cis-REGULATORY ANALYSIS OF THE SEA URCHIN *delta* GENE: VALIDATING THE ARCHITECTURE OF THE GENE REGULATORY NETWORK MODEL FOR MICROMERE LINEAGE SPECIFICATION

Thesis by

Roger Revilla-i-Domingo

In Partial Fulfillment of the Requirements

for the Degree of

Doctor of Philosophy

California Institute of Technology

Pasadena, California

2007

(Defended May 22, 2007)

To my father, my mother,

and my brothers, Dani, David and Ferran,

who provided the foundation and support

that made this possible

ACKNOWLEDGEMENTS

First of all I want to thank my advisor, Eric Davidson. There are many reasons for which I am grateful to him. Some are especially important to me. I want to thank him first for opening the door of his lab to a "backpacker" with no knowledge of biology whatsoever. Secondly, because at key points of this PhD he has believed in my ability to learn biology and to do experiments more than I have myself. Third, and most important to me, for having illuminated a path toward my deepest intellectual dream: that of understanding how life works. In my view he has opened a whole scientific life in front of me, both in a practical and in an intellectual sense. Finally, I want to thank him for his scientific intuition that has impressed me throughout the course of this PhD, and which I can best illustrate with an email I received from him on August 3, 2006, just a few months before discovering that *hesC*, a HES class factor, was the *repressor of micromeres*:

> "Dear Roger
> I still think the Id misadventure is not a total loss. Most likely it is interacting with an E2A type transcriptional activator, and so it will lead us to the activator if we can determine what its partner is. E2A proteins bind E-box target sites. But another class of bHLH factors binds a subset of these same sites and acts as a dominant repressor, the HES factors (sometimes in complex with E2A also). [...] R of mic could be a HES class factor, Id a level titrator of the E2A activator and the E2A and HES see the same site, sort of like the scratch drawings we were making last night. [...].
> Eric"

I also want to especially thank Ellen Rothenberg. Her teaching, her interests, and our numerous discussions have been an invaluable source of inspiration. Through her I have discovered the fascinating scientific questions to which I hope to devote my academic future. I am extremely grateful to her for the enormous help and very precious guidance.

Special thanks also to Jaume Gabarro, a mentor long before I started this PhD, who first motivated me to explore the world of scientific research, and whose work and ideas have had a great impact on my views of science.

I want to thank my thesis committee for their assistance throughout this PhD. And the Davidson lab in general for creating an incredibly diverse and fertile environment in which to grow, technically and intellectually, as a scientist. Special thanks to Paola Oliveri. Her work set the ground for the project that has constituted this thesis; she taught me how to microinject sea urchin embryos; and she has collaborated with me in what has turned out to be a very enjoyable and fruitful part of my PhD. I am also especially grateful to Joel Smith, my companion in my intellectual explorations. I thank him for sharing with me his scientific thoughts and ideas, and for feeding me tons of interesting papers. To Jongmin Nam for extremely useful discussions and criticism of the fourth chapter of this thesis in particular, and for teaching me how to scare people in general. To Veronica Hinman for joining me in the “punishment room”; and for publishing a paper that inspired me enormously. To Samdar Ben-Tabou de-Leon, for being right in telling me that the previous version of the introduction to this thesis needed to be completely rewritten. To Sagar Damle for his unbeatable WMISH technical service, for accompanying me in the scuba diving course, and for letting me take out from his drawer more than I put in. To Stefan Materna for the runs and hikes. To Albert Erives for being my first teacher in the lab. To Andy Ransick for teaching me the most useful microscopy techniques, like how to role alive embryos on a slide. To Julie Hahn for enormous help with preparing mutation constructs. To Jane Rigg for those many little things, and some not so little ones, that many times allowed me to make it over the hurdles. I will

especially remember that she gave me the first bowls and cups I ever owned when I arrived in Pasadena with empty pockets. And to Deanna Thomas for her patience with my picky and always rushed ordering requests.

I also very especially want to thank those who have been closest to me in the lab, physically but also, I believe, at a personal level: the past and present members of 017K. Jane Wyllie for understanding me from the very first day. Cesar Arenas-Mena for an attitude toward science that I admire. Takuya Minokawa for never getting upset (except once) to my asking way too many questions; for teaching me an enormous number of technical tricks; and most importantly for the daily discussions, through which I discovered vast and fascinating new horizons in biology. Qiang Tu for being a real “soul mate”; for being born the same day as me, May 26, 1975; for the bike trip to Mexico border, definitely my best adventure in California; and most of all, for his quiet, but always present readiness to help a friend. And finally, my most special thanks to the one person in the lab that has walked with me every step of this PhD: Pei Yun Lee. I thank her most of all for her unconditional and continued signs of friendship, which have more than once been the only good news of some tough weeks.

Many thanks to all my friends from Caltech. First and foremost, to Dr. Gilberto Hernandez for putting me back on track on many occasions, and because from him I have learned why the Spartans were seldom defeated. To Maria Eugenia Hernandez for sharing with me adventures, misadventures, stories and dreams. To Agustin Rodriguez for his views on humankind, which have influenced my own. To Javier for staying at my place every once in a while, and for talking to me about his projects. To Macarena for painting me once. To Gabi, Mihoko and Nonia for many nice lunches and brunches. To

my neighbors Juanse and Clari for providing free internet access. And also to Beto, Leila, Lavi, Theo, Roya, Julia, Jernej, Debrah and Paola.

Thanks to my friends from UCLA, whose parties have been the best: especially to Lan, but also to Paco, Nacho, Patricia, Laura, Marc, Marta, David, Israel, Nikos, Tom, and many others.

I also thank my friends from Manresa for not having forgotten me after all these years: el Xevi, el Xul.le, el Jorba, l'Elena Lloses, el Pol and la Celia, la Mariona, l'Emma, l'Elena Sixtu, el Marc, la Nadina, el Joan, el Dominguix, el Ferro, el Potra, la Celeste, la Nuria, l'Oriol, la Nuria, el Tom and many others. Also Albert Armenteres for being a great teacher.

Finally, I am very especially thankful to my family, for their strong moral support and encouragement. To my grandparents: avi Pasqual, iaia Maria, avi Albert and iaia Lola. To my uncles and cousins: Mariu, Merce, Aleix, Queralt, Gerard, Jaume, Montserrat, Pol, Dunia, Lis, Carme, Nil, Carla and Ricard. And to my sisters-in-law, Carla and Africa, for their precious friendship, and for making my brothers happy.

ABSTRACT

During the specification of the endomesoderm of the sea urchin embryo, mesodermal and endodermal cell types derive from common progenitors. The Delta signal, a ligand of the Notch receptor, serves as the spatial cue that triggers the segregation between these two fates. Expression of the *delta* gene exclusively in the micromere lineage early in development is essential for Delta to be able to correctly serve this role. According to a model of the gene regulatory network (GRN) underlying this process, the mechanism by which the micromere lineage is specified as a distinct domain, and by which the *delta* gene is expressed exclusively there, depends on a double repression system. A gene encoding a transcriptional repressor, *pmar1,* is activated specifically in the micromeres, where it represses transcription of a second repressor that is otherwise active globally. Zygotic expression of *delta* and micromere specific control genes depends on ubiquitous activators, and localization in the micromere lineage depends on repression by the second repressor everywhere else. In this model the second repressor is an unidentified gene, the existence of which is implied by numerous experiments. The work presented in this thesis experimentally validates the double repression architecture for micromere lineage specification and localization of *delta* expression. To prove the existence of the double repression system a genomic screen was devised to identify the gene playing the role of the second repressor. *hesC*, a transcription factor of the HES family, was found to be this gene. It is expressed at the right time and place, and its function is to repress micromere specific regulatory genes. To show that expression of *delta* in the micromere lineage depends on ubiquitous activators and HesC-dependent repression, the relevant *cis*-

regulatory module (CRM) was recovered. This CRM, named R11, is shown to be able to drive the expression of a reporter gene exclusively in the micromere lineage at the right time. Dissection of R11 and its response to blockade of *hesC* expression show that R11 expression depends on ubiquitously present activators, and on HesC-dependent repression everywhere except the micromere lineage.

TABLE OF CONTENTS

Dedication iii

Acknowledgements iv

Abstract viii

Table of Contents x

Introduction 1

Chapter 1: *Developmental Gene Network Analysis* 9

Chapter 2: *R11: A* cis*-Regulatory Node of the Sea Urchin Embryo Gene Network that Controls Early Expression of* SpDelta *in Micromeres* 38

Chapter 3: *A "Missing Link" in the Sea Urchin Embryo Gene Regulatory Network: hesC and the Double Negative Specification of Micromeres* 77

Chapter 4: *Regulatory Functions in the* delta *R11* cis*-Regulatory Element* 97

Conclusions 115

Appendix 1: *A Provisional Regulatory Gene Network for Specification of Endomesoderm in the sea Urchin Embryo* 120

Appendix 2: *A Genomic Regulatory Network for Development* 150

Appendix 3: *Supplementary Material* 161

INTRODUCTION

During the process of development an enormous amount of complexity arises from a single egg. Through specification, a large number of domains, each expressing a distinct set of genes, are established in a coordinated manner in time and space. This requires a sophisticated capability of processing information. The spatial information provided by asymmetries in the egg needs to be translated into the institution of distinct domains. At each succeeding stage, spatial and temporal cues from preceding stages need to be interpreted, and new cues need to be correctly positioned so that each domain can be further partitioned. A fundamental question, a small aspect of which will be addressed in this thesis, is how the genome controls this process.

The genomic loci of spatial and temporal information processing are the *cis*-regulatory modules (CRMs) that control when and where each gene is to be expressed (Davidson, 2006). The inputs are transcription factors localized in time and/or space in the embryo, which bind specific sequences within the CRM. Presence or absence of each input in the nuclei of each cell at each stage of development determines whether the gene is to be expressed or switched off. Maternally localized factors in the egg, and intercellular signaling molecules serve as spatial and temporal cues. These contribute to the control of gene expression by affecting the availability of specific transcription factors in specific nuclei. Because the expression of each transcription factor and signaling molecule is itself controlled by other transcription factors and signaling molecules, the mechanism by which the genome controls the specification process takes the form of a network of interactions among regulatory genes (Davidson, 2006).

A model for the gene regulatory network (GRN) controlling one particular process of development, namely, the specification of the endomesoderm in the sea urchin embryo was published (see appendices 1 and 2 of this thesis). The experiments on which this model was based are reviewed in chapter 1 of this thesis ("Developmental Gene Network Analysis"). Figure 0.1 illustrates the process of endomesoderm specification in the sea urchin embryo. Ultimately the endomesoderm consists of the skeletogenic mesenchyme, a few other mesodermal structures, and the endodermal gut (Fig. 0.1D). By the seventh cleavage (Fig. 0.1A), the cell lineages of the sea urchin embryo have been segregated into a canonical set of territories, each of which is destined to give rise to distinct cell types and in each of which a specific set of genes is already running (reviewed by (Davidson, 2006)). The animal pole half of the embryo now consists of blastomeres that produce only cells types ultimately found in the oral, aboral, and apical neurogenic ectoderm. The lower half consists of the veg1 ring, their sister cells of the veg2 ring immediately below, and the large and small mircromeres at the vegetal pole. The large micromeres will produce all the cells of the skeletogenic mesenchyme lineage, and the progeny of the veg1 and veg2 will produce the rest of the endomesoderm. At the swimming-blastula stage (Fig. 0.1B), the veg2 lineage has been segregated into two distinct domains: the inner veg2 ring consists of cells that will give rise to mesodermal cell types; and the rest of the veg2 domain will give rise to endodermal cells (Ruffins and Ettensohn, 1996; Ruffins and Ettensohn, 1993). At the mesenchyme blastula stage (Fig. 0.1C), the skeletogenic mesenchyme cells have ingressed into the blastocoel as primary mesenchyme cells (PMCs). After this, the veg1 progeny will become specified as endoderm (Logan and McClay, 1997), and gastrulation and skeletogenesis will follow.

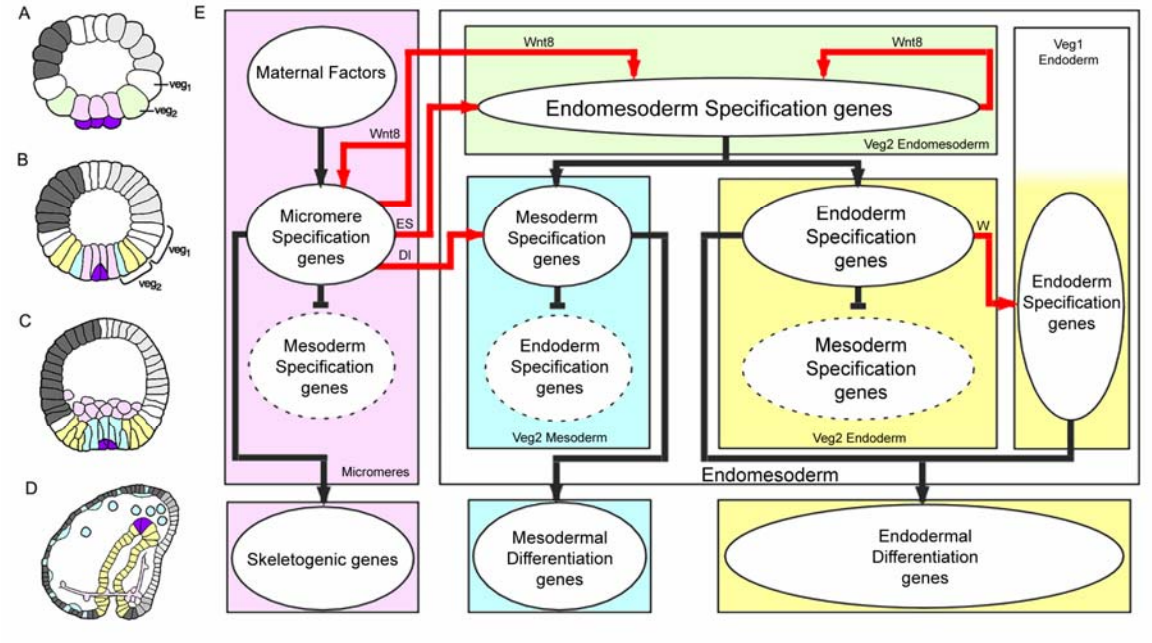


Fig. 0.1: Endomesoderm specification in the sea urchin embryo. (A-D) Schematic diagrams of sea urchin embryos displaying specified domains. The color coding shows the disposition of specified endomesoderm components: lavender indicates skeletogenic lineage; dark purple indicates small micromere precursors of adult mesoderm; green indicates endomesoderm lineage that later gives rise to endoderm, yellow, and mesoderm, blue; light grey indicates oral ectoderm; dark grey indicates aboral ectoderm; white indicates regions yet to be specified at the stages shown. **(A)** 7th cleavage embryo (about 10 h after fertilization). **(B)** Blastula stage embryo at about 9th cleavage (about 15 h after fertilization). **(C)** Mesenchyme blastula stage embryo (about 24 h after fertilization). **(D)** Late gastrula stage embryo (about 55 h after fertilization). The drawing shows the later disposition of all the endomesodermal cell types about midway through embryonic morphogenesis. **(E)** Process diagram describing endomesoderm specification events in the sea urchin embryo. Boxes represent domains of specification according to the color coding of their background. Ovals represent sets of genes that execute a particular developmental function. Arrows indicate that the set of genes in the oval where the arrow originates, triggers the developmental function executed by the genes in the oval where the arrow ends. In particular, red arrows represent signaling events. Barred lines indicate repression of the developmental function executed by the genes in the oval where the barred line ends. Developmental time in the process diagram runs from top to bottom in accordance with the stages represented by the schematic diagrams A-D. Abbreviations: ES, Early Signal; Dl, Delta; W, Wnt8.

Fig. 0.1E show a diagram that describes the specification events and the genetic functions that underlie the process just described. Two of these events are important for what follows. The first one relates to the specification of the skeletogenic mesenchyme lineage. These cells are autonomously specified (reviewed by Davidson, 2006). The spatial cue that triggers their specification consists of maternal factors that are localized at the vegetal pole of the egg. The second event is the segregation between the non-skeletogenic mesodermal cell types and the endodermal cell types from common progenitors. The spatial cue that triggers this event is a signaling molecule, Delta (Dl in Fig. 0.1E). The gene encoding this signal is exclusively expressed in the micromere lineage from late cleavage and during blastula stage. Localization of *delta* expression in these cells at this time is essential. Between 7^{th} and 9^{th} cleavage, the Delta signal activates a Notch receptor in adjacent endomesodermal (veg2) cells, and this is required for normal specification of mesodermal fate (McClay et al., 2000; Sweet et al., 1999; Sweet et al., 2002). Thus, the cells of the veg2 territory immediately adjacent to the micromere descendants are specified as mesoderm, while the rest of the cells of the veg2 territory will become endoderm.

The genomic apparatus that uses the spatial information in the egg to correctly position the expression of the Delta signal is the focus of this thesis. According to the endomesoderm GRN model, the mechanism by which the micromere lineage is specified as a distinct domain, and by which the *delta* gene is expressed exclusively there, depends on a double negative gate (Fig. 0.2A; Oliveri et al. 2002; with updates from (Ettensohn et al., 2003). Immediately after the micromeres are born, they express a gene, *pmar1*, in response to the maternal factors localized in the vegetal pole of the egg. This gene

encodes a transcriptional repressor. A second repressor, named *repressor of micromeres*, or *r of mic*, is proposed to be zygotically expressed everywhere in the embryo, except in the micromere lineage, where it is repressed by Pmar1. R of mic in turn represses the zygotic expression of *delta* and of at least three regulatory genes (*alx1*, *ets,* and *tbr*) which are responsible for the activation of the rest of the micromere skeletogenic program. The zygotic expression of *delta*, *alx1*, *ets,* and *tbr* depends on ubiquitously present activators, and its localization in the micromere lineage depends on repression by R of mic everywhere else in the embryo (Fig. 0.2A).

