not reveal any obvious differences in localization of the GFP reporter as compared to module 10 injected embryos (Fig, 3.2D); 97% of GFP expressing embryos displayed endomesoderm expression (Table 3.1), and careful observation showed that GFP localized to both endoderm and mesoderm cells of the vegetal plate. To quantify GFP RNA, embryos injected with 10 or Otxmut were collected various stages of development and QPCR was performed on embryos in each group. The time course generated from QPCR experiments showed that GFP RNA levels in Otxmut injected

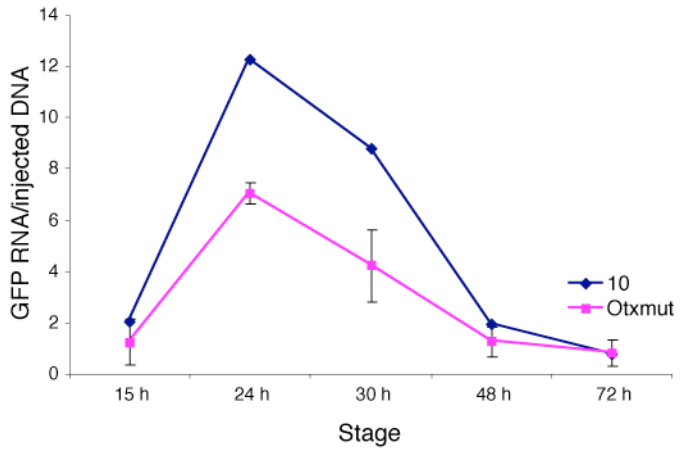


Figure 3.4. Otx binding site mutations led to decrease in reporter expression. Time courses of embryos injected with module 10 (blue) or Otxmut (pink). Error bars on the Otxmut curve represent two standard deviations from four embryo batches, normalized to module 10 values at each point.

embryos at 24 h were 57% of the WT reporter (Fig. 3.4). Therefore while mutation of Otx binding sites did not affect spatial expression of module 10, it did lower the level of transcription.

Module 10 is downstream of the Notch signaling pathway

Notch signaling was perturbed with a dominant negative form of the Su(H) protein (dNSu(H)). dNSu(H) contains mutations that led to its inability to bind DNA, and has been postulated to function by acting as a "sink" for the Notch intracellular domain and preventing the signal from being transduced to the nucleus (Ransick and Davidson, 2006). Co-injection of 10 and dNSu(H) RNA (10+dNSu(H)) resulted in decreased reporter expression to 47% of WT levels at 24 h (Fig. 3.5A). Similar results were observed in embryos co-injected with 10 and Notch MASO (data not shown). As an internal control, expression of the known Notch target *gcm* was also quantified. Observation of mesenchyme blastulas injected with 10+dNSu(H) did not detect differences in spatial expression of the GFP reporter with 97% of GFP expressing embryos displaying endomesoderm expression (Fig. 3.2E and Table 3.1).

Su(H) directly regulates module 10

Using the consensus motif YRTGRGAD (Ransick and Davidson, 2006) we identified a single putative Su(H) binding site. However, mutation of this binding site did

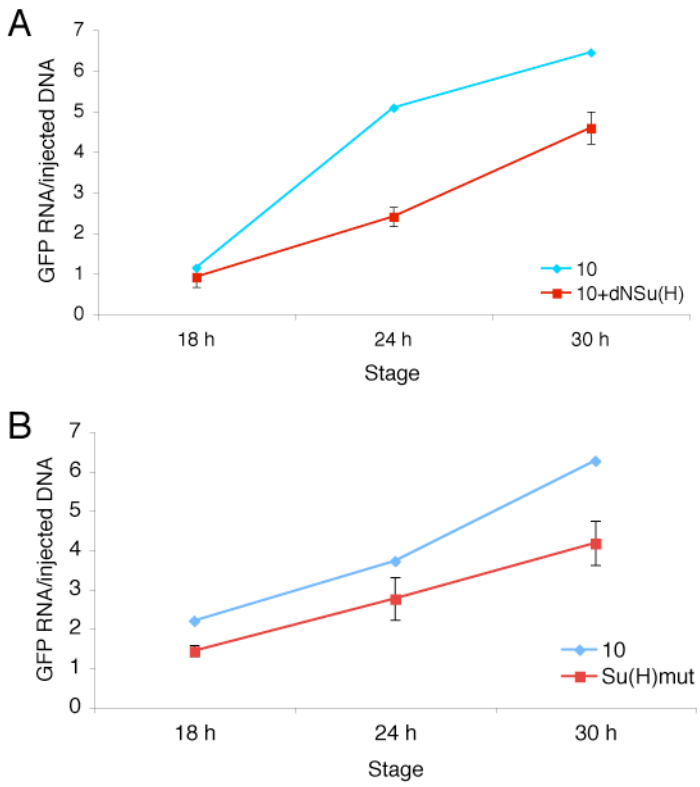


Figure 3.5. ***gatae*** **is directly regulated by Su(H).** Time courses for embryos injected with (A) 10 (blue) and 10+dNSu(H) (red); (B) 10 (blue) and Su(H)mut (red). Error bars on the curves in (A) and (B) correspond to two standard deviations from four embryo batches, normalized to module 10 values at each point.

result in any change in reporter expression (data not shown). As slightly different variants of Su(H) binding sites have been identified in different organisms , and the fact that the consensus Su(H) binding site is not always strictly followed (Flores et al., 2000), we undertook a more detailed analysis of possible Su(H) binding sites in module 10. Using a position weight matrix of mouse Su(H) binding sites (Tun et al., 1994), we performed a search in module 10 for any allowable sequence combinations. Each putative binding site was assigned a score based on the probability that a particular nucleotide occured at that position. Using this method we identified nine promising

binding sites (Fig. 3.1), and generated a construct in which all ten Su(H) binding sites were mutated (Su(H)mut). QPCR on embryos injected with 10 and Su(H)mut showed that Su(H)mut injected embryos displayed a 25% decrease in GFP RNA molecules as compared to WT (Fig. 3.5B). Similar to dNSu(H) perturbations, no differences in spatial expression between 10 and Su(H)mut injected embryos were detected, with only one out of 247 GFP positive embryos that did not exhibit endomesoderm expression (Fig. 3.2F and Table 3.1).

Otx and Notch signaling function additively to regulate gatae expression

Since the mutation of either Otx of Su(H) binding sites did not affect module 10's spatial expression, the regulation of *gatae* could not be mediated via Boolean AND logic. To explore the possibility that OR logic was utilized in *gatae* regulation, a construct lacking both Otx and Su(H) binding sites (OtxSu(H)mut) was generated and injected into fertilized eggs. 99% of OtxSu(H)mut embryos expressed GFP in the vegetal plate in a pattern similar to the WT reporter (Fig. 3.2G and Table 3.1). QPCR demonstrated that OtxSu(H)mut injected embryos expressed GFP at 50% of WT levels in the 24 h embryo (Fig. 3.6A). This represents a further reduction in GFP RNA levels comparison to the 57% for Otxmut and 75% for Su(H)mut injected embryos, the embryos still expressed GFP, suggesting that a third input is necessary for the regulation of *gatae.*

To determine if the third input was also Notch dependent, we co-injected the Otxmut construct and dNSu(H) mRNA (Otxmut+dNSu(H)), thereby removing Otx in *cis* and Notch in *trans*. Strikingly, a complete abolishment of GFP expression was observed