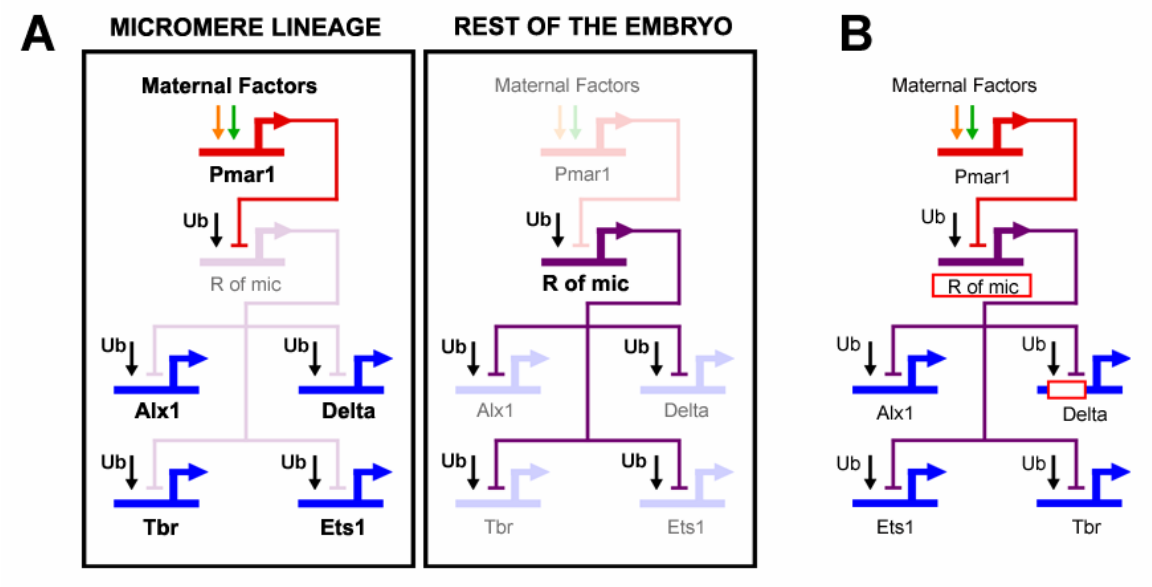


Fig. 0.2: The double negative gate for micromere lineage specification and localization of *delta* expression. (A) GRN model. Within the micromere lineage a distinct specification program is activated. In the rest of the embryo, the same program is actively repressed by R of mic. Genes that are active in the respective domain are shown in strong color. Genes that are inactive are shown in light color. **(B)** The red rectangles represent predictions of the GRN model.

The double negative gate of Fig. 0.2A is an explicit representation of how the genome processes spatial information and thereby controls the specification of the

micromere lineage and the expression of *delta*. It is a subcircuit of the GRN, i.e., a set of linkages with a particular developmental job (Davidson, 2006). Its architecture is revealing. The use of two repressors in regulatory tandem, and ubiquitous activators, is not the only way to produce a localized expression pattern. The alternative is of course localized expression of activators. But these two GRN architectures are not functionally equivalent. The double negative gate provides de facto, the active repression of regulatory states outside the correct domain of their expression. Thus, it acts as an "exclusion effect" (Oliveri and Davidson, 2007), actively ensuring silence of target genes in ectopic locations while at the same time ensuring their expression in correct locations.

A remarkable aspect of the subcircuit of Fig. 0.2A is that key components of it are predictions of the GRN model. Fig. 0.2B indicates two such predictions. One is critical to the specification of the micromere lineage in general: the existence of R of mic. The other one is critical specifically to the localization of *delta* expression in this lineage: that expression of *delta* in the micromere lineage depends on ubiquitous activators and on repression by R of mic. Both predictions are implied by numerous experimental observations (Oliveri et al., 2002). First, Pmar1 is expressed in the micromere lineage before zygotic expression of *delta*, *tbr*, *ets* and *alx1* starts in the same domain. Second, if expression of Pmar1 is forced to occur globally, then *delta*, *tbr*, *ets*, *alx1* (and downstream genes) are transcribed in all cells of the embryo, and all cells thereby adopt skeletogenic micromere fate. Third, exactly the same outcome follows if an mRNA encoding a dominantly repressive Engrailed fusion of the Pmar1 protein is globally expressed. Fourth, interfering with the expression of *ets*, *tbr* or *alx1* has no effect on the expression of *delta* or of each other at the relevant developmental stage. It follows that

the pmar1 gene product naturally acts as a repressor; that *delta*, *tbr*, *ets* and *alx1* are controlled by ubiquitous activators; and that localization of expression of these genes to the micromere lineage in normal embryos depends on their repression by R of mic everywhere else in the embryo. In particular, the possibility that any of these three genes is upstream of *delta*, or of each other, is ruled out.

To prove that the double negative gate for micromere lineage specification exists, and that it is responsible for the localization of expression of *delta* in the micromere lineage, it is necessary to experimentally validate the predictions of Fig. 0.2B. This means: a) to find the gene playing the role of *r of mic*; and b) to recover the relevant *delta* CRM and to demonstrate that it executes the predicted regulatory functions, i.e., ubiquitous activation and R of mic-dependent repression.

In this thesis I set out to validate the predictions of Fig. 0.2B. The first step was to recover the CRM that drives the expression of *delta* in the micromere lineage at the right time. I then could verify that the recovered CRM responds to the Pmar1 repression system as is predicted by the model. This work is described in chapter 2. It confirms that the localization of *delta* expression in the micromere lineage is transcriptionally controlled.

The second step was to find *r of mic* among all transcription factors in the sea urchin genome. I then could confirm that its properties and its function in the specification of the micromere lineage are as predicted by the GRN model. This work is presented in chapter 3.

The third step was to confirm that the CRM recovered in chapter 2 executes the predicted regulatory functions: activation by ubiquitously present factors, and R of mic-dependent repression. This is described in chapter 4.

The work described in chapter 4 strongly supports, but does not demonstrate, that the interaction between HesC and the recovered CRM is direct, as predicted by the GRN model. A demonstration that this is the case is the subject of ongoing work.

REFERENCES

Davidson, E. (2006). "The Regulatory Genome: Gene Regulatory Networks In Development and Evolution." Academic Press,

Ettensohn, C. A., Illies, M. R., Oliveri, P., and De Jong, D. L. (2003). Alx1, a member of the Cart1/Alx3/Alx4 subfamily of Paired-class homeodomain proteins, is an essential component of the gene network controlling skeletogenic fate specification in the sea urchin embryo. *Development* **130,** 2917-2928.

Logan, C., and McClay, D. (1997). The allocation of early blastomeres to the ectoderm and endoderm is variable in the sea urchin embryo. *Development* **124,** 2213-2223.

McClay, D., Peterson, R., Range, R., Winter-Vann, A., and Ferkowicz, M. (2000). A micromere induction signal is activated by beta-catenin and acts through notch to initiate specification of secondary mesenchyme cells in the sea urchin embryo. *Development* **127,** 5113-5122.

Oliveri, P., Carrick, D. M., and Davidson, E. H. (2002). A Regulatory Gene Network That Directs Micromere Specification in the Sea Urchin Embryo. *Developmental Biology* **246,** 209-228.

Oliveri, P., and Davidson, E. H. (2007). DEVELOPMENT: Built to Run, Not Fail 10.1126/science.1140979. *Science* **315,** 1510-1511.

Ruffins, S., and Ettensohn, C. (1996). A fate map of the vegetal plate of the sea urchin (Lytechinus variegatus) mesenchyme blastula. *Development* **122,** 253-263.

Ruffins, S. W., and Ettensohn, C. A. (1993). A Clonal Analysis of Secondary Mesenchyme Cell Fates in the Sea Urchin Embryo. *Developmental Biology* **160,** 285-288.

Sweet, H., Hodor, P., and Ettensohn, C. (1999). The role of micromere signaling in Notch activation and mesoderm specification during sea urchin embryogenesis. *Development* **126,** 5255-5265.

Sweet, H. C., Gehring, M., and Ettensohn, C. A. (2002). LvDelta is a mesoderm-inducing signal in the sea urchin embryo and can endow blastomeres with organizer-like properties. *Development* **129,** 1945-1955.

CHAPTER 1

Developmental Gene Network Analysis

Roger Revilla-i-Domingo and Eric H. Davidson

Published in *International Journal of Developmental Biology* **47:** 695-703 (2003)

ABSTRACT

The developmental process is controlled by the information processing functions executed by the *cis*-elements that regulate the expression of the participating genes. A model of the network of *cis*-regulatory interactions that underlies the specification of the endomesoderm of the sea urchin embryo is analyzed here. Although not all the relevant interactions have yet been uncovered, the model shows how the information processing functions executed by the *cis*-regulatory elements involved can control essential functions of the specification process, such as transforming the localization of maternal factors into a domain-specific program of gene expression; refining the specification pattern; and stabilizing states of specification. The analysis suggests that the progressivity of the developmental process is also controlled by the *cis*-regulatory interactions unraveled by the network model. Given that evolution occurs by changing the program for development of the body plan, we illustrate the potential of developmental gene network analysis in understanding the process by which morphological features are maintained

and diversify. Comparison of the network of *cis*-regulatory interactions with a portion of that underlying the specification of the endomesoderm of the starfish illustrates how the similarities and differences provide insights into how the programs for development work, and how they evolve.

KEY WORDS: *Gene network, genetic program, evolution and development, genomic regulatory system, sea urchin*

INTRODUCTION

The genetic programs that control the processes by which the body plans of animals are built were invented, and shaped, by the evolutionary process. How these programs work, if nothing else, is a matter of great curiosity. Because gene networks constitute the control systems for development, analysis of such networks explains both the process of development and the process by which development has evolved (Davidson, 2001).

Ultimately, development is the process by which the body plans of animals are laid down. Distinct cell types are produced in particular spatial domains, each with particular structural properties given by the distinct programs of gene expression that the cells execute. Through the process of specification each domain in the embryo obtains its developmental identity. Once specified, each domain will run through a progression of states of regulatory gene expression, leading to the establishment and ultimately the stabilization of the terminal programs of gene expression that give each cell type its unique properties.

Spatial cues are always required in order to trigger specification in development. These spatial cues sometimes consist of localized maternal regulatory factors that are distributed to particular cells with the egg cytoplasm, and are partitioned during cleavage. Alternatively they can also consist of signaling ligands produced by other cells, in consequence of their own prior state of specification. Ultimately, these spatial cues affect the course of events in development by causing the activation (or repression), in a certain region of the embryo, of particular genes encoding transcription factors. Through this process, new, more refined, domains of specification are created, and the complexity of the embryo increases. But although it is the spatial cues that trigger the events of spatial specification, the locus of programmatic control for each developmental event is the sequence of the particular *cis*-regulatory elements that respond to the inputs presented (Davidson, 2001).

cis-Regulatory elements can recognize the presence or absence of those transcription factors for which they contain specific binding sites. According to the set of inputs presented in each cell, the *cis*-regulatory elements of given genes control the expression of the gene in each domain of the embryo. Of particular importance are genes encoding transcription factors, and their *cis*-regulatory elements. Spatial information is translated by the *cis*-regulatory elements of these genes into distinct states of regulatory gene expression. It is the network of all these *cis*-regulatory interactions that is ultimately responsible for driving the process of development. To fully understand how the process of development is programmed in the genomic DNA, it will be necessary to unravel the network of regulatory interactions, and to analyze the information processing functions executed by each *cis*-regulatory element (Davidson, 2001).

The experiments reviewed here represent a step toward the goal of determining the complete network of DNA-based interactions that underlie one particular major process of development, namely, the specification of the endomesoderm of the sea urchin embryo. Given that evolution occurs by changing the program for development of the body plan, we also illustrate briefly how developmental gene network analysis sheds light on the process by which morphological features are maintained and diversify.

UNRAVELING THE GENE REGULATORY NETWORK THAT UNDERLIES THE PROCESS OF ENDOMESODERM SPECIFICATION IN THE SEA URCHIN EMBRYO

The armature of the network

Figure 1.1 illustrates the process of endomesoderm specification in the sea urchin embryo (Fig. 1.1A-D), and it shows a diagram (Fig. 1.1E) that describes the specification events and the genetic functions that underlie this process.

Ultimately, the endomesoderm consists of the endodermal gut, the skeletogenic mesenchyme and several other mesodermal cell types, including pigment cells (Fig. 1.1D). By the seventh cleavage cycle (Fig. 1.1A), the cell lineages of typical sea urchin embryos have been segregated into a canonical set of territories, each of which is destined to give rise to certain distinct cell types (Hörstadius, 1939; Cameron *et al*., 1987, 1991), and in each of which a distinct set of genes is already running (reviewed by Davidson *et al*., 1998; Davidson, 2001). The upper or animal pole half of the embryo now consists of blastomeres that produce only the cell types ultimately found in the oral

and aboral ectoderm. The lower half consists of the veg1 ring, their sister cells of the veg2 ring immediately below, and the large and small micromeres at the vegetal pole. In the undisturbed embryo, the large micromeres (the population of cells colored lavender in the diagram) will produce all the cells of the skeletogenic mesenchyme lineage, and the progeny of veg1 and veg2 will produce the rest of the endomesoderm. At the ciliated swimming-blastula stage (Fig. 1.1B), the veg2 lineage has been segregated into two distinct domains: the inner veg2 ring consists of cells that will give rise to mesodermal cell types, including pigment cells; and the rest of the veg2 domain will give rise to endodermal cells (Ruffins and Ettensohn, 1993, 1996). At the mesenchyme blastula stage (Fig. 1.1C), the skeletogenic mesenchyme cells have ingressed into the blastocoel, leaving behind a now fully specified central disc of prospective mesodermal cell types, and peripheral to them, the endodermal precursors (reviewed by Davidson *et al*., 1998). After this, the adjacent veg1 progeny will become specified as endoderm as well (Logan and McClay, 1997), and gastrular invagination ensues.

The mechanisms that trigger each one of the specification events that are symbolized by the colors in Fig. 1.1A-D are now reasonably well understood. The micromere lineage is autonomously specified as soon as these cells are formed at fourth cleavage (reviewed by Davidson *et al*., 1998). The spatial cues that trigger their specification are maternally localized. As soon as they are born, the micromeres emit a signal that, together with spatial cues that are autonomously localized, triggers the specification of the surrounding veg2 cells to endomesodermal fate (Ransick and Davidson, 1993, 1995). The segregation of veg2 between mesodermal and endodermal domains depends on a second signaling event from the micromeres that takes place at 7th-

9th cleavage, and is executed by the ligand Delta (Sherwood and McClay, 1999; Sweet *et al*., 1999; McClay *et al*., 2000; Sweet *et al*., 2002). The cells in the inner veg2 ring, which are exposed to the Delta signal from the micromeres, are specified as mesoderm. The rest of the veg2 cells will acquire endodermal fate. The result is that the initial crude pattern of specification, which defines veg2 as endomesoderm, has now been refined into two distinct specification states. Finally, another signaling event from the veg2 endoderm triggers the specification of the surrounding veg1 also as endoderm (Logan and McClay, 1997; Ransick and Davidson, 1998).

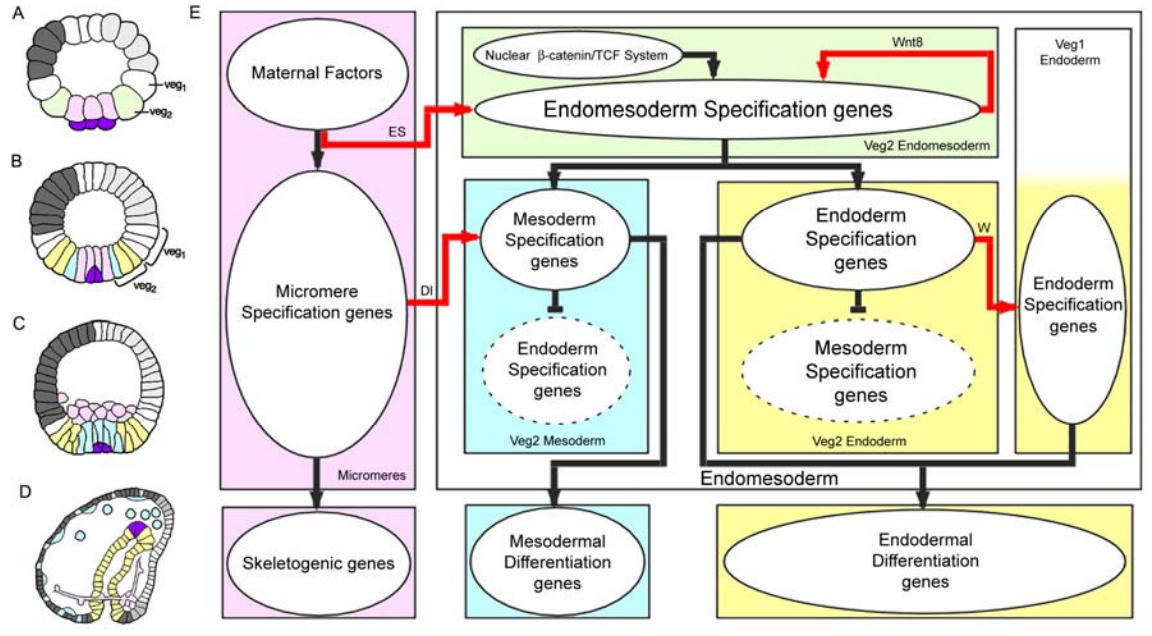


Fig. 1.1. Endomesoderm specification in the sea urchin embryo. (A-D) Schematic diagrams of sea urchin embryos displaying specified domains, from Davidson *et al*. (2002b). The color coding shows the disposition of specified endomesoderm components: Lavender indicates skeletogenic lineage; dark purple indicates small micromere precursors of adult mesoderm; green indicates endomesoderm lineage that later gives rise to endoderm, yellow, and mesoderm, blue; light grey indicates oral ectoderm; dark grey indicates aboral ectoderm; white indicates regions yet to be specified at the stages shown. **(A)** 7th cleavage embryo (about 10 h after fertilization). **(B)**

Blastula stage embryo at about 9th cleavage (about 15 h after fertilization). **(C)** Mesenchyme blastula stage embryo (about 24 h after fertilization). **(D)** Late gastrula stage embryo (about 55 h after fertilization). The drawing shows the later disposition of all the endomesodermal cell types about midway through embryonic morphogenesis. **(E)** Process diagram describing endomesoderm specification events in the sea urchin embryo. Boxes represent domains of specification according to the color of their background. The color coding represents the same endomesoderm components as in the schematic diagrams A-D. Ovals in the boxes represent sets of genes that execute certain developmental function. Arrows indicate that the set of genes in the oval where the arrow originates triggers the developmental function executed by the genes in the oval where the arrow ends. In particular, red arrows represent signaling events. Barred lines indicate repression of the developmental function executed by the genes in the oval where the barred line ends. Developmental time in the process diagram runs from top to bottom in accordance with the stages represented by the schematic diagrams A-D. "ES" stands for "Early Signal"; "Dl" stands for "Delta"; "W" stands for "Wnt8." Evidence is reviewed in Davidson *et al.* (2002a), and from P. Oliveri, A. Ransick, D.R. McClay and E.H. Davidson, unpublished data.

The knowledge summarized in Fig. 1.1E provides us with the armature on which the network of gene interactions is subsequently built. It tells us what specification functions must be executed by the genes in each domain: for example we know that the genes in the lavender box (Fig. 1.1E) must be able to translate the maternally localized spatial cues into a skeletogenic program of differentiation, and they must also be able to cause expression of the ligand Delta; and that the genes in the blue box must be able to listen to the spatial information given by the Delta signal in order to create a state of specification on which the mesodermal differentiation program is then installed.

The process diagram of Fig. 1.1E also serves another purpose. It tells us how we can interfere specifically with a certain specification event or domain, which is an essential tool in the enterprise of building the regulatory network, as we see below.

Useful as the knowledge contained in Fig. 1.1E might be, it should be made clear that this knowledge by itself does not provide us with any real understanding of the developmental process. Figure 1.1E by itself fails to show us the explicit mechanisms of specification, the instructions followed by each cell on its way to becoming specified. These instructions are encoded in the genomic DNA. It is the goal of the following to unravel the network of DNA-based interactions from which the instructions for development can be read.

Building the network of *cis*-regulatory interactions

In order to clothe with real genes the armature of interactions indicated in Fig. 1.1E, a major gene discovery effort was undertaken by performing several differential macroarray screens (Rast *et al*., 2000). The goal of each of these screens was to isolate cDNA transcripts that are differentially expressed in a given domain of the endomesoderm. To this end, different specification events were interfered with so as to generate populations of RNA transcripts lacking given classes of endomesodermal sequence, and these populations were compared to normal embryo RNA or to RNA from embryos in which the RNA populations contained larger amount of endomesodermal sequences than normal. By using a very sensitive subtractive hybridization technology on these populations of transcripts, probes were created in which sequences differentially expressed in the chosen endomesodermal domain were greatly enriched. These probes were then used to screen large-scale arrays of ~10^5 clone cDNA libraries (macroarrays) (Rast *et al*., 2000).

In order to determine the interactions among the different genes, a large-scale perturbation analysis was carried out, in which the expression of many genes was individually altered experimentally, and the effect on all other relevant genes in the network was then measured by quantitative polymerase chain reaction (QPCR) (Davidson *et al*., 2002a). Given the *cis*-regulatory interactions predicted by the QPCR experiments, direct *cis*-regulatory analysis is used to test the predicted network linkages, and in certain instances to unravel the key information processing functions executed by the relevant *cis*-regulatory elements.

THE *CIS*-REGULATORY NETWORK: THE CONTROL SYSTEM FOR THE SPECIFICATION PROCESS

A model for the process of endomesoderm specification is shown in Fig. 1.2 in the form of a network diagram that combines all significant perturbation data; information on time and place of gene expression, as determined by whole mount *in situ* hybridization (WMISH) and QPCR measurements; *cis*-regulatory data where available; and all the underlying information of experimental embryology.

At each *cis*-regulatory element in the model predicted regulatory interactions with the products of other genes in the network are indicated. Therefore each one of these predicted interactions can be experimentally tested by determining the presence and function of the relevant binding sites in the relevant *cis*-regulatory elements. The importance of this point is worth emphasizing. It means that eventually the *cis*-regulatory network can be turned into a solid, experimentally confirmed structure.

Even though not all the *cis*-regulatory interactions that underlie the specification of the endomesoderm of the sea urchin embryo have yet been identified, and even though not all the identified interactions have yet been tested, the model of Fig. 1.2 allows us to see how the network of *cis*-regulatory interactions controls the specification process. The model shows how the *cis*-regulatory interactions control the specification functions that need to be executed for the different domains of the endomesoderm of the sea urchin to become what they become.

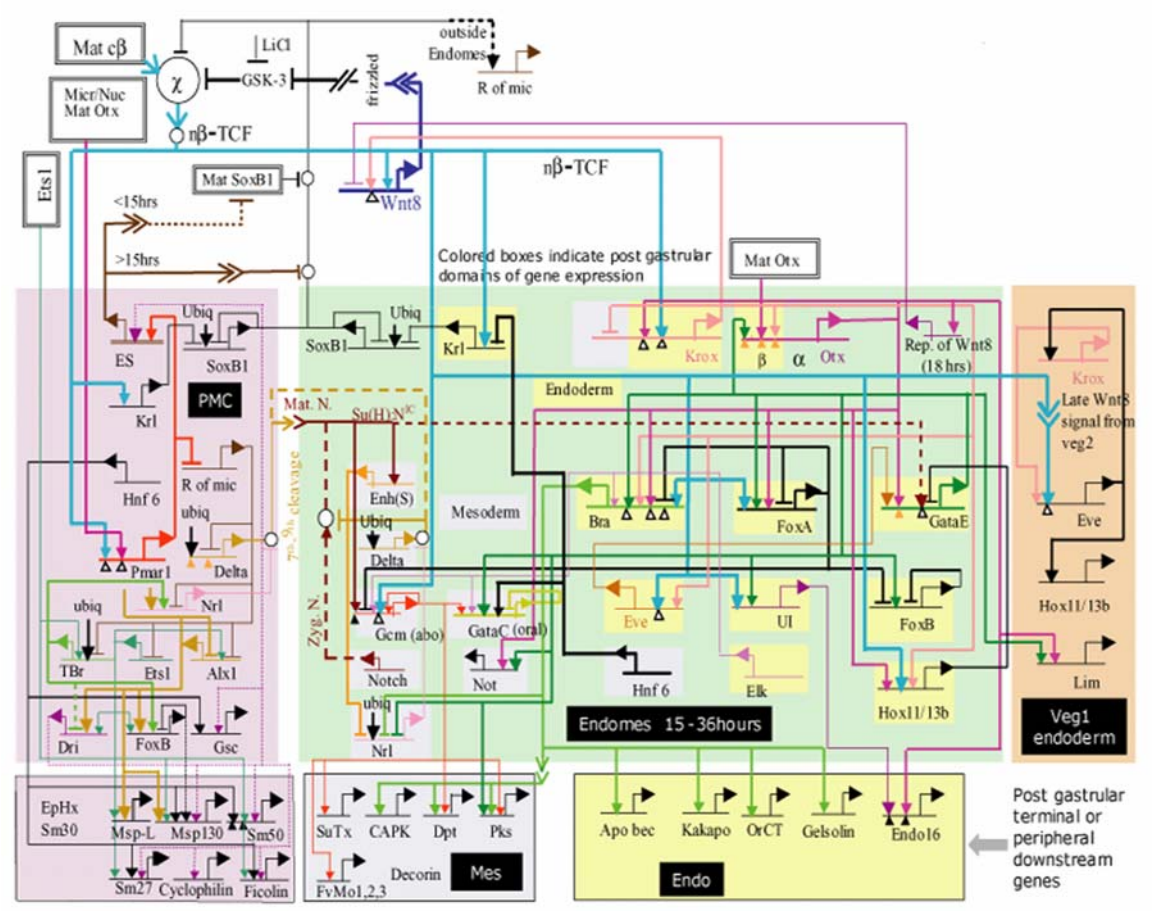


Fig. 1.2. Regulatory gene network model for endomesoderm specification from fertilization to just before gastrulation. This is a recent version of the model originally presented by Davidson *et al*. (2002a, 2002b). The current version of the model and the perturbation data on which it is based are available at www.its.caltech.edu/~mirsky/endomes.htm (End-mes Gene

Network Update) and www.its.caltech.edu/~mirsky/qpcr.htm (End-mes Network QPCR Data), respectively. Short horizontal lines from which bent arrows extend represent *cis*-regulatory elements responsible for expression of the genes named beneath the line. The arrows and barred lines indicate the inferred normal function of the input (activation or repression), as deduced from changes in transcript levels due to the perturbations. Each input arrow constitutes a prediction of specific transcription factor target site sequence(s) in the *cis*-regulatory control element. Dotted lines indicate inferred but indirect relationships. Arrows inserted in arrow tails indicate intercellular signaling interactions. Large open ovals represent cytoplasmic biochemical interactions at the protein level. The spatial domains are color coded as in Fig. 1.1, and genes are placed therein according to their loci of expression. The interactions at the top of the diagram, with no background color, are very early interactions. The rectangles in the lower tier of the diagram show downstream differentiation genes. "Ubiq" indicates an inferred ubiquitously active positive input. "Mat cβ" indicates maternal cytoplasmic β-catenin. "nβ/TCF" indicates nuclear β-catenin complexed with TCF. For further details see Davidson *et al*. (2002a, 2002b) and www.its.caltech./~mirsky/endomes.htm. For evidence see text, Davidson *et al.* (2002a, 2002b), Oliveri *et al*. (2002), Ransick *et al*., (2002), Rast *et al*., (2002), www.its.caltech./~mirsky/endomes.htm.

Interpreting the spatial cues: Specification of the micromeres

The network model of Fig. 1.2 indicates the mechanism by which maternal spatial cues in the micromeres are interpreted and translated into the specification state that is specific to the micromere lineage.

The genes *tbr, alx* and *ets*, are all known to activate a number of genes that are responsible for the differentiation of the micromere lineage into skeletogenic cells [Kurokawa *et al*., 1999; Fuchikami *et al*., 2002; Ettensohn *et al*., 2003 and www.its.caltech.edu/~mirsky/qpcr.htm (End-mes Network QPCR Data)]. Early in development, these three skeletogenic regulators are all kept silent everywhere in the embryo by a repressor gene (*r of mic*). At this time, *delta*, which is responsible for executing one of the micromere-specific developmental functions, is also repressed

everywhere in the embryo by the same repressor gene. Immediately after the micromeres are born at 4th cleavage, the *pmar1* gene is activated specifically in these cells. This gene has a repressor function that shuts down the expression of "*r of mic*". Now, *delta*, and the skeletogenic regulators *tbr, alx* and *ets* are allowed to be expressed exclusively in the micromeres, and as a result the skeletogenic program is set in train (Oliveri *et al*., 2002).

The mechanism just described ensures that once the *pmar1* is activated, the micromere specification program will be installed without the need for any further spatial cues. If *pmar1* is ectopically expressed everywhere in the embryo, the skeletogenic regulator *tbr*, the signaling ligand Delta, and the skeletogenic differentiation gene *sm50* are all also expressed everywhere, and the whole embryo is now expressing the functions normally executed only by the cells of the micromere lineage (Oliveri *et al*., 2002, 2003). The fact that *pmar1* is sufficient to establish the skeletogenic program, together with the fact that *pmar1* is activated by factors that are all either maternally present or autonomously localized in the micromere nuclei, tells us why the micromeres are autonomously specified. The most important general point is that the explanation of this embryological phenomenon is now provided in terms of the genomically encoded map of *cis*-regulatory interactions.

Refining the specification pattern: Specification of the pigment cells

The portion of the network in the diagram of Fig. 1.3 tells us the mechanism by which the pigment cells are specified and ultimately differentiated, according to the network model. The pigment cells arise specifically from the mesodermal cells of the veg2 domain (Ruffins and Ettensohn, 1993, 1996). The Delta signaling ligand produced by the

micromeres between 7th and 9th cleavage serves as the spatial cue that triggers the segregation of the mesodermal and endodermal fates of veg2 descendant cells (Fig. 1.3 A-B). Expression of the ligand Delta in the micromere descendants activates a Notch (N) receptor in the adjacent veg2 cells, which is required for normal specification of mesodermal fate in these cells (Sweet *et al*., 1999; McClay *et al*., 2000; Sweet *et al*., 2002). Localization of the Delta signal in the micromere descendants depends on the operation of the *pmar1* repression system, as explained above and illustrated in the diagram of Fig. 1.3. The response of Delta to the *pmar1* repression system depends on the *cis*-regulatory element named R11 (Fig. 1.3D-H) (R. Revilla-i-Domingo and E. Davidson, unpublished data). In normal embryos R11 drives expression of a reporter construct in the micromere descendants. When "*r of mic*" is repressed everywhere in the embryo by ectopic expression of *pmar1*, the *delta* gene is activated in every cell (Fig. 1.3 E-F), and in the same embryos R11 also drives expression of the reporter construct everywhere (Fig. 1.3 G-H) (R. Revilla-i-Domingo and E. Davidson, unpublished data).

Expression of the *gcm* gene begins in the single ring of mesoderm progenitor cells that directly receive the Delta micromere signal (Fig. 1.3B). As shown in the diagram of Fig. 1.3, activation of this gene depends on inputs from both the Notch signal transduction pathway, activated by the Delta signal, and (directly or indirectly) the nuclear β-catenin/TCF system (see diagram of Fig. 1.3), which is active in the whole of veg2 (Davidson *et al*., 2002a and A. Wikramanayake, unpublished data). The expression of *gcm*, therefore, reflects the creation of the new mesoderm-endoderm border, which did not exist before the Delta signal was received from the micromeres. The *cis*-regulatory element of *gcm* is responsible for integrating the spatial information provided by the

inputs from the Notch transduction pathway, and the β-catenin/TCF system. In normal embryos this element drives the expression of a reporter construct in a localized region in the vegetal plate. But if a portion of this element, containing binding sites for the Notch transduction pathway, is eliminated, expression of the reporter construct is expanded to a broader region that includes the whole of the veg2 domain (A. Ransick and E. Davidson, unpublished data). In other words, now the *cis*-regulatory element that controls *gcm* expression is 'blind' to the mesoderm-endoderm border established by the activation of the Notch transduction pathway.

Ultimately, the gene *gcm* is expressed in the pigment cells (a prominent subset of the veg2 mesodermal cell types), where it activates a number of differentiation genes (see diagram of Fig. 1.3), the products of some of which are likely to be required for synthesis of the red quinone pigment that these cells produce (Davidson *et al*., 2002b; Ransick *et al*., 2002; Calestani *et al*., 2003). If translation of *gcm* transcripts is blocked experimentally, the perturbed embryos show a perfectly normal morphology, except that they have no pigment cells (A. Ransick and E. Davidson, unpublished data).

The portion of the network depicted in Fig. 1.3 is a piece of the genetic program encoded in the *cis*-regulatory genomic sequence. It consists of a transcriptional apparatus, including R11 element, that localizes the Delta signal, and another transcriptional apparatus, including the Notch responsive element of the *gcm* gene, that interprets the signal. It explains why the cells in the inner ring of the veg2, and no others, give rise to pigment cells. And it also explains why elimination of expression of a single player in the program, *gcm*, results in the absence of the pigment cells. The overall function of this portion of the network is, first, to create a new domain of specification in

the embryo (the veg2 mesoderm), by setting a new border in the specification pattern; and then to install the program for pigment cell differentiation in the cells of the new domain. Other similar network subelements not yet resolved are undoubtedly responsible for differentiation of additional mesodermal cell types.