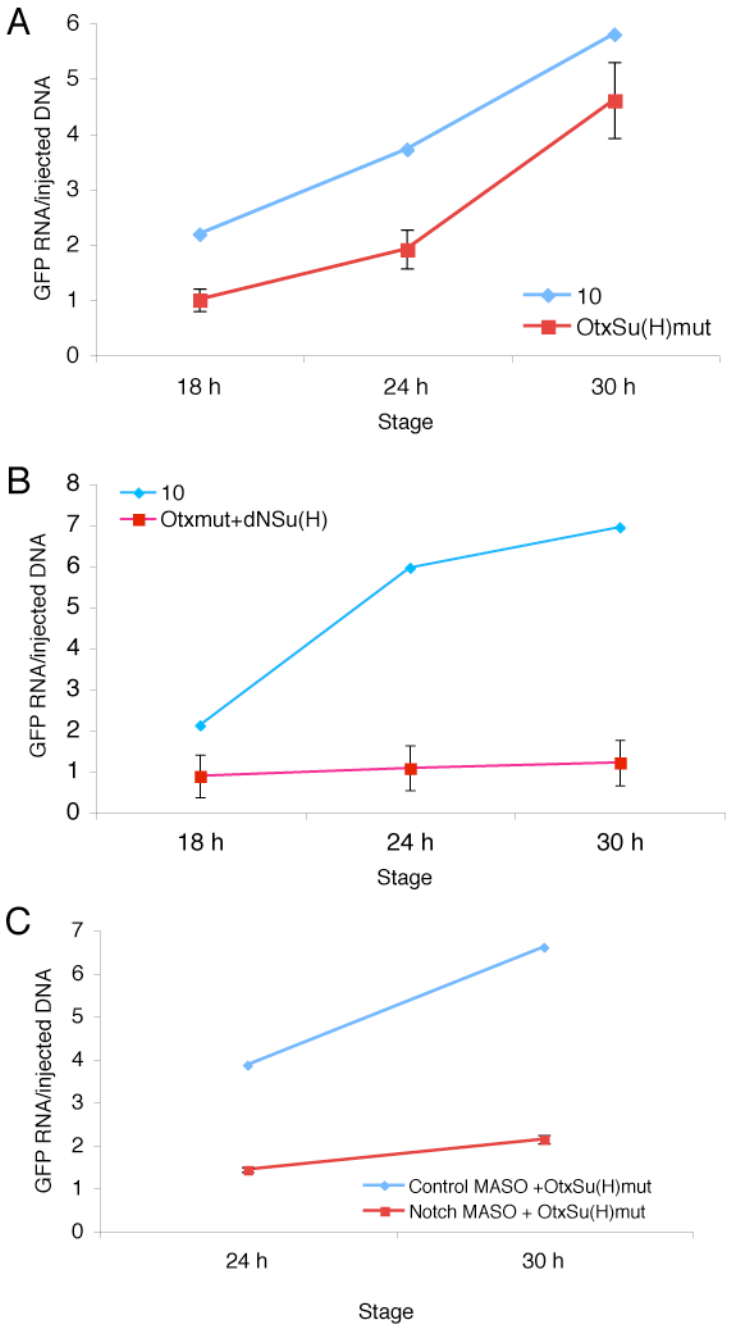


Figure 3.6. Otx and Notch signaling regulate *gatae* through OR logic. Embryos were injected with (A) 10 (blue) and OtxSu(H)mut (red); (B) 10 (blue) and Otxmut+dNSu(H) (red); (C) OtxSu(H)mut + Control MASO (blue) and OtxSu(H)mut + Notch MASO (red). Error bars on the lower curve correspond to two standard deviations from (A) and (B), four, and (C), two embryo batches, normalized to module 10 (A) and (B) or Control MASO + OtxSu(H)mut (C) values at each point.

in Otxmut+dNSu(H) injected embryos at mesenchyme blastula (Fig. 3.2H). QPCR on Otxmut+dNSu(H) is congruent with the GFP fluorescence data; the curve for Otxmut+dNSu(H) was completely flat, with GFP expression only at background levels (Fig. 3.6B). In a parallel argument, perturbation of Notch signaling in OtxSu(H)mut embryos should give the same result, which in fact we did observe by co-injecting OtxSu(H)mut and Notch MASO (Fig. 3.6C). This demonstrated that the third input is also downstream of Notch signaling, but not directly mediated through Su(H). Furthermore, Otx and Notch signaling regulate *gatae* by OR logic, whereby each input contributes to *gatae* expression levels but either one is able to recapitulate the entire repertoire of spatial expression.

DISCUSSION

Regulation of Spgatae

In this paper we have undertaken a *cis*-regulatory analysis of *gatae*'s early module, which is necessary and sufficient for *gatae* expression during the blastula stages. By using the predictions from the sea urchin endomesoderm GRN, then analyzing each input through the use of *trans* perturbations and *cis* binding site mutations, we were able to determine that the two inputs predicted by the endomesoderm GRN, *otx* and Notch signaling, occur by direct binding of the Otx and Su(H) transcription factors to module 10. Even though the removal of both inputs were able to eliminate reporter expression, binding site mutations did not abolish activity, suggesting that another unknown input is involved in *gatae* regulation. Co-injection of Otxmut and dNSu(H) RNA has shown that this heretofore unknown input is also downstream of Notch signaling.

Verification of the otx-gatae *cross-regulatory loop*

Auto and cross regulatory feedback loops are postulated to stabilize the gene expression program and ensure "lockdown" of the developmental state. In the sea urchin GRN, two cross-regulatory loops occur during endomesoderm specification. The first feedback loop involves *blimp1/krox* and *otx* and functions up to 18 h. Otx starts to activate *gatae* expression at 15 h, and by 18 h Gatae protein has accumulated to sufficient levels to activate *otx*. The end result is that those two genes are now locked in a cross-

regulatory embrace. Yuh et al. (2004) demonstrated that Gatae binding sites in *otx* module 15 bound the Gatae protein and mutations of those binding sites greatly reduced reporter expression. In this study we observed that the reciprocal interaction is also direct, demonstrating that one of the most important features of the endomesoderm GRN is wired at the DNA level.

The participation of GATA factors in cross-regulatory loops have been observed in many organisms. Notable is *gata6* and *nkx2.5* in chick heart development, in which these two genes have been shown to directly activate each other (Davis et al., 2000; Lee et al., 2004). Cross-regulatory feedback loops involving *gatae* in endoderm have been observed in zebrafish between *gata5* and *casanova*, *Xenopus gata5/6* and *sox17α*, and mouse *gata4* and *sox7* (Alexander et al., 1999; Futaki et al., 2004; Kikuchi et al., 2001; Murakami et al., 2002; Sinner et al., 2006). All these suggest an evolutionary conserved role of GATA factors as activators of the transcriptional program and stabilizers of the developmental state.

Authenticity and Completeness of endomesoderm GRN

As a model, the endomesoderm GRN needs to be constantly verified, expanded and updated. There are two major areas as to the refinement of the network: authenticity and completeness. Authenticity has been addressed through *cis*-regulatory analysis of various genes. Results from the *cis*-regulatory analysis of *gatae*, combined with data from other ongoing *cis*-regulatory projects suggest that the endomesoderm GRN is largely accurate in predicting direct gene interactions (Minokawa et al., 2005; Ransick

and Davidson, 2006; Yuh et al., 2004). The issue of completeness is more difficult to approach, simply because it is impossible to know what linkages are absent from a GRN rather than if an existing connection is accurate. In the case of *gatae*, the missing link appears to be due to a lack of completeness of the network. With the completed sequence of the *Strongylocentrotus purpuratus* genome (Sodergren et al., 2006) and a computational survey of transcription factors expressed during embryogenesis (Howard-Ashby et al., 2006), we are now well equipped with tools towards a complete endomesoderm GRN.

Use of OR logic in gene regulation

gatae is regulated by OR logic rather than the more commonly observed AND logic. Mutating binding sites for the either Otx or Su(H) did not eliminate reporter expression; rather each input by itself is able to drive expression in the endoderm and mesoderm. The elimination of one input, whether in *cis* or *trans*, only affects the amplitude of expression. *gatae* is not the only GATA factor regulated in this way. The *C. elegans end-1* gene, one of eleven GATA factors encoded in its genome, is regulated by *pop-1* and *skn-1* through OR logic as well. Mutation of a Pop-1 binding site was insufficient to eliminate reporter expression, however, if the Pop-1 mutation was coupled with skn-1 RNAi, reporter expression was virtually eliminated. A more common observation is the regulation of differentiation genes by OR logic. For example, the activation of the albumin promoter by Gata4 and Hnf3 works in this way (Bossard and Zaret, 1998).

An intriguing question is raised as to why *gatae*, unlike many other transcription factors, is regulated by OR logic. The choice of AND or OR logic may depend on sensitivity to the protein levels. In the case of differentiation genes, the level of enzymes are often critical for the function of an organ. For many transcription factors, their absolute concentrations are not important once they reach a threshold such that the factor can now bind to DNA, with regulation of their downstream targets mediated by the strength of existing binding sites. GATA factors are known to interact with other proteins that modulate their transcriptional activity (Fossett et al., 2001; Tsang et al., 1997). In addition, their associations with different cofactors are concentration dependent and of different strengths (Lu et al., 1999; Morin et al., 2000), suggesting that their protein levels are critical for their function. The use of OR logic in *gatae* regulation may provide a means of fine tuning transcription factor levels in different space and time.

ACKNOWLEDGEMENTS

We thank Cathy Yuh for technical help with gel shift experiments and Jongmin Nam for many useful discussions. We also appreciate Bill Klein for his generosity in providing us with the Otx-en construct and the Otx antibody. And as always, we are deeply grateful toward Pat Leahy for his assistance in acquiring and caring for the sea urchins, and to the animals who have sacrificed their lives in the pursuit of this work.