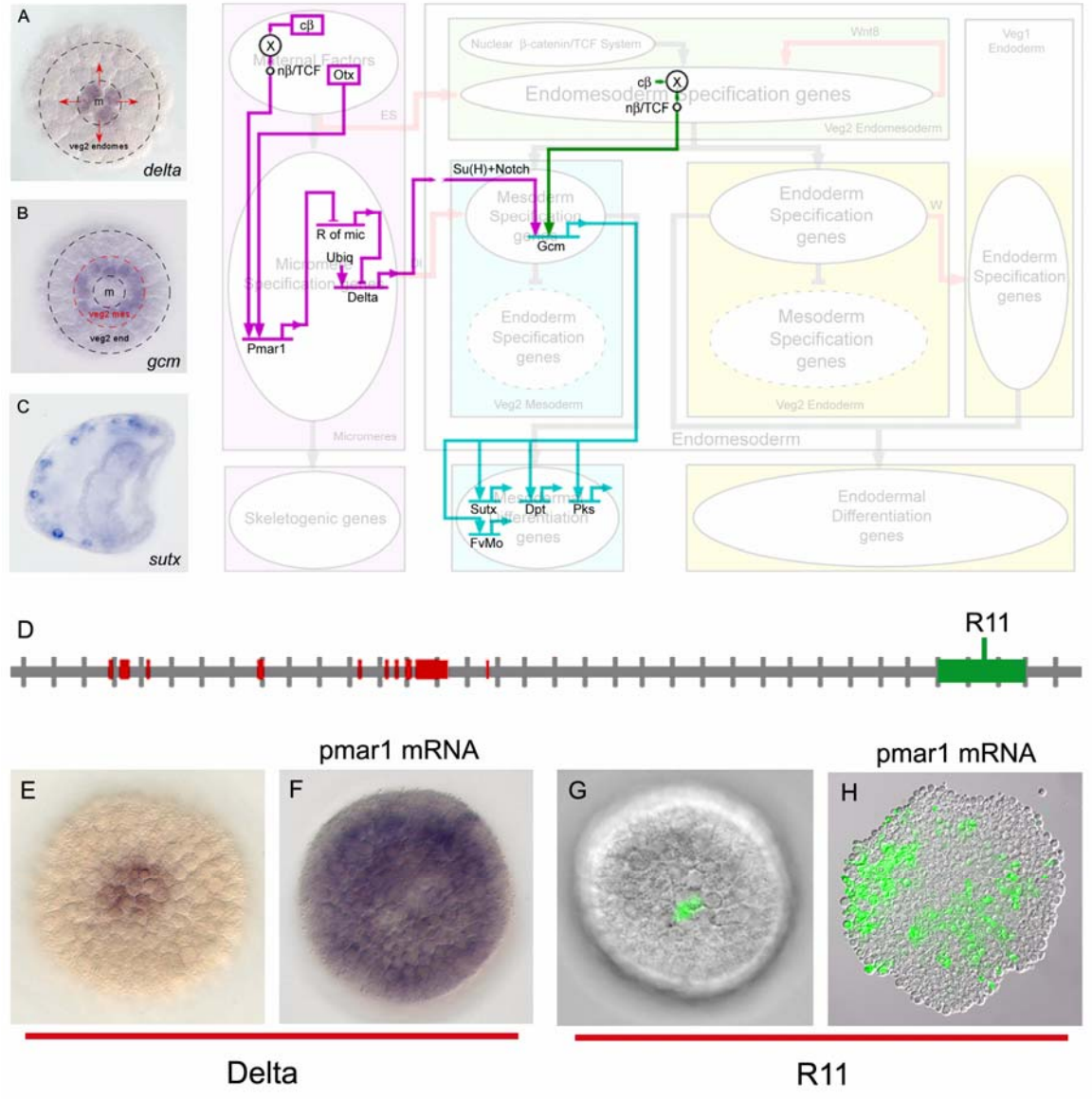


Fig. 1.3. Segregation of the veg2 domain into mesodermal and endodermal territories and installation of the pigment cell differentiation program. The diagram shows key interactions, extracted from the model of Fig 1.2, that control the segregation of the veg2 domain and the

installation of the pigment cell differentiation program. The dimmed background shows the process diagram of Fig 1.1E to indicate the domains where the interactions shown happen, and the developmental functions that the genes shown execute. **(A)** Between 7th and 9th cleavage the micromeres express the signaling ligand Delta (Oliveri *et al*., 2002; Sweet *et al.,* 2002). The figure shows a whole mount *in situ* hybridization photomicrograph, from P. Oliveri, displaying the expression of delta gene 12 h after fertilization (around 8th cleavage). "m" indicates micromeres domain. Red arrows indicate the signaling event from the micromeres to the surrounding veg2 endomesodermal cells. **(B)** The veg2 endomesodermal cells that receive the Delta signal from the micromeres become specified as mesoderm, and express the gene *gcm*; the rest of the veg2 endomesodermal cells become specified as endoderm (Sherwood and McClay, 1999; Sweet *et al*., 1999; McClay *et al*., 2000; Ransick *et al*., 2002; Sweet *et al*., 2002). The figure shows a whole mount *in situ* hybridization photomicrograph, modified from Ransick *et al*. (2002), displaying the expression of *gcm* gene 12 h after fertilization (around 8th cleavage). The red dotted circle indicates the newly formed border that segregates the veg2 domain into mesodermal and endodermal territories. **(C)** Ultimately, a subset of the veg2 mesodermal cells differentiate into pigment cells, and express the gene *sutx* (Calestani *et al*., 2003), among other pigment cell differentiation genes. The figure shows a whole mount *in situ* hybridization photomicrograph, modified from Calestani *et al*. (2003), displaying the expression of *sutx* gene in a gastrula stage embryo. **(D-H)** The *cis*-regulatory element R11 controls the localization of *delta* gene expression in the micromeres. **(D)** R11 element consists of a sequence of genomic DNA near the coding sequence of the Delta gene. Each tic on the horizontal grey line representing genomic sequence demarcates 1 kb from the previous tic. 5' direction is to the left. Red blocks on the sequence indicate positions of the *delta* gene coding sequence. The green box on the sequence indicates the position of the R11 element. **(E-F)** *pmar1* mRNA injection results in *delta* expression everywhere in the embryo. The figures show whole mount *in situ* hybridization photomicrographs, modified from Oliveri *et al*. (2002), comparing the expression of *delta* gene in normal blastula stage embryos **(E)**, and embryos that have been injected with *pmar1* mRNA **(F)**. **(G-H)** R11 element is responsible for localizing the expression of *delta* gene in the micromeres of normal embryos, and for driving the expression of the gene in every cell of embryos that have been injected with *pmar1* mRNA (R. Revilla-i-Domingo and E. Davidson, unpublished data). The photomicrographs compare the expression of the GFP reporter gene in blastula stage embryos that have been injected with R11 reporter construct **(G)**, and embryos that have been injected with *pmar1* mRNA in addition to R11 reporter construct **(H)**.

Stabilizing states of specification: The endoderm

Figure 1.4 illustrates the process by which the veg2 endoderm is specified. The veg2 lineage is born at 6th cleavage. By this time, the two spatial cues that trigger the specification of veg2 as endomesoderm are already operating. These initial cues consist of the autonomous nuclearization of β-catenin, which is a cofactor of the Tcf transcription regulator required for Tcf to function as a gene activator, and the early micromere signal (Ransick and Davidson, 1993, 1995; Logan *et al*., 1999). Two regulatory subcircuits execute the process by which the zygotic transcriptional apparatus interprets these initial cues, and by which it establishes an endomesodermal state of specification (Fig. 1.4A). The β-catenin/Tcf input activates the *krox* gene (Davidson *et al*., 2002b). This gene stimulates expression of *wnt8* gene and one of the transcription units of the *otx* gene. Wnt8 is a ligand which activates the β-catenin/Tcf system, and is itself a target of the β-catenin/Tcf input. This implies an autoreinforcing Tcf control loop, which is set up within the endomesodermal domain once this is defined (Davidson *et al*., 2002a). So, the result of the stimulation of *wnt8* expression, first by the β-catenin/Tcf system and later by *krox*, is to transfer control of the β-catenin/Tcf system from the autonomous cytoplasmic mechanism by which its activity was initiated to a zygotically controlled, intercellular signaling mechanism operating among the cells of the endomesoderm. The "community effect" (as defined by Gurdon, 1988; Gurdon *et al*., 1993) established by this regulatory subcircuit (dark blue connections in Fig. 1.4A) takes the cells out of a condition of alternative transcriptional possibility that is their initial condition, and locks them into a stable state of gene expression.

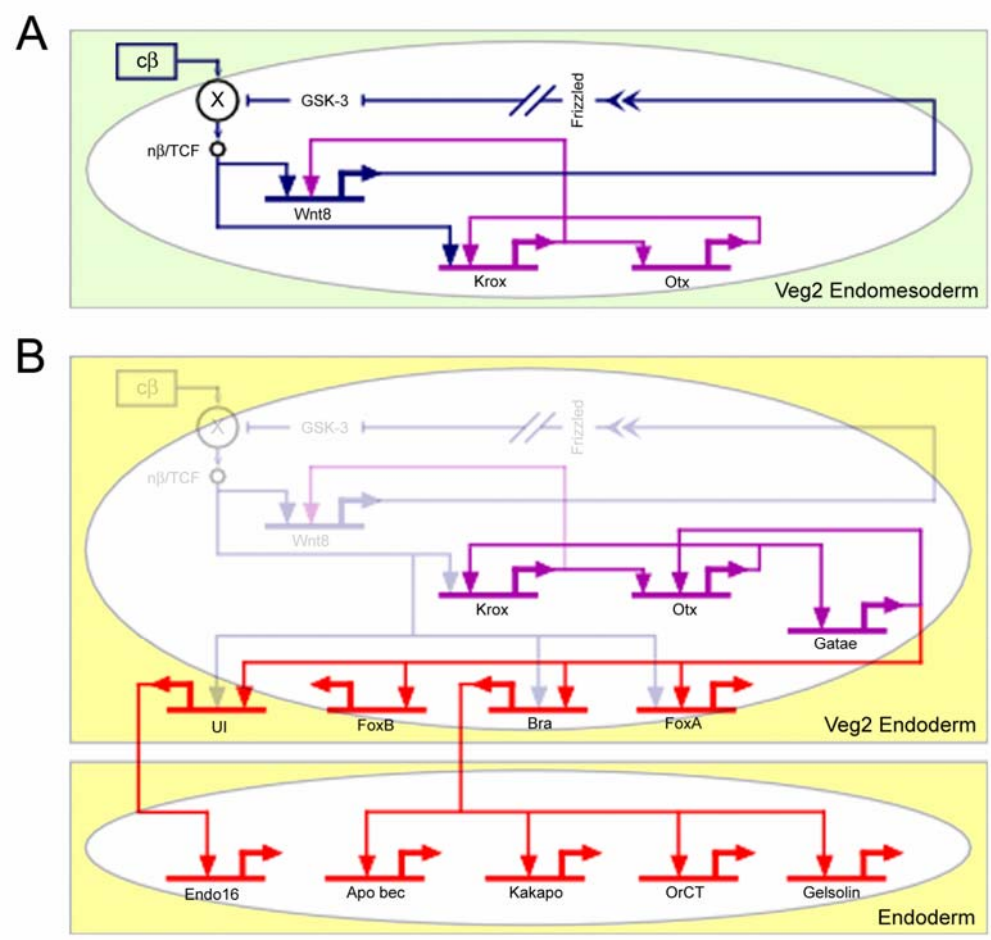


Fig. 1.4. Stabilization of the endomesoderm specification state and installation of the endoderm differentiation program. The diagram shows key interactions, extracted from the model of Fig 1.2, that control the stabilization of the endomesoderm state of specification and the installation of the endoderm differentiation program. **(A)** The box with green background shows the interactions that operate in the veg2 endomesoderm domain up to about 9th cleavage. Nuclearization of β-catenin is autonomous, and results in the activation of two regulatory subcircuits. Dark blue subcircuit: Wnt8 intercellular signaling among cells of the veg2 domain stimulates the nuclearization of β-catenin and establishes a "community effect," which defines and locks the endomesodermal state of specification in the veg2 cells. Purple subcircuit: *krox* and *otx* cross-regulate, which results in a reinforcing loop that renders the endomesoderm state of specification independent of the initial inputs. **(B)** The box labeled "Veg2 Endoderm" shows the interactions that operate in the veg2 endoderm domain, from about 9th cleavage to mesenchyme blastula stage. *Gatae* is added to the *krox-otx* feedback loop (purple interactions), and together with β-catenin/TCF system, installs the endoderm specification program (red interactions). When

β-catenin/TCF/Wnt8 inputs disappear, the stabilization loop maintains the endodermal specification program active, which eventually results in the activation of endodermal differentiation genes (lower box in the diagram labeled "Endoderm").

The *otx* gene stimulates expression of the *krox* gene. A regulatory subcircuit consisting of *otx* and *krox* cross-regulation produces a transcription-level stabilization of the endomesodermal regulatory state (purple connections in Fig. 1.4A) (Davidson *et al*., 2002a). The *otx* gene also provides an input into the *gatae* gene, which in turn has an input back into *otx* gene. This is a further positive feedback that links the *gatae* gene into the stabilization circuitry (purple connections in Fig. 1.4B). The *gatae* gene plays an important role in endoderm specification (red connections in Fig. 1.4B), since, together with the β-catenin/Tcf system, it is responsible for the activation of many of the known endodermal regulators, including the *bra, foxA* and *ui* genes (Davidson *et al*., 2002a and P. Y. Lee and E. Davidson, unpublished data). The FoxA transcription factor is a repressor that has multiple roles in the spatial control of gene expression patterns in the endoderm; Bra results in the activation of endodermal differentiation genes which are involved in cell motility and are needed for gastrulation and invagination to occur (Gross and McClay, 2001; Rast *et al*., 2002); the UI factor directly controls expression of *endo-16* (Yuh *et al*., 2001), which encodes a differentiation protein that is secreted in the lumen of the midgut. The crucial role that *gatae* plays in the specification of the endoderm explains the phenotype shown by embryos in which translation of the *gatae* transcripts has been blocked. This treatment produces a severe interference with endoderm specification and gut development (P. Y. Lee and E. Davidson, unpublished data).

During the late blastula stage, β-catenin disappears from the nuclei of the veg2 endodermal domain (Logan *et al*., 1999). But by this time, a network of stable intergenic interactions has been installed, so that the β-catenin inputs used earlier to set up transcriptional specification are no longer needed (Fig. 1.4B).

We see here that the *cis*-regulatory interactions control the operation of at least three different regulatory devices that are directly responsible for establishing at least part of the endoderm differentiation program. The first device consists of the "community effect," which first defines and then locks on the endomesodermal specification state in the veg2 domain (dark blue connections in Fig. 1.4A). The second device depends on a feedback loop, including *krox* and *otx* (purple connections in Fig. 1.4A), which generates a robust and resilient regulatory structure in the already defined endomesoderm domain. The third device consists of the addition of *gatae* to the *krox-otx* feedback loop (purple connections in Fig. 1.4B), which ensures the operation of many endodermal regulatory genes in the endoderm. The result is a control system that drives the specification process forward as a progression of states, and it prevents it from reversing direction when the initial cues that trigger the specification process disappear. Progressivity and stability are fundamental properties of the developmental process. They derive from regulatory devices consisting of assemblages of *cis*-regulatory interactions.

UNDERSTANDING DEVELOPMENT AND EVOLUTION

Developmental and evolutionary processes both have their root in the heritable genomic regulatory programs that determine how the body plan of each species is built (Davidson,

2001). It has been clear for a long time that the evolution of body plans has occurred by change in the genomic programs for the development of these body plans (Britten and Davidson, 1971), and it is now clear that we need to consider this in terms of change in the regulatory devices that execute these programs. The bilaterians all rely on essentially the same repertoire of regulatory genes to control the developmental organization of their body plans. Analysis of *cis*-regulatory networks affords the means to focus on the significance of preserved uses of these genes, and on the exact consequences of differences in their use (Davidson, 2001).

Figure 1.5 compares the way certain genes are utilized in the specification of the endomesoderm of two different bilaterians, namely, the sea urchin and the starfish. All genes in Fig. 1.5, except for *tbr*, are central elements that control the specification of the endoderm in the sea urchin (see Fig. 1.2 and Fig. 1.4). The *tbr* gene, on the other hand, is activated exclusively in the micromere derived skeletogenic cells (see Fig. 1.2) (Croce *et al*., 2001; Fuchikami *et al*., 2002; Oliveri *et al*., 2002). Its regulation depends on other genes specifically expressed in the micromere lineage (Oliveri *et al*., 2002), and in turn, it drives expression of larval skeletogenic differentiation genes (Davidson *et al*., 2002a; Oliveri *et al*., 2002 and www.its.caltech.edu/~mirsky/endomes.htm). While the formation of the endoderm is at least superficially similar in the two species (Fig. 1.5A), starfish embryos do not have a micromere lineage, nor do they produce a larval skeleton (Fig. 1.5A).

Figure 1.5B shows that the *cis*-regulatory interactions that constitute the endodermal three-gene stabilizing loop in the sea urchin (see Fig. 1.4B), is found in identical form in the starfish (connections in bold in Fig. 1.5B) (Hinman *et al*., 2003).

This set of identical *cis*-regulatory interactions must serve conserved evolutionary roles, since the possibility of convergence is ruled out by the number of similar functional starfish and sea urchin *cis*-regulatory interactions.

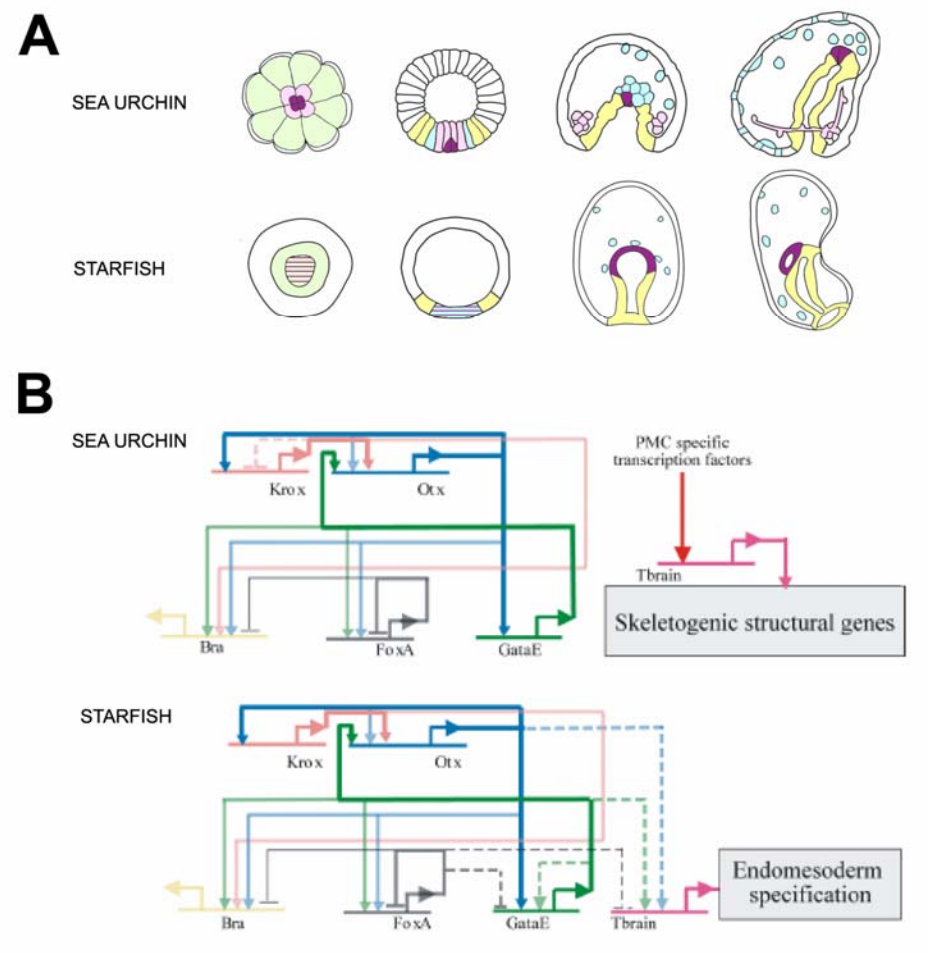


Fig. 1.5. Comparison of sea urchin and starfish gene regulatory networks. The figure compares portions of the gene regulatory networks underlying the specification of the endomesoderm in the sea urchin and the starfish embryos. **(A)** Comparison of the fate maps. Schematic diagrams of sea urchin embryos (top row) and starfish embryos [lower row, modified from Hinman *et al*. (2003)] at selected stages. Stages are (from left to right): cleavage/early blastula stage; blastula stage; gastrula stage; and early larval stage. Color coding indicates the fate of domains of cells through development: lavender indicates cells that will become skeletogenic; green indicates cells that will contribute to mesoderm and endoderm; blue indicates cells that will become mesodermal; purple indicates cells of the mesoderm that specifically will become

coelomic cells; purple stripes indicate domains that might contain a subset of cells that will contribute exclusively to coelomic cells; yellow indicates cells that will become endodermal. **(B)** Comparison of portions of the underlying gene regulatory networks. The top diagram, corresponding to the sea urchin, is extracted from Fig. 1.2. The bottom diagram, corresponding to the starfish, is from Hinman *et al*. (2003). Regulatory connections are represented as described in Fig. 1.2. In this figure dashed lines indicate a regulatory connection observed in sea urchin not present in starfish, or *vice versa*. The positive feedback loops between *krox*, *otx* and *gatae* that are present in both echinoderms are highlighted in bold.

Sea urchins and starfish have diverged for at least 500 million years (Sprinkle and Kier, 1987; Smith, 1988; Bowring and Erwin, 1998). The reinforcing loop is therefore a regulatory device that was invented at least about 500 million years ago, and that has been conserved in at least two independently evolving lineages during all this time. 500 million years represents a very long genomic divergence, in the sense that comparisons of starfish and sea urchin DNA sequences around orthologous regions do not show any conservation distinguishable from random occurrence between the *cis*-regulatory elements, even when the genes are similarly regulated (V. Hinman and E. Davidson, unpublished data). The preservation of this regulatory device suggests that the function it serves in the specification process must be essential. As we have already seen, in the sea urchin the regulatory feedback loop between *krox* and *otx* genes generates a robust regulatory structure in the endomesoderm domain, and the addition of the *gatae* gene to this feedback loop ensures and maintains the operation of many endodermal regulatory genes after the initial transient inputs have disappeared (Davidson *et al*., 2002a and P.Y. Lee and E. Davidson, unpublished data). In the starfish, *gatae* also drives the expression of many endodermal regulatory genes (Hinman *et al*., 2003), and in many other bilaterians, members of the Gata family of transcription regulatory genes are required for

gut development (Reuter, 1994; Maduro *et al*., 2002; Patient and McGhee, 2002). What makes the reinforcing loop especially useful, and hence likely to be preserved during evolution, may therefore be that it controls the installation and stabilization of the expression of the *gatae* gene in the endoderm (Hinman *et al*., 2003). Other intergenic feedback loops are used across the Bilateria to serve similar functions. For example a reinforcing feedback loop is found in the *hox* gene network that controls rhombomere specification in the mouse hindbrain (Nonchev *et al*., 1996; Barrow *et al*., 2000), in the regulatory network for tracheal placode specification in *Drosophila* (Zelzer and Shilo, 2000), and in specification of the oral ectoderm in sea urchin embryos (Amore *et al*., 2003), among others. It seems a general property of the developmental process to use feedback loops as a mechanism to achieve the progressivity of the process.

The *tbr* gene, on the other hand, is used in completely different ways in the starfish and sea urchin embryos (Fig. 1.5B). It is required for the formation of the archenteron in the starfish embryo, and its expression is under the control of endodermal regulators (Otx, Gatae) (Hinman *et al*., 2003), whereas it is involved solely in skeletogenic functions in the sea urchin embryo (Croce *et al*., 2001; Fuchikami *et al*., 2002; Oliveri *et al*., 2002 and www.its.caltech.edu/~mirsky/endomes.htm). The skeletogenic micromere lineage is a relatively recent echinoid invention (Wray and McClay, 1988; Tagawa *et al*., 2000). This suggests that in the sea urchin the skeletogenic use of *tbr* may have been coopted from an adult skeletogenic regulatory system, while an original embryonic endomesodermal regulatory element was lost (Hinman *et al*., 2003).

If indeed the larval skeletogenic lineage is the result of a cooption from the adult skeletogenic regulatory system, it represents an example of how a regulatory subroutine

can be "wired" into the specification system as the result of evolutionary change. How the intrinsic behavior of the subroutine is preserved in the new context, and how the rest of the developmental control system can cope with this change without disrupting its workability, speaks directly to the intrinsic robustness of the subroutine, and the robustness of the developmental process in general. Regulatory networks serve as the link between development and evolution. They provide a new means to address specific questions about the robustness of the developmental process, and about the preservation of aspects of the process through evolutionary time. Questions such as these can only be answered by considering evolution and development together.

CONCLUSIONS

Gene network analysis identifies the mechanisms that control and operate the program for the developmental process. This will be true for all aspects of the developmental process that are required to generate the species-specific body plan. To address some of the general and fundamental questions about the process of development, though, will require understanding evolution. Because gene regulatory networks underlie the processes of both development and evolution, unraveling their architecture in appropriately chosen species will be the key to understanding how genomes control development and how they evolve.

ACKNOWLEDGMENTS

We are very grateful to Dr. Takuya Minokawa for very useful criticism and fruitful discussions. We especially thank Dr. Paola Oliveri for providing a photomicrograph displaying the expression of *delta* gene 12 h after fertilization, and for useful discussions. We are particularly indebted to Dr. Veronica Hinman for allowing us to discuss her unpublished data regarding the comparison of sea urchin and starfish genomic sequences, and for useful criticism; to Dr. Andy Ransick for allowing us to discuss his unpublished data regarding the analysis of the *cis*-regulatory element of the *gcm* gene; and also to Pei Yun Lee for allowing us to discuss her unpublished data regarding the role of the *gatae* gene in the specification of the endoderm of the sea urchin embryo, and for helpful discussions. This research was supported by NIH Grants HD37105 and GM61005 and by NASA/Ames Grant NAG2-1587.

REFERENCES

AMORE, G., YAVROUIAN, R.G., PETERSON, K.J., RANSICK, A., McCLAY, D.R. and DAVIDSON, E.H. (2003). *Spdeadringer*, a sea urchin embryo gene required separately in skeletogenic and oral ectoderm gene regulatory networks. *Dev. Biol*., in press.

BARROW, J., STADLER, H. and CAPECCHI, M. (2000). Roles of Hoxa1 and Hoxa2 in patterning the early hindbrain of the mouse. *Development* 127: 933-944.

BOWRING, S. and ERWIN, D. (1998). A new look at evolutionary rates in deep time: Uniting paleontology and high-precision geochronology. *GSA Today* 8: 2.

BRITTEN, R.J. and DAVIDSON, E.H. (1971). Repetitive and non-repetitive DNA sequences and a speculation on the origins of evolutionary novelty. *Quart. Rev. Biol*. 46: 111-138.

CALESTANI, C., RAST, J.P. and DAVIDSON, E.H. (2003). Isolation of mesoderm specific genes in the sea urchin embryo by differential macroarray screening. *Development* 130, in press.

CAMERON, R.A., HOUGH-EVANS, B.R., BRITTEN, R.J. and DAVIDSON, E.H. (1987). Lineage and fate of each blastomere of the eight-cell sea urchin embryo. *Genes & Dev.* 1: 75-85

CAMERON, R.A., FRASER, S.E., BRITTEN, R.J. and DAVIDSON, E. H. (1991). Macromere cell fates during sea urchin development. *Development* 113: 1085-1092

CROCE, J., LHOMOND, G., LOZANO., J.-C. and GACHE, C. (2001). Ske-T, a T-box gene expressed in the skeletogenic mesenchyme lineage of the sea urchin embryo. *Mech. Dev*. 107: 159-162.

DAVIDSON, E.H. (2001). *Genomic Regulatory systems: Development and Evolution*. San Diego, Academic Press.

DAVIDSON, E.H., CAMERON, R.A. and RANSICK, A. (1998). Specification of cell fate in the sea urchin embryo: Summary and some proposed mechanisms. *Development* 125: 3269-3290.

DAVIDSON, E.H. *et al*. (2002a). A provisional regulatory gene network for specification of endomesoderm and the sea urchin embryo. *Dev. Biol*. 246: 162-190.

DAVIDSON, E.H. *et al*. (2002b). A genomic regulatory network for development. *Science* 295: 1669-1678.

ETTENSOHN, C.A., ILLIES, M.R., OLIVERI. P. and DE JONG, D.L. (2003). Alx1, a member of the Cart1/Alx3/Alx4 subfamily of Paired-class homeodomain proteins, is an essential component of the gene network controlling skeletogenic fate specification in the sea urchin embryo. *Development* 130, 2917-2928.

FUCHIKAMI, T. *et al*. (2002). T-brain homologue (HpTb) is involved in the archenteron induction signals of micromere descendant cells in the sea urchin embryo. *Development* 129: 5205-5216.

GROSS, J.M. and McCLAY, D.R. (2001). The role of Brachyury (T) during gastrulation movements in the sea urchin *Lytechinus variegates. Dev. Biol*. 239: 132-147.

GURDON, J. (1988). A community effect in animal development. *Nature* 336: 772-774.

GURDON, J., KATO, K. and LEMAIRE, P. (1993). The community effect, dorsalization and mesoderm induction. *Curr. Opin. Genet. Dev*. 3: 662-667.

HINMAN, V.F., NGUYEN, A., CAMERON, R.A. and DAVIDSON, E.H. (2003). Developmental gene regulatory network architecture across 500 MYA of echinoderm evolution. In press.

HÖRSTADIUS, S. (1939). The mechanics of sea urchin development, studied by operative methods. *Biol. Rev. Cambridge Philos. Soc*. 14: 132-179.

KUROKAWA, D., KITAJIMA, T., MITSUNAGA-NAKATSUBO, K., AMEMIYA, S., SHIMADA, H. and AKASAKA, K. (1999). EpEts, an ets-related transcription factor implicated in primary mesenchyme cell differentiation in the sea urchin embryo. *Mech. Dev*. 80: 41-52.

LOGAN, C.Y. and McCLAY, D.R. (1997). The allocation of early blastomeres to the ectoderm and endoderm is variable in the sea urchin embryo. *Development* 124: 2213-2223.

LOGAN, C.Y., MILLER, J.R., FERKOWICZ, M. and McCLAY, D.R. (1999). Nuclear beta-catenin is required to specify vegetal cell fates in the sea urchin embryo. *Development* 126: 345-357.

MADURO, M.F., LIN, R. and ROTHMAN, J.H. (2002). Dynamics of developmental switch: Recursive intracellular and intranuclear redistribution of *Caenorhabditis*

elegans POP-1 parallels Wnt-inhibited transcriptional repression. *Dev. Biol*. 248: 128-142.

McCLAY, D.R., PETERSON, R., RANGE, R., WINTER-VANN, A. and FERKOWICZ, M. (2000). A micromere induction signal is activated by beta-catenin and acts through notch to initiate specification of secondary mesenchyme cells in the sea urchin embryo. *Development* 127: 5113-5122.

NONCHEV, S. *et al*. (1996). Segmental expression of Hoxa-2 in the hindbrain is directly regulated by Krox-20. *Development* 122: 543-554.

OLIVERI, P., CARRICK, D.M. and DAVIDSON, E.H. (2002). A regulatory gene network that directs micromere specification in the sea urchin embryo. *Dev. Biol*. 246: 209-228.

OLIVERI, P., DAVIDSON, E.H. and McCLAY, D.R. (2003). Activation of *pmar1* controls specification of micromeres in the sea urchin embryo. *Dev. Biol*. 258: 32-43.

PATIENT, R.K. and McGHEE, J.D. (2002). The GATA family (vertebrates and invertebrates). *Curr. Opin. Genet. Dev*. 12: 416-422.

RANSICK, A. and DAVIDSON, E.H. (1993). A complete second gut induced by transplanted micromeres in the sea urchin embryo. *Science* 259: 1134-1138.

RANSICK, A. and DAVIDSON, E.H. (1995). Micromeres are required for normal vegetal plate specification in sea urchin embryos. *Development* 121: 3215-3222.

RANSICK, A. and DAVIDSON, E.H. (1998). Late specification of veg1 lineages to endodermal fate in the sea urchin embryo. *Dev. Biol*. 195: 38-48.

RANSICK, A., RAST, J.P., MINOKAWA, T., CALESTANI, C. and DAVIDSON, E.H. (2002). New early zygotic regulators expressed in endomesoderm of sea urchin embryos discovered by differential array hybridization. *Dev. Biol*. 246: 132-147.

RAST, J.P., AMORE, G., CALESTANI, C., LIVI, C.B., RANSICK, A. and DAVIDSON, E.H. (2000). Recovery of developmentally defined gene sets from high-density cDNA macroarrays. *Dev. Biol*. 228: 270-286.

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.

REUTER, R. (1994). The gene *serpent* has homeotic properties and specifies endoderm versus ectoderm within the *Drosophila* gut. *Development* 120: 1123-1135.

RUFFINS, S.W. and ETTENSOHN, C.A. (1993). A clonal analysis of secondary mesenchyme cell fates in the sea urchin embryo. *Dev. Biol*. 160: 285-288.

RUFFINS, S.W. and ETTENSOHN, C.A. (1996). A fate map of the vegetal plate of the sea urchin (*Lytechinus variegatus*) mesenchyme blastula. *Development* 122: 253-263.

SHERWOOD, D. and McCLAY, D.R. (1999). LvNotch signaling mediates secondary mesenchyme specification in the sea urchin embryo. *Development* 126: 1703-1713.

SMITH, A. (1988). *Echinoderm Phylogeny and Evolutionary Biology*. Oxford, Clarendon Press.

SPRINKLE, J. and KIER, P. (1987). *Fossil Invertebrates*. Cambridge, MA, Blackwell Science.

SWEET, H., HODOR, P. and ETTENSOHN, C. (1999). The role of micromere signaling in Notch activation and mesoderm specification during sea urchin embryogenesis. *Development* 126: 5255-5265.

SWEET, H.C., GEHRING, M. and ETTENSOHN, C.A. (2002). LvDelta is a mesoderm-inducing signal in the sea urchin embryo and can endow blastomeres with organizer-like properties. *Development* 129: 1945-1955.

TAGAWA, K., HUMPHREYS, T. and SATOH, N. (2000). T-Brain expression in the apical organ of hemichordate tornaria larvae suggests its evolutionary link to the vertebrate forebrain. *J. Exp. Zool*. 288: 23-31.

WRAY, G. and McCLAY, D.R. (1988). The origin of spicule-forming cells in a 'primitive' sea urchin (*Eucidaris tribuloides*) which appears to lack primary mesenchyme cells. *Development* 103: 305-315.

YUH, C.-H., BOLOURI, H. and DAVIDSON, E.H. (2001). *cis*-Regulatory logic in the *endo16* gene: Switching from a specification to a differentiation mode of control. *Development* 128: 617-629.

ZELZER, E. and SHILO, B.-Z. (2000). Interaction between the bHLH-PAS protein Trachealess and the POU-domain protein Drifter, specifies tracheal cell fates. *Mech. Dev*. 91: 163-173.