REFERENCES

Alexander, J., Rothenberg, M., Henry, G. L., and Stainier, D. Y. (1999). casanova plays an early and essential role in endoderm formation in zebrafish. *Dev Biol* **215,** 343-57.

Bossard, P., and Zaret, K. S. (1998). GATA transcription factors as potentiators of gut endoderm differentiation. *Development* **125,** 4909-17.

Calestani, C., Rast, J. P., and Davidson, E. H. (2003). Isolation of pigment cell specific genes in the sea urchin embryo by differential macroarray screening. *Development* **130,** 4587-96.

Davidson, E. H., Rast, J. P., Oliveri, P., Ransick, A., Calestani, C., Yuh, C. H., Minokawa, T., Amore, G., Hinman, V., Arenas-Mena, C., Otim, O., Brown, C. T., Livi, C. B., Lee, P. Y., Revilla, R., Rust, A. G., Pan, Z., Schilstra, M. J., Clarke, P. J., Arnone, M. I., Rowen, L., Cameron, R. A., McClay, D. R., Hood, L., and Bolouri, H. (2002a). A genomic regulatory network for development. *Science* **295,** 1669-78.

Davidson, E. H., Rast, J. P., Oliveri, P., Ransick, A., Calestani, C., Yuh, C. H., Minokawa, T., Amore, G., Hinman, V., Arenas-Mena, C., Otim, O., Brown, C. T., Livi, C. B., Lee, P. Y., Revilla, R., Schilstra, M. J., Clarke, P. J., Rust, A. G., Pan, Z., Arnone, M. I., Rowen, L., Cameron, R. A., McClay, D. R., Hood, L., and Bolouri, H. (2002b). A provisional regulatory gene network for specification of endomesoderm in the sea urchin embryo. *Dev Biol* **246,** 162-90.

Davis, D. L., Wessels, A., and Burch, J. B. (2000). An Nkx-dependent enhancer regulates cGATA-6 gene expression during early stages of heart development. *Dev Biol* **217,** 310-22.

Flores, G. V., Duan, H., Yan, H., Nagaraj, R., Fu, W., Zou, Y., Noll, M., and Banerjee, U. (2000). Combinatorial signaling in the specification of unique cell fates. *Cell* **103,** 75-85.

Fossett, N., Tevosian, S. G., Gajewski, K., Zhang, Q., Orkin, S. H., and Schulz, R. A. (2001). The Friend of GATA proteins U-shaped, FOG-1, and FOG-2 function as negative regulators of blood, heart, and eye development in Drosophila. *Proc Natl Acad Sci U S A* **98,** 7342-7.

Futaki, S., Hayashi, Y., Emoto, T., Weber, C. N., and Sekiguchi, K. (2004). Sox7 plays crucial roles in parietal endoderm differentiation in F9 embryonal carcinoma cells through regulating Gata-4 and Gata-6 expression. *Mol Cell Biol* **24,** 10492-503.

Gan, L., Mao, C. A., Wikramanayake, A., Angerer, L. M., Angerer, R. C., and Klein, W. H. (1995). An orthodenticle-related protein from Strongylocentrotus purpuratus. *Dev Biol* **167,** 517-28.

Howard-Ashby, M., Materna, S. C., Brown, C. T., Chen, L., Cameron, R. A., and Davidson, E. H. (2006). Gene families encoding transcription factors expressed in early development of Strongylocentrotus purpuratus. *Dev Biol* **300,** 90-107.

Kikuchi, Y., Agathon, A., Alexander, J., Thisse, C., Waldron, S., Yelon, D., Thisse, B., and Stainier, D. Y. (2001). casanova encodes a novel Sox-related protein necessary and sufficient for early endoderm formation in zebrafish. *Genes Dev* **15,** 1493-505.

Lee, K. H., Evans, S., Ruan, T. Y., and Lassar, A. B. (2004). SMAD-mediated modulation of YY1 activity regulates the BMP response and cardiac-specific expression of a GATA4/5/6-dependent chick Nkx2.5 enhancer. *Development* **131,** 4709-23.

Lee, P. Y., and Davidson, E. H. (2004). Expression of Spgatae, the Strongylocentrotus purpuratus ortholog of vertebrate GATA4/5/6 factors. *Gene Expr Patterns* **5,** 161-5.