CHAPTER 2

R11: A *cis*-Regulatory Node of the Sea Urchin Embryo Gene Network that Controls Early Expression of *SpDelta* in Micromeres

Roger Revilla-i-Domingo, Takuya Minokawa, and Eric H. Davidson

Published in *Developmental Biology* **274:** 438-451 (2004)

ABSTRACT

A gene regulatory network (GRN) controls the process by which the endomesoderm of the sea urchin embryo is specified. In this GRN the program of gene expression unique to the skeletogenic micromere lineage is set in train by activation of the *pmar1* gene. Through a double repression system this gene is responsible for localization of expression of downstream regulatory and signaling genes to the cells of this lineage. One of these genes, *delta*, encodes a Notch ligand, and its expression in the right place and time is crucial to the specification of the endomesoderm. Here we report a *cis*-regulatory element, R11, that is responsible for localizing the expression of *delta* by means of its response to the *pmar1* repression system. R11 was identified as an evolutionarily conserved genomic sequence located about 13 kb downstream of the last exon of the *delta* gene. We demonstrate here that this *cis*-regulatory element is able to drive the expression of a reporter gene in the same cells and at the same time that the

endogenous *delta* gene is expressed, and that temporally, spatially, and quantitatively it responds to the *pmar1* repression system just as predicted for the *delta* gene in the endomesoderm GRN. This work illustrates the application of *cis*-regulatory analysis to the validation of predictions of the GRN model. In addition, we introduce new methodological tools for quantitative measurement of the output of expression constructs, that promise to be of general value for *cis*-regulatory analysis in sea urchin embryos.

Keywords: cis-Regulatory element; Gene regulatory network; *delta*; Endomesoderm specification; Sea urchin

INTRODUCTION

In the process of development a network of gene regulatory interactions underlies each specification event (Davidson et al., 2002a). These interactions occur at genomic *cis*-regulatory elements which respond to the set of inputs (i.e., transcription factors) presented in each cell, and which control the expression of each gene, in each domain of the embryo. The properties of the set of all relevant *cis*-regulatory elements ultimately determine the architecture of the gene regulatory network (GRN) that underlies embryonic specification.

An explicit model of the GRN directing the specification of the distinct endodermal and mesodermal cell types of the sea urchin embryo has been published (Davidson et al., 2002a,b; reviewed by Oliveri and Davidson, 2004). This model predicts inputs to the *cis*-regulatory elements of the many genes involved, based on an extensive

experimental perturbation analysis. The full explanatory power of the model, however, can only be achieved when we have in our hands the key fragments of genomic DNA that execute the *cis*-regulatory interactions predicted by the model. These *cis*-regulatory elements will serve to provide the ultimate tests for the correctness of the model. Also, their identification will eventually make possible experiments in which chosen parts of the network of *cis*-regulatory interactions can be deliberately modified, thereby highlighting the roles of specific portions of the circuitry.

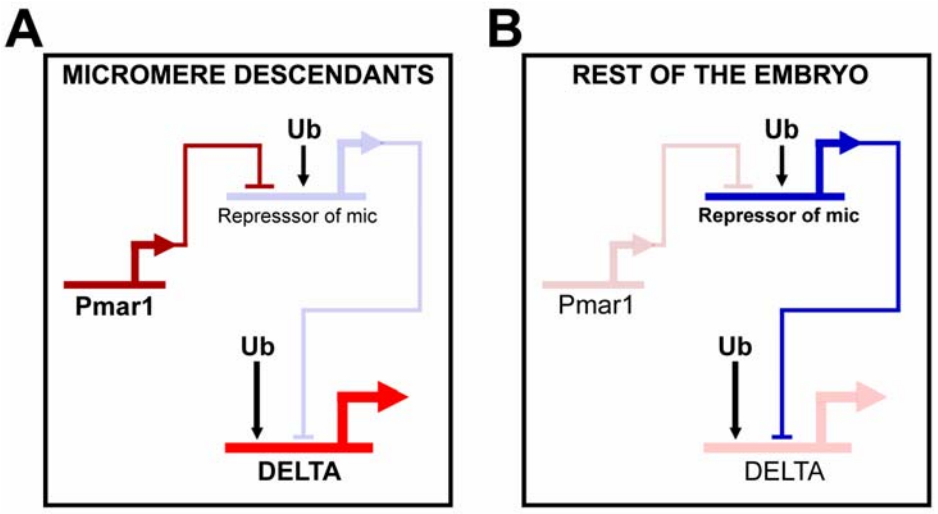


Fig. 2.1. Network interactions predicted to be responsible for expression of *delta* in micromere lineage cells (modified from Davidson et al. (2002b), and Oliveri et al., (2002)). Thick horizontal lines from which bent arrows extend represent *cis*-regulatory elements responsible for expression of the genes named beneath the lines. *cis*-Regulatory elements represented in dimmed color indicate that the gene they control is silent. *cis*-Regulatory elements represented in full color indicate that the gene they control is active. The arrows and barred lines indicate the inferred normal function of the input (activation or repression). (A) In the micromere lineage the *pmar1* gene is active, and it represses a gene encoding a yet unknown, otherwise globally expressed repressor (*repressor of mic*), resulting in the activation of *delta* exclusively in these cells. (B) In the rest of the embryo, *delta* is kept silent by *repressor of mic*. Ub, ubiquitous activator.

Oliveri et al. (2002) demonstrated that the program of gene expression specific to the skeletogenic primary mesenchyme cell (PMC) lineage is set in train by the *pmar1* gene, acting through a double repression system. Two developmental functions that are specific to the PMC lineage are set in action as a direct consequence of the operation of this repression system. The first of these is the emission of the Delta signal, which serves as a spatial cue that triggers the specification of mesodermal cell types from the common endomesodermal progenitor cells. Expression of the ligand Delta between 7^{th} and 9^{th} cleavages in the micromere lineage, the precursors of the PMCs, activates a Notch receptor in the adjacent endomesodermal (veg_2) cells, and this is required for normal specification of mesodermal fate in these cells (Sweet et al., 1999; McClay et al., 2000; Sweet et al., 2002). Thus the cells of the veg_2 territory immediately adjacent to the micromere descendants are specified as mesoderm; the rest of the cells of the veg_2 territory will become endoderm. The GRN model predicts that expression of *delta* in the micromere lineage depends on activating factors that are ubiquitously present (Fig. 2.1). The normally exclusive expression of this gene in the micromere lineage depends on a repressor ("Repressor of mic" in Fig. 2.1) that is also active everywhere, except in this lineage. There the *pmar1* gene product in turn represses the gene encoding the otherwise ubiquitous repressor. The second developmental function executed specifically by the cells of the PMC lineage is to give rise to the skeletogenic mesenchyme of the postgastrular embryo. The regulatory genes *tbr*, *alx1* and *ets1* are all known to contribute to the activation of a number of biomineralization genes that are responsible for the skeletogenic differentiation of the micromere lineage (Kurokawa et al., 1999; Fuchikami et al., 2002; Oliveri et al, 2002; Ettensohn et al., 2003). The GRN model predicts that

these three regulatory genes are expressed specifically in the micromere descendants due to *cis*-regulatory interactions that include the same mechanism used to localize the expression of *delta*, i.e., the *pmar1* repression system summarized in Fig. 2.1. In particular, this prediction rules out the possibility that any of these three genes is upstream of *delta* or of each other, in agreement with the fact that none of these three genes affects the expression of *delta* or of each other (Oliveri et al., 2002).

The goal of the present study was to test the GRN model by identifying a fragment of genomic DNA from the *Strongylocentrotus purpuratus delta* gene, here referred to as *delta*, that executes the predicted *cis*-regulatory interactions. We first set ourselves to recover the *cis*-regulatory element that drives the expression of *delta* in the micromere descendants at the right time. We were then able to ask whether it responds to the *pmar1* repression system as in the GRN model prediction.

MATERIALS AND METHODS

Isolation and analysis of BAC clones containing Spdelta *and* Lvdelta *genes*

A BAC clone, named 046A16, containing the *delta* gene had been obtained earlier. BAC clones, named 020B17 and 071J09, containing the *Lvdelta* gene were recovered by cross-species hybridization of a *Lytechinus variegatus* BAC genomic library (Cameron et al., 2000). The partial sequence of a *delta* cDNA clone, obtained by Zhu et al. (2001), was used to design the probe for the cross-species hybridization. This probe was obtained by PCR amplification from the cDNA clone (left primer: 5'-acaacagctgcagggacatt-3'; right primer: 5'-acatggtccgacacactgat-3').

BAC clones of both species were sequenced by DOE's Joint Genome Institute. These sequences are available at www.sugp.caltech.edu (under Resources/Annotation). The exons of the *delta* gene in *Strongylocentrotus purpuratus* and *L. variegatus* BAC clones were identified using the sequence of both a partial *S. purpuratus* and a complete *L. variegatus* cDNA clones (Sweet et al., 2002). The sequences were annotated using the SUGAR software package (Brown et al., 2002). This software was used to identify coding sequences of genes neighboring *delta* in the BAC clones.

Comparison of the genomic sequence around the delta *genes of* S. purpuratus *and* L. variegatus

The FamilyRelations software package (Brown et al., 2002) was used to compare the BAC sequences of *S. purpuratus* and *L. variegatus*. Window sizes used in the comparison ranged from 10 bp to 200 bp. The pairwise view of the software was used to identify conserved regions. The Dot Plot view was used in some cases to identify the boundaries of the conserved regions found.

Preparation of reporter constructs

Selected regions R1 through R12 of the BAC clone 046A16 of *S. purpuratus* were amplified by means of PCR. The relevant sequences were amplified from the BAC clone by using the "Expand High Fidelity PCR System" (Roche). Primers used for the amplifications were equipped with restriction digest anchors. The sequence of the primers used for the amplification of region R11 were: Left primer – 5' aagtaggtaccatgccaacatgaagatgc 3'; Right primer – 5' taagtgagctccacgtctcgtctcgtttaat 3'.

Reporter constructs R1-GFP through R12-GFP were prepared by cloning the amplified regions R1 thru R12, respectively, into the multiple cloning site of the universal *S. purpuratus* expression vector EpGFPII (Cameron et al., 2004). That the correct sequences had been cloned was confirmed by restriction mapping. The vector EpGFPII contains the region around the start of transcription of the *endo16* gene (from -117 to +20). The activity of this basal promoter element has been described in detail elsewhere (Yuh and Davidson, 1996; Yuh et al., 1996; 1998). The EpGFPII expression vector also contains the coding sequence of the GFP protein. All reporter constructs were linearized by restriction digestion upstream of the cloned fragment.

Animals and microinjection of reporter constructs

Microinjection solutions were prepared containing 350-1000 molecules/pl of the reporter construct to be microinjected, together with 4- to 9-fold molar excess of HindIII-digested carrier sea urchin DNA and 0.12 M KCl (Franks et al., 1990).

Gametes from *S. purpuratus* maintained in our year-round culture system were obtained and microinjected as described by Rast (2000). This protocol is essentially based on the original protocol by McMahon et al. (1985) with significant modifications. The volume of solution microinjected into the embryos was estimated by observing the size of the disturbance produced in the egg cytoplasm. We aimed at microinjecting a volume of 2 pl or 5 pl of solution depending on the experiment. Experiments were carried out in which nominally 700, 1200, 2500 or 4000 molecules of the reporter construct were microinjected into the eggs.

Microinjected embryos were reared at 14 ºC to various developmental stages. Some embryos were reared that had not been microinjected, and that had been obtained from the same female and prepared in a similar way as the microinjected embryos. These uninjected embryos were used to control for possible developmental anomalies caused by microinjection.

Simultaneous microinjection of R11-GFP reporter construct and* pmar1 *mRNA

The preparation of *pmar1* mRNA by plasmid transcription was performed as described (Oliveri et al., 2002). Microinjection solutions were prepared containing 400 molecules/pl of R11-GFP reporter construct and 22 ng/µl of *pmar1* mRNA, together with 7-fold molar excess of HindIII-digested carrier sea urchin DNA and 0.12 M KCl. Nuclease-free water was used to prepare the microinjection solutions. ~5 pl of the microinjection solution was microinjected into the embryos using the same method as described above for the microinjection of reporter constructs.

Determination of GFP expression in microinjected embryos

Microinjected embryos were visualized on an epifluorescence Axioskop 2 Plus microscope (Zeiss, Hallbergmoos, Germany), equipped with the recording device AxioCam MRm (Zeiss). Expression of GFP in each embryo was determined by the presence of cells fluorescing at a level significantly higher than background. For each GFP-expressing embryo, the location of the GFP-expressing cells was determined according to the morphology of the embryo.

Images were collected and processed in Adobe Photoshop.

Quantification of R11-GFP DNA in microinjected embryos

The Sigma "GenElute Mammalian Total RNA Miniprep Kit" is designed to isolate total RNA. Along with RNA, however, small amounts of DNA are also recovered. This was exploited to quantify the R11-GFP DNA in microinjected embryos in which the GFP expression level was also to be quantified. RNA and DNA were isolated, as described in the manufacturer's manual, from samples of 100-150 embryos that had been microinjected with the R11-GFP reporter construct and/or *pmar1* mRNA. Samples were not digested with DNase I, so that the extracted DNA remained in the samples for quantification. Quantitative PCR (QPCR) was conducted using primer sets designed to amplify products of 125 to 150 bp of the coding sequence of GFP (GFP primer set) and the coding sequence of the *foxb* gene (*foxb* primer set). For sequences of primers see http://sugp.caltech.edu/resources/methods/q-pcr.psp. Amplification reactions were analyzed on an ABI 5700 sequence detection system using SYBR Green chemistry (PE Biosystems, Foster City, CA). Reactions were run in triplicate with samples from two embryos. Thermal cycling parameters were 95 ºC for 30 s, 60 ºC for 1 min, 40 cycles. The number of molecules of R11-GFP DNA per embryo was estimated by using the *foxb* gene as an internal standard; we know that there are two copies of the *foxb* gene per cell (Luke et al., 1997).

Quantification of GFP,* delta *and* pmar1 *mRNA in microinjected embryos

Samples for which the amount of R11-GFP construct DNA had been measured were then treated with DNase I using the DNA-free kit (Ambion, Austin, TX), as described in the manufacturer's manual, in order to remove all existing DNA. QPCR was conducted as described above to confirm that no DNA remained in the samples.

cDNA was prepared from the samples by reverse transcription-PCR (RT-PCR). The TaqMan Reverse Transcription Reagents Kit (Applied Biosystems, Foster City, CA) was used for this purpose. 38.5 µl of the RNA preparation was used in a 100 µl reverse transcription reaction (note, though, that in more recent experiments 30 µl were used instead of 38.5 µl, and this seems to improve the efficiency of the RT-PCR).

QPCR was conducted as described in the previous section using primer sets designed to amplify products of 125 to 150 bp of the cDNA generated from *18S* ribosomal RNA, GFP mRNA, *ubiquitin* mRNA, *Spz12*-1 mRNA, *delta* mRNA and *pmar1* mRNA (for primer sequences, see http://sugp.caltech.edu/resources/methods/q-pcr.psp). Amplification reactions were analyzed as described above. Reactions were run in triplicate with cDNA from 4-6 embryos. For all QPCR experiments, the data from each cDNA sample were normalized against the *ubiquitin* mRNA and/or *18s* rRNA levels, which are known to remain relatively constant during the developmental stages used (Nemer et al., 1991; Ransick et al., 2002). Absolute quantification of the number of *ubiquitin* and/or *18s* rRNA transcripts in uninjected embryos was obtained by using *Spz12-1* as an internal standard. The number of *Spz12-1* transcripts in embryos of the relevant stages had been measured earlier by RNA titration (Wang et al., 1995). The number of *ubiquitin* and/or *18s* rRNA transcripts was then used for absolute

quantification of the number of GFP and *delta* mRNA transcripts in microinjected embryos.

RESULTS

Spatial and temporal expression pattern of delta *during endomesoderm specification*

Sweet et al. (2002) showed that in *Lytechinus variegatus* the *delta* gene is expressed starting at around 7th cleavage in the micromere descendants. As the PMCs ingress into the blastocoel, the expression of *Lvdelta* in the micromere descendants disappears, and expression starts in the presumptive secondary mesenchyme cells (SMCs). Whole mount *in situ* hybridization (WMISH) experiments carried out by Oliveri et al. (2002) indicated a similar pattern of expression in *S. purpuratus.* The *delta* gene is expressed in the micromeres starting no later than 8 h after fertilization, and the transcripts remain in their descendants at 18 h. We carried out further WMISH experiments which show that in *S. purpuratus, delta* transcripts remain present in the micromere lineage until these cells ingress into the blastocoel at 20 h (data not shown). At this time, expression of *delta* ceases in the micromere lineage, and as reported for the *Lvdelta* gene (Sweet et al., 2002), expression is then activated in the presumptive SMCs (data not shown). By 24 h, expression of *delta* is seen only in the presumptive SMCs.

To further refine the time at which *delta* expression starts, we measured the levels of *delta* mRNA at several stages of development by means of QPCR. As shown in Fig. 2.2A these experiments indicate that *delta* is first expressed between 6 and 8 h after fertilization. Our objective was then to identify the genomic element(s) that are

responsible for the specific expression of *delta* in the micromere lineage, from 6-8 h to 20 h after fertilization.