Lee, P. Y., Nam, J., and Davidson, E. H. (2007). Exclusive developmental functions of *gatae cis*-regulatory modules in the *Strongylocentrotus purpuratus* embryo.

Li, X., Wikramanayake, A. H., and Klein, W. H. (1999). Requirement of SpOtx in cell fate decisions in the sea urchin embryo and possible role as a mediator of beta-catenin signaling. *Dev Biol* **212,** 425-39.

Lu, J. R., McKinsey, T. A., Xu, H., Wang, D. Z., Richardson, J. A., and Olson, E. N. (1999). FOG-2, a heart- and brain-enriched cofactor for GATA transcription factors. *Mol Cell Biol* **19,** 4495-502.

McMahon, A. P., Flytzanis, C. N., Hough-Evans, B. R., Katula, K. S., Britten, R. J., and Davidson, E. H. (1985). Introduction of cloned DNA into sea urchin egg cytoplasm: replication and persistence during embryogenesis. *Dev Biol* **108,** 420-30.

Minokawa, T., Wikramanayake, A. H., and Davidson, E. H. (2005). cis-Regulatory inputs of the wnt8 gene in the sea urchin endomesoderm network. *Dev Biol* **288,** 545-58.

Morin, S., Charron, F., Robitaille, L., and Nemer, M. (2000). GATA-dependent recruitment of MEF2 proteins to target promoters. *Embo J* **19,** 2046-55.

Murakami, A., Ishida, S., and Dickson, C. (2002). GATA-4 interacts distinctively with negative and positive regulatory elements in the Fgf-3 promoter. *Nucleic Acids Res* **30,** 1056-64.

Pancer, Z., Rast, J. P., and Davidson, E. H. (1999). Origins of immunity: transcription factors and homologues of effector genes of the vertebrate immune system expressed in sea urchin coelomocytes. *Immunogenetics* **49,** 773-86.

Ransick, A., and Davidson, E. H. (2006). cis-regulatory processing of Notch signaling input to the sea urchin glial cells missing gene during mesoderm specification. *Dev Biol* **297,** 587-602.

Rast, J. P., Cameron, R. A., Poustka, A. J., and Davidson, E. H. (2002). brachyury Target genes in the early sea urchin embryo isolated by differential macroarray screening. *Dev Biol* **246,** 191-208.

Revilla-i-Domingo, R., Minokawa, T., and Davidson, E. H. (2004). R11: a cis-regulatory node of the sea urchin embryo gene network that controls early expression of SpDelta in micromeres. *Dev Biol* **274,** 438-51.

Sinner, D., Kirilenko, P., Rankin, S., Wei, E., Howard, L., Kofron, M., Heasman, J., Woodland, H. R., and Zorn, A. M. (2006). Global analysis of the transcriptional network controlling Xenopus endoderm formation. *Development* **133,** 1955-66.

Sodergren, E., Weinstock, G. M., Davidson, E. H., Cameron, R. A., Gibbs, R. A., Angerer, R. C., Angerer, L. M., Arnone, M. I., Burgess, D. R., Burke, R. D., Coffman, J. A., Dean, M., Elphick, M. R., Ettensohn, C. A., Foltz, K. R., Hamdoun, A., Hynes, R. O., Klein, W. H., Marzluff, W., McClay, D. R., Morris, R. L., Mushegian, A., Rast, J. P., Smith, L. C., Thorndyke, M. C., Vacquier, V.

D., Wessel, G. M., Wray, G., Zhang, L., Elsik, C. G., Ermolaeva, O., Hlavina, W., Hofmann, G., Kitts, P., Landrum, M. J., Mackey, A. J., Maglott, D., Panopoulou, G., Poustka, A. J., Pruitt, K., Sapojnikov, V., Song, X., Souvorov, A., Solovyev, V., Wei, Z., Whittaker, C. A., Worley, K., Durbin, K. J., Shen, Y., Fedrigo, O., Garfield, D., Haygood, R., Primus, A., Satija, R., Severson, T., Gonzalez-Garay, M. L., Jackson, A. R., Milosavljevic, A., Tong, M., Killian, C. E., Livingston, B. T., Wilt, F. H., Adams, N., Belle, R., Carbonneau, S., Cheung, R., Cormier, P., Cosson, B., Croce, J., Fernandez-Guerra, A., Geneviere, A. M., Goel, M., Kelkar, H., Morales, J., Mulner-Lorillon, O., Robertson, A. J., Goldstone, J. V., Cole, B., Epel, D., Gold, B., Hahn, M. E., Howard-Ashby, M., Scally, M., Stegeman, J. J., Allgood, E. L., Cool, J., Judkins, K. M., McCafferty, S. S., Musante, A. M., Obar, R. A., Rawson, A. P., Rossetti, B. J., Gibbons, I. R., Hoffman, M. P., Leone, A., Istrail, S., Materna, S. C., Samanta, M. P., Stolc, V., Tongprasit, W., et al. (2006). The genome of the sea urchin Strongylocentrotus purpuratus. *Science* **314,** 941-52.