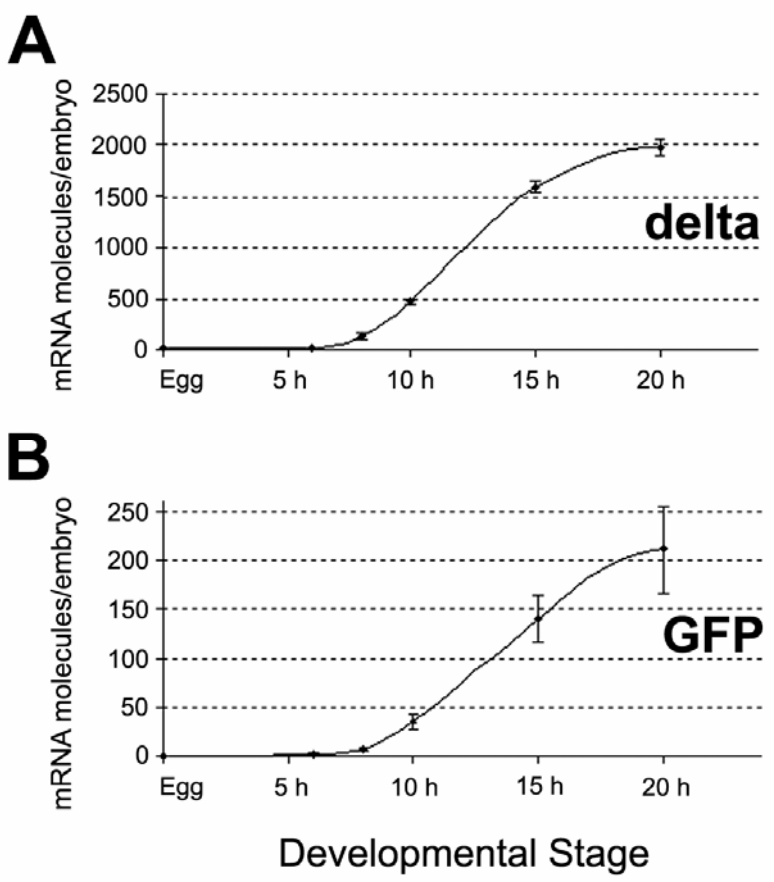


Fig. 2.2. Temporal expression pattern of endogenous *delta* gene compared to the temporal expression pattern of GFP mRNA from the R11-GFP reporter construct. (A) QPCR data indicating levels of *delta* mRNA at different developmental stages. Experimental data are indicated by dots. The line joining these dots is inferred. The error bars represent one standard deviation. Note: For the sake of accuracy in the comparison, the levels of *delta* mRNA were measured in the same sample of embryos as in (B). Although these embryos had been injected with R11-GFP, measurement of the levels of *delta* mRNA in uninjected embryos of the same batch showed that injection of R11-GFP has no effect in the levels of *delta* mRNA. (B) QPCR data indicating levels of GFP mRNA in the same samples of embryos as in (A). Similar temporal expression patterns were obtained using embryos from three different females. The absolute levels of GFP mRNA vary extensively between different experiments, depending on the number

of R11-GFP DNA molecules incorporated in the genome of the microinjected embryos in each case. The timing at which GFP mRNA expression starts, nevertheless, is accurately reproduced in each experiment.

Genomic sequence surrounding the delta *gene*

The sequence in the vicinity of the *delta* gene was annotated in order to determine the regions where its *cis*-regulatory system might likely be found. A BAC clone containing the *delta* gene was sequenced, and the positions of the *delta* exons in this clone are indicated in Fig. 2.3A. This BAC clone contains the complete 2394 bp of Delta coding sequence, divided into 11 exons, which together extend over almost 15 kb of the genome. Application of the SUGAR annotation package (Brown et al., 2002), revealed the presence of the coding sequence of an unnamed gene about 37 kb upstream of the *delta* start of translation, and another gene is predicted about 33 kb downstream of the termination of the *delta* coding sequence (Fig. 2.3A). Therefore the *cis*-regulatory regions that control the expression of *delta* are likely to reside within the 85 kb of genomic sequence between the genes identified upstream and downstream of *delta*.

Identification of conserved genomic sequences as putative cis*-regulatory elements*

We compared the relevant genomic region of *S. purpuratus* with the orthologous region of the *L. variegatus* genome in order to identify conserved sequence patches. *S. purpuratus* and *L. variegatus* diverged about 50 million years ago, and this distance has been shown to be useful for the identification of putative *cis*-regulatory elements, which are recognized as significantly conserved sequence elements (Yuh et al., 2002; 2004). To this end *L. variegatus* BAC clones containing the coding sequence of *Lvdelta* were

obtained by cross-species hybridization of a *L. variegatus* BAC genomic library, and sequenced. Analysis of the *S. purpuratus* and *L. variegatus* genomic sequences with the SUGAR software revealed 70 kb of *L. variegatus* BAC sequence which is orthologous to the *S. purpuratus* genomic sequence around *delta*. These 70 kb of genomic sequence extend, in the *S. purpuratus* genome, from the next gene upstream of *delta* to about 18 kb downstream of the termination of the *delta* coding sequence (Fig. 2.3A).

The 70 kb of orthologous genomic sequence was scanned computationally for short conserved sequence regions using the FamilyRelations software package (Brown et al., 2002). This tool allows for the detection of sequence similarities above a chosen criterion within sliding windows set at chosen window sizes. Figure 2.3A shows a pairwise view of this comparison. In this view every red line connecting the *S. purpuratus* and *L. variegatus* sequences indicates an interspecific sequence similarity at the chosen criterion; in the case of Fig. 2.3A it represents the presence of a sequence stretch of 20 bp that is identical in the two species. Given the stringency of the criterion chosen, only regions with very high similarity are detected.

The comparison of the two orthologous sequences was also visualized using a dot plot view. Figure 2.3B shows a small portion of such a view. Each dot indicates the presence of a sequence of 10 bp in which at least 9 bp are identical in the two sequences. The low stringency of the criterion used in Fig. 2.3B results in a high level of noise due to random matches. These random matches appear as isolated dots, while sequence similarities corresponding to "true" conservation can be distinguished by their diagonal continuity. The dot plot view offers an important advantage with respect to the pairwise view, in that it better shows the structure of the sequence similarities. Thus we see that

most of the conserved stretches in Fig. 2.3A consist of isolated blocks of very well conserved sequence, with sharp boundaries, surrounded by very poorly conserved sequence. Fig. 2.3B shows one of these blocks.

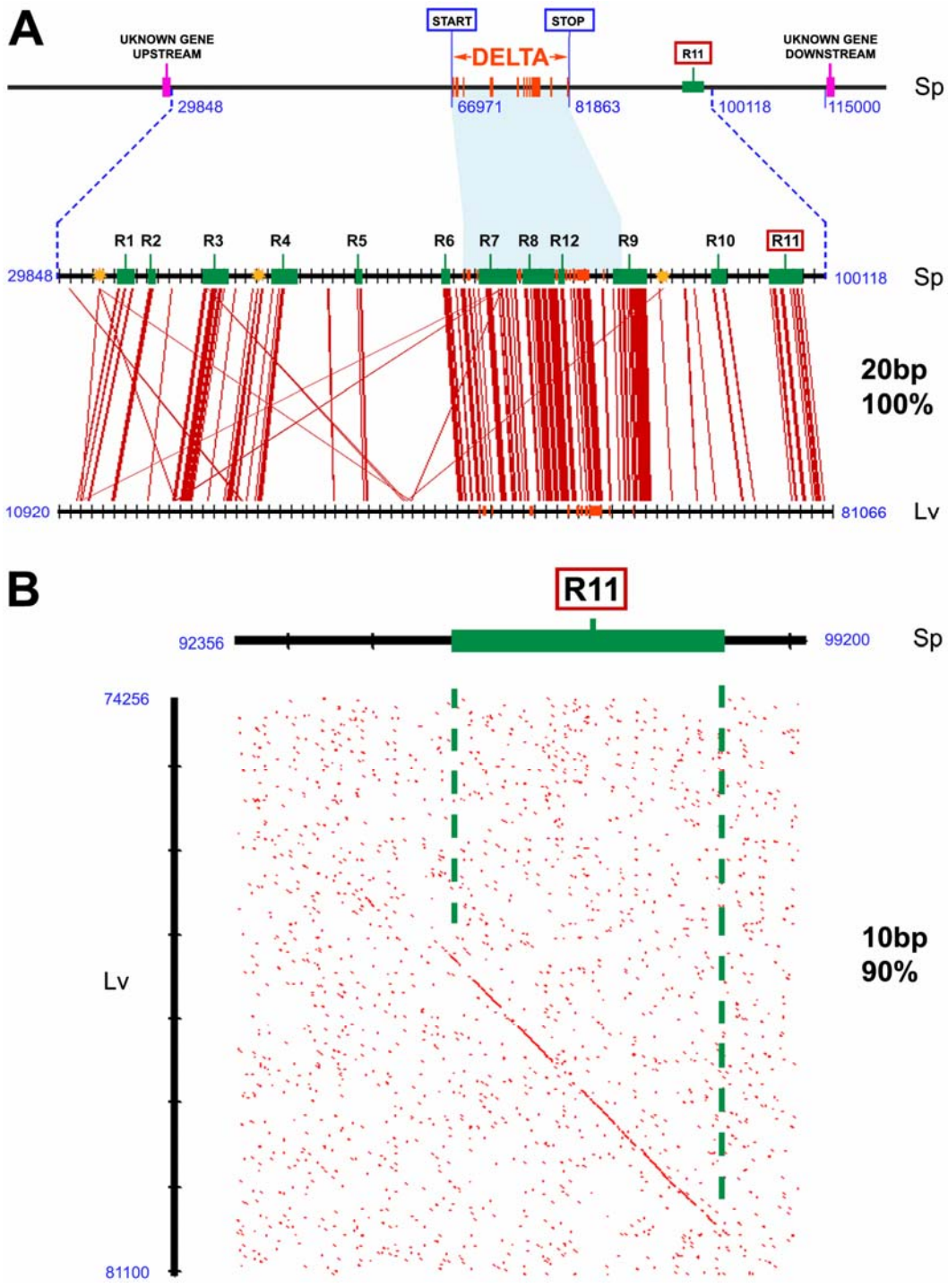


Fig 2.3. Comparative interspecific sequence analysis. (A) Map of the *S. purpuratus* genomic sequence around the *delta* gene (top), and pairwise view of a FamilyRelations comparison of *S. purpuratus* (*Sp*) and *L. variegatus* (*Lv*) orthologous genomic sequences around the *delta* gene. Horizontal black lines represent these BAC sequences. Coordinate positions in the respective BAC clones are indicated. Pink blocks indicate the position of other genes immediately upstream and downstream of *delta* in the *S. purpuratus* genome. Orange blocks indicate the positions of the coding sequence of *delta*, as obtained from sequenced cDNA clones and by comparison to the coding sequence of *Lvdelta*. START indicates start of translation, STOP, the coding sequence termination. The two blue dashed lines indicate the limits of the *S. purpuratus* genomic sequence that was compared to the orthologous *L. variegatus* genomic sequence. The shaded area indicates the region of the genome of *S. purpuratus* between the start of translation and the coding sequence termination. Each tic on the sequences demarcates 1 kb from the previous tic. The red lines connecting the two BAC sequences indicate interspecific sequence similarities, here consisting of 100% identity for a sliding window of 20 bp. Yellow stars indicate sequence similarities that contain simple sequences, e.g., microsatellites. Numbered green boxes indicate the sequence regions that were selected to be tested experimentally. (B) Dot Plot view of part of the FamilyRelations comparison in (A) but using a different criterion. In this case each dot indicates interspecific similarities, consisting of 90% identity in the sequence of the two species, for a sliding window of 10 bp. Here the *S. purpuratus* sequence is on the horizontal axis, and the *L. variegatus* sequence is on the vertical axis.

Conserved blocks with significant similarity were chosen and analyzed in detail using the Mapping Closup function of FamilyRelations. Regions consisting of simple sequence (e.g., microsatellites; yellow stars in Fig. 2.3A), regions consisting of coding sequence (orange blocks in Fig. 2.3A), and conserved regions shorter than 100 bp were excluded from further analysis. The remaining conserved patches were considered putative *cis*-regulatory elements of the *delta* gene. A total of 12 such regions, named R1 through R12 (green blocks in Fig. 2.3A), were selected for experimental test of *cis*-regulatory function during the relevant developmental stages.

The R11 DNA fragment accurately generates the early expression pattern of the* delta *gene

To test the *cis*-regulatory function of the selected conserved regions R1-R12, we prepared constructs R1-GFP-R12-GFP. Each construct was microinjected into embryos, and expression of GFP was monitored at several stages between fertilization and mesenchyme blastula stage. In the present report we focus exclusively on region R11. As the following work shows R11 generates the early expression pattern of *delta*. The *cis*-regulatory activities of the remaining conserved regions, and the overall organization of the *delta* gene, will be discussed elsewhere, since while some of these constructs are active they do not generate the phase of expression we are interested in the present report.

Table 2.1 ("R11" column) indicates the locations where GFP expression was observed at three different stages of development, in embryos that had been microinjected with the R11-GFP reporter construct. Images of some representative embryos are shown in Figs. 2.4 (A-F). In interpreting these data, we have to bear in mind two technical points: first, due to the time it takes for the GFP to be translated and for the chromophore to form, there is a delay of about 4 h from the time the mRNA accumulates to when fluorescence becomes detectable; second, that exogenous DNA is incorporated in mosaic fashion in microinjected sea urchin embryos. Within minutes after injection into the egg cytoplasm linear DNA molecules are ligated together to form one or a few very large, end-to-end concatenates (McMahon et al., 1985). Then, early in cleavage, an exogenous DNA concatenate is incorporated randomly into the genome of usually one blastomere (Flytzanis et al., 1985; Hough-Evans et al., 1988; Livant et al, 1991). Once incorporated, the exogenous DNA replicates together with the endogenous DNA, and is

inherited by the progeny of the host cells (Flytzanis et al., 1987; Franks et al., 1988; Livant et al., 1991). As a consequence, each of the microinjected embryos will have one or a few clones of cells that contain exogenous DNA, and that therefore have the possibility to express the reporter gene.

Table 2.1. Expression of GFP in embryos microinjected with the R11-GFP reporter construct and in embryos simultaneously microinjected with *pmar1* mRNA

Stage	Injection	R11			R11+pmar1		
			% TOTAL	% Expr		% TOTAL	% Expr
Blastula (15h-17h)	TOTAL	231			117		
	Expressing	67	29%		54	46%	
Early Mesenchyme Blastula (20h-22h)	TOTAL	515			107		
	Expressing	182	35%		93	87%	
	Ingressing PMCs	179	35%	98%	37	35%	40%
	Blastula Wall Cells	6	1%	3%	73	68%	78%
Mesenchyme Blastula (24h-26h)	TOTAL	897			125		
	Expressing	344	38%		120	96%	
	PMCs	316	35%	92%			
	Vegetal Plate	29	3%	8%			
	Presumtive Ectoderm	20	2%	6%			

'% TOTAL' means % of embryos respect to the 'TOTAL' number of embryos observed; '% Expr' means % of embryos respect to the number of 'Expressing' embryos.

Notes:

1) At early mesenchyme blastula stage, cells were scored as ingressing PMCs if they were inside the blastocoel, ingressing into the blastocoel, or immediately next to the cells ingressing into the blastocoel.

2) Values shown have been obtained by summing over all the experiments carried out in which no anomalies were observed in the development of the microinjected embryos.

3) Each value in the table derives from experiments carried out using eggs from at least three different females. In some cases, eggs from as many as 20 different females were used.

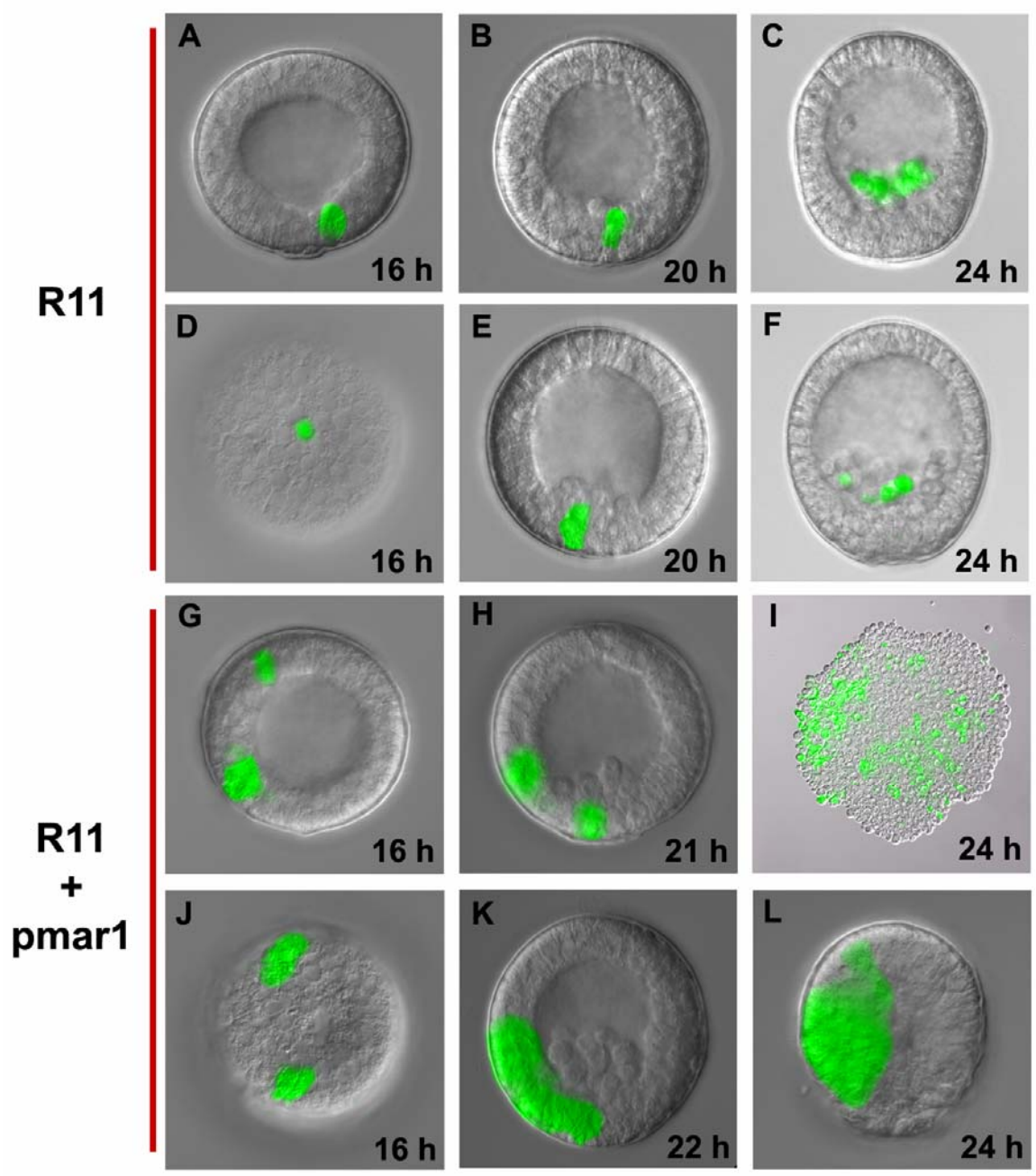


Fig. 2.4. Spatial GFP expression pattern driven by R11-GFP reporter construct. Fluorescence images superimposed on bright field images of embryos microinjected with R11-GFP reporter construct, and cultured to the developmental stage indicated at the lower right corner of each image. (A-F) Embryos microinjected with R11-GFP reporter construct only. (G-L) Embryos simultaneously microinjected with R11-GFP reporter construct and *pmar1*mRNA. Note that the embryo in I was slightly squeezed to show all the expressing cells in the same focal plane.