Tsang, A. P., Visvader, J. E., Turner, C. A., Fujiwara, Y., Yu, C., Weiss, M. J., Crossley, M., and Orkin, S. H. (1997). FOG, a multitype zinc finger protein, acts as a cofactor for transcription factor GATA-1 in erythroid and megakaryocytic differentiation. *Cell* **90,** 109-19.

Tun, T., Hamaguchi, Y., Matsunami, N., Furukawa, T., Honjo, T., and Kawaichi, M. (1994). Recognition sequence of a highly conserved DNA binding protein RBP-J kappa. *Nucleic Acids Res* **22,** 965-71.

Yon, J., and Fried, M. (1989). Precise gene fusion by PCR. *Nucleic Acids Res* **17,** 4895.

Yuh, C. H., Dorman, E. R., Howard, M. L., and Davidson, E. H. (2004). An otx cis-regulatory module: a key node in the sea urchin endomesoderm gene regulatory network. *Dev Biol* **269,** 536-51.

CONCLUSIONS

In this work I have characterized both the function and regulation of the *Spgatae* gene. *gatae* is a prime example of how a gene could be studied completely using a network approach, which began with the description of its expression pattern, followed by the determination of its position and role in the endomesoderm GRN. The inputs predicted by the endomesoderm GRN pertaining to *gatae* were then verified through *cis*-regulatory analysis, including the important cross-regulatory node with *otx*.

However, some questions still remain to be answered. Binding site mutation analyses on module 10 demonstrated that the *notch* input into *gatae* includes direct and indirect components. At this time the nature of the indirect *notch* input is unknown. However, this transcription factor is predicted to be expressed in the endomesoderm in the mesenchyme blastula stage. Its expression should be initiated before that of *gatae*, and its perturbation would lead to a downregulation of *gatae* expression. The completion of the sea urchin genome sequence and the comprehensive study of transcription factors expressed in the embryo will facilitate the identification of this factor. Previous work has shown that the 3´ end of module 10, which contains all the Otx binding sites and some Su(H) sites, was capable of driving vegetal specific expression on its own. The mutation of said Otx and Su(H) binding sites in this 3´ fragment eliminated all GFP reporter expression, implying that the unknown input is mediated through the 5´ end of module 10.

The second question pertains to the spatial regulation of *gatae*. Perturbation of either *otx* expression or Notch signaling alone did not affect the spatial expression of

gatae, suggesting that both inputs play a role in its spatial regulation. Notch signaling confers the ability to restrict spatial expression, whereas the situation surrounding *otx* is more complex. At this time it is unclear which of the *otx* transcription units is upstream of *gatae*. The *otx-α* transcription unit is expressed in the vegetal pole during mesenchyme blastula, while the *otx-β1/2* transcription unit is expressed in both the endomesoderm and oral ectoderm. Perturbations of *otx* function were performed with an Otx-en fusion protein, which did not discriminate between the different transcription units. Therefore depending on the *otx* transcription unit responsible for *gatae* activation, *gatae* spatial expression can either be completely accounted for by Otx activation, or requires an additional repressive input.

One possibility of such a possible repressive interaction is the Tcf protein. In the presence of Wnt signaling, Tcf associates with nuclear β-catenin to activate endomesoderm genes. It is unlikely that Tcf activation based on Wnt signaling is responsible for the activation of *gatae*, due to the fact that the expression of *wnt* and presence of nuclear β-catenin are detected in *gatae* expressing cells 6 h before *gatae* expression initiation. However, there remains the possibility that the Tcf/Groucho repressive complex might be involved in the negative restriction of *gatae* to the endomesoderm. Module 10 contains five putative binding sites for the Tcf transcription factor, whose mutations would determine whether such a repressive interaction is utilized in the spatial regulation of *gatae*.