As shown in Table 2.1, GFP fluorescence was observed at blastula stage (15-17 h after fertilization) in a significant number of embryos (29%), indicating that R11 had driven the expression of GFP mRNA at least as early as 11 h. At the 15-17 h blastula stage, it is often impossible to identify distinct cell types in the embryos by morphological observation alone. The small micromeres can sometimes be distinguished, however, thereby indicating the vegetal pole of the embryo. Those embryos expressing GFP in which this identification was possible, showed that GFP fluorescence is always localized to cells immediately next to the small micromeres (Fig. 2.4A). Whenever expression of GFP was observed, it was confined to a small region of the embryo (Figs. 2.4A and D).

At early mesenchyme blastula stage (20-22 h after fertilization), as the micromere descendants begin their ingression into the blastocoel, the embryos expressed GFP either in the ingressing cells (Fig. 2.4B), or in underlying cells, which from their position appear about to ingress (Fig. 2.4E; Table 2.1).

At later mesenchyme blastula stage (24-26 h after fertilization), when all cells of the micromere lineage have completed ingression into the blastocoel, the PMCs, vegetal plate cells, and ectodermal cells of the embryo can be clearly distinguished. At this stage, expression of GFP was seen almost exclusively in the PMCs (Table 2.1, Fig. 2.4C and F).

These results indicated that the R11-GFP construct drives expression of GFP in cells of the micromere lineage beginning sometime before 11 h postfertilization. No cells other than the micromere descendants accumulate significant levels of GFP mRNA, even transiently. Arnone and Davidson (1997) showed that GFP is very stable in these embryos, and therefore the fluorescence seen at any given stage of development is the

sum of all prior episodes of expression. Had GFP transcripts been transiently accumulated to significant levels in cells other than those of the micromere lineage any time before 20 h postfertilization, we would have seen fluorescence in the descendants of those cells at mesenchyme blastula stage. To check for this, embryos expressing during earlier blastula stages were kept alive and individually monitored for GFP expression until they reached mesenchyme blastula stage. All these embryos expressed in PMCs at mesenchyme blastula stage (data not shown). Importantly in no case did expression disappear between these two stages, demonstrating that indeed GFP fluorescence is stable and not transient, and most importantly, that the only cells showing GFP fluorescence throughout the blastula stage are precursors of the PMCs.

Hough-Evans et al. (1988) and Livant et al. (1991) showed that incorporation of exogenous DNA happens most often at the 3rd or 4th cleavage stages. According to this we would expect from the lineage map (Davidson, 1986; Cameron et al., 1987) that in only 35 to 40% of the microinjected embryos would exogenous DNA be incorporated in the cells of the micromere lineage. Consistent with this, previous *cis*-regulatory studies on a gene encoding a biomineralization protein specific to PMCs yielded exactly this frequency of expressing embryos (Makabe et al., 1995). Similarly, Table 2.1 shows that the fraction of embryos expressing R11-GFP between 20 and 26 h was 35-38%.

It remained to be seen whether the developmental time course of GFP mRNA expression driven by R11 accurately mimics the temporal expression pattern of *delta*. To resolve this we compared the levels of GFP mRNA to those measured for *delta* at several stages of development, in embryos that had been microinjected with the R11 construct (Fig. 2.2). These data show that transcription of GFP mRNA begins between 6 h and 8 h

after fertilization, and continues to increase up to at least 20 h after fertilization (Fig. 2.2B). The time course almost exactly resembles the temporal expression pattern of the endogenous *delta* gene from fertilization to 20 h postfertilization. Temporally as well as spatially, the expression pattern driven by R11 accurately recapitulates the early expression pattern of *delta*.

The R11 expression pattern depends on operation of the* pmar1 *repression system

If the sequence element R11 is responsible for localizing the expression of *delta* to the micromere descendants, we would expect that it should contain binding sites for those transcription factors that control the expression of *delta*. Therefore, if the predictions of the network model are correct, we would expect that R11 should contain binding sites for activating factors that are ubiquitously present in the embryo, and that it should respond to a repressor, expression of which is prevented in micromere descendants by the *pmar1* gene product (Fig. 2.1; Oliveri et al., 2002). Thus we would expect that ectopic expression of *pmar1* in cells other than micromere descendants should result in R11-driven expression of GFP in those cells; global expression of *pmar1* should result in GFP expression everywhere.

Global expression can be effected by microinjection of *pmar1* mRNA into fertilized eggs (Oliveri et al., 2002). To examine the effect on the expression pattern generated by R11, we analyzed GFP expression in embryos that had been microinjected simultaneously with the R11-GFP reporter construct and with *pmar1* mRNA. Results are shown in Table 2.1 and Fig. 2.4, which compare the expression of GFP driven by the R11-GFP reporter construct in normal embryos (Table 2.1, column "R11"; Fig. 2.4A-F),

and in embryos globally expressing *pmar1* mRNA (Table 2.1, column "R11+ *pmar1*"; Fig. 2.4G-L). At the 15-17 h blastula stage, the number of embryos expressing GFP is significantly higher in embryos with ectopic *pmar1* expression (46%), than in normal embryos (29%). Most strikingly, some of these embryos displayed expression in several patches of cells, located on opposite sides of the embryo (Fig. 2.4G and J). This was never observed in normal embryos (Fig. 2.4A and D), and it indicated that expression of R11-GFP is no longer localized to the micromere lineage in embryos that ectopically express *pmar1* mRNA.

At early mesenchyme blastula stage the majority of embryos bearing ectopic *pmar1* mRNA expressed GFP in cells other than the micromere lineage: 78% of the expressing embryos display ectopic GFP expression when *pmar1*is expressed ectopically, whereas only 3% do so normally (Table 2.1). Figure 2.4 (H and K) clearly illustrate this effect. In these embryos expression of R11-GFP is observed in cells of the blastula wall, in addition to the ingressing cells that normally express the construct.

At 24-26 h after fertilization, the morphology of embryos undergoing global *pmar1* expression (Fig. 2.4I and L) is no longer normal (Fig. 2.4C and F), probably because all cells in the embryo have been transformed to PMC fate (Oliveri et al., 2002, 2003). It is now impossible to distinguish the cells that would have normally become PMCs from the rest of the cells. Note that at this stage almost all (96%) the embryos globally expressing *pmar1* display GFP (Table 2.1). So high a percentage of embryos expressing GFP can be expected only if R11 activates expression of GFP in any cell of the embryo where the exogenous R11-GFP DNA happens to be integrated. Also, the large size of the clones expressing GFP in these embryos (Fig. 2.4I and L) is also

consistent with the conclusion that R11 drives expression of GFP in all cells that also express *pmar1* mRNA.

The results shown in Table 2.1 and Fig. 2.4 are consistent with the hypothesis that wherever *pmar1* is active, the regulatory activity of R11 is derepressed, as predicted by the GRN model (Fig. 2.1). Since *pmar1* is normally transcribed only in the micromere descendants, R11 normally drives expression of GFP only in these cells, but when *pmar1* is expressed ectopically in all cells of the embryo, expression of GFP driven by R11 is also expanded to the whole embryo. The response of R11 to global expression of *pmar1* thus accurately recapitulates the response of the endogenous *delta* gene to the same perturbation, which causes expression of *delta* to expand to all cells (Oliveri et al., 2002). This equivalence provides strong support for the claim that the R11 element suffices to generate the control functions that govern *delta* expression in the cells of the micromere lineage.

Measurement of Incorporated Exogenous DNA and its Transcriptional Activity

To measure quantitatively the derepression of R11-GFP caused by global expression of *pmar1* mRNA, we developed what is essentially a new method of assessing expression of exogenous constructs *in vivo*. This relies on use of QPCR to assess both the amount of incorporated DNA and the amount of transcript generated from it in the experimental embryos.

An important preliminary consideration is that the amount of GFP mRNA that will be transcribed in a sample of embryos microinjected with the R11-GFP reporter construct will depend on the overall number of DNA molecules that happen to be

incorporated. Furthermore, as shown earlier (Livant *et al.*, 1990), we may assume that for a small amount of incorporated R11-GFP DNA, the amount of GFP mRNA transcribed will be linearly dependent on the number of incorporated R11-GFP DNA molecules. By "small" here is meant much smaller than the number of molecules of R11-GFP DNA required to saturate the transcription of GFP mRNA due to titration of the regulatory factors. Under these conditions the activity of R11-GFP reporter constructs in the different samples can be compared, by normalizing the absolute number of GFP mRNA molecules in each sample to the number of R11-GFP DNA molecules incorporated per embryo in that sample.

The diagram in Fig. 2.5A describes how the method was carried out. Total RNA was isolated along with small amounts of DNA from samples of embryos microinjected with R11-GFP. The number of molecules of R11-GFP DNA per embryo in these samples was estimated using QPCR. The single copy *foxB* gene was used as an internal standard to assess the number of genomes recovered, as described in Materials and methods. To quantify the levels of GFP mRNA the samples were treated with DNase I in order to remove all existing DNA. cDNA was then prepared from the sample and the levels of GFP mRNA were measured by QPCR. To confirm that the method we used consistently recovers genomic DNA as well as RNA, we tested the nucleic acids isolated from over 50 samples of embryos microinjected with R11-GFP reporter construct, and from more than 10 samples of embryos that had not been microinjected. In all samples the *foxb* sequence was amplified to detectable levels, indicating that sufficient genomic DNA had always been recovered (data not shown). GFP DNA was detected in all samples of embryos

microinjected with R11-GFP reporter construct, but not in any samples of uninjected embryos (data not shown).

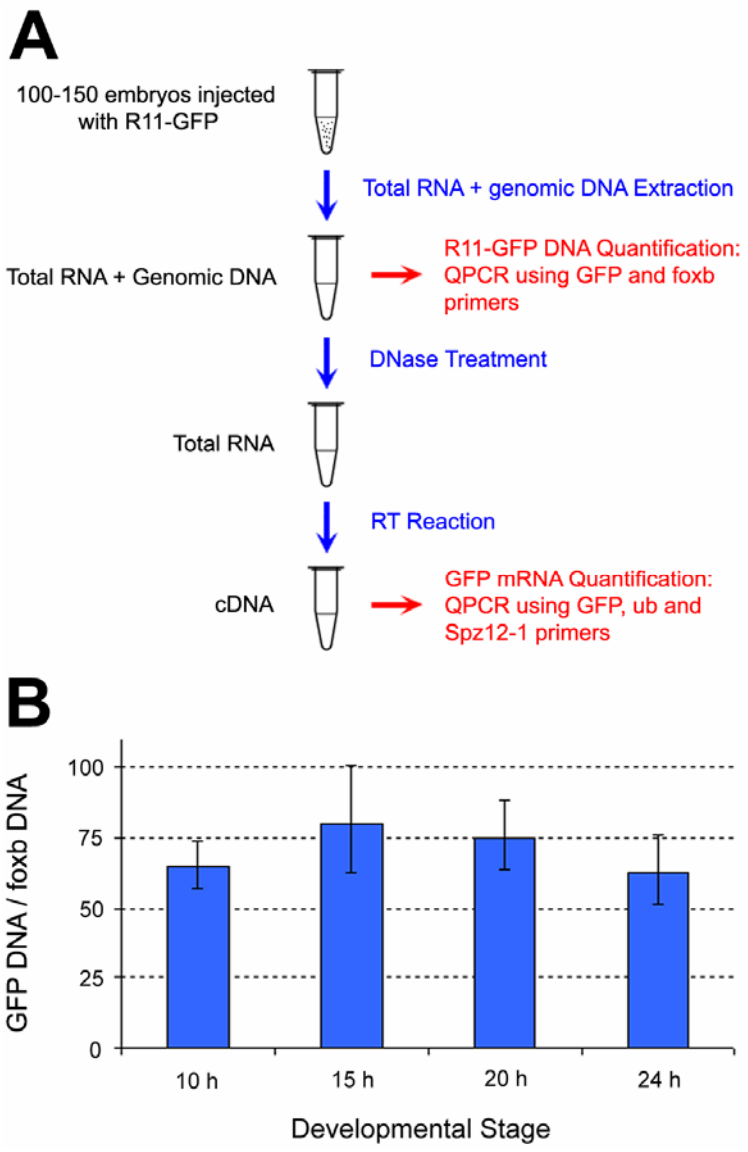


Fig. 2.5. A QPCR-based method to quantify the activity of exogenous constructs *in vivo*. (A) Schematic diagram describing the main steps of the method. (B) Demonstration that incorporated exogenous DNA replicates together with genomic DNA. The bars show the ratio of the number of DNA copies detected by the GFP primer set to the number of copies detected by the *foxb* primer set, at the indicated times postfertilization. Measurements were made on samples from which

total RNA and small amounts of DNA had been isolated using the "GenElute Mammalian Total RNA Miniprep Kit" (Sigma).

To confirm that the R11-GFP DNA detected by this method consisted mainly of DNA that had been incorporated into the genomes of the microinjected embryos, measurements of the number of molecules of R11-GFP DNA were made at several developmental stages. Figure 2.5B shows the relative amounts of DNA detected by the GFP primer set and the *foxb* primer set. We see that from the 10 h to 24 h stages, the ratio of R11-GFP DNA to *foxb* DNA remains constant. The genomic DNA (and hence the *foxb* DNA) is replicated about three times between 10 h and 24 h of development, and therefore the R11-GFP DNA must have been replicated together with the genomic DNA during this time. Only DNA that is incorporated into the genome of the microinjected embryos is replicated in sea urchin embryos (Flytzanis et al., 1985). Our result is the same as obtained by Franks *et al.* (1988) for injected expression constructs, using a different method. According to the data in Fig. 2.5B, we can estimate that at 10 h after fertilization there are ~10,000 molecules of R11-GFP DNA per embryo, and at 24 h there are ~60,000 molecules of R11-GFP DNA. The amount of DNA estimated at the 10 h stage represents ~15 times the amount of DNA microinjected in each embryo (~700 molecules of DNA were microinjected); and the amount estimated at the 24 h stage represents ~85 times the number of DNA molecules microinjected. Therefore, if any unincorporated exogenous DNA is detected at all, it represents an insignificant proportion of the detected DNA. The amount of exogenous DNA measured as incorporated into the genomes of the embryos of Fig. 2.5B is in fact the amount that would be present if most

of the microinjected DNA had been incorporated into a single blastomere genome between 3rd and 4th cleavage, as expected (Hough-Evans et al., 1988).

We have not calculated the efficiency of the RNA kit used in isolating genomic DNA (see Materials and methods). It is possible that the efficiency of this kit in recovering genomic DNA varies from sample to sample. But it is important to note that even were that the case, it would not affect these measurements, because the use of the internal *foxb* DNA standard renders the results independent of the absolute fraction of genomic DNA recovered. We need only assume that no part of the genome is isolated with a different systematic efficiency than any other part.

Timing and magnitude of the effect of ectopic* pmar1 *on R11 expression

The effect of global *pmar1* mRNA expression on the activity of R11-GFP, normalized to the amount of incorporated DNA is shown in Fig. 2.6A, and on the level of *delta* mRNA in Fig. 2.6C. In Fig. 2.6B the normalized activities of R11-GFP of Fig. 2.6A have all been multiplied by the number of R11-GFP DNA molecules incorporated in the control sample expressing the endogenous *pmar1* gene normally. This gives a direct comparison of the amounts of transcript that would have been produced had all the samples contained the same amount of exogenous DNA. The values in Fig. 2.6B still reflect normalized activities, and the advantage of this representation is that it allows us to compare the relative levels of GFP mRNA at different stages. More importantly this representation is equivalent to that of Fig. 2.6C, and therefore Fig. 2.6B can be directly compared to Fig. 2.6C.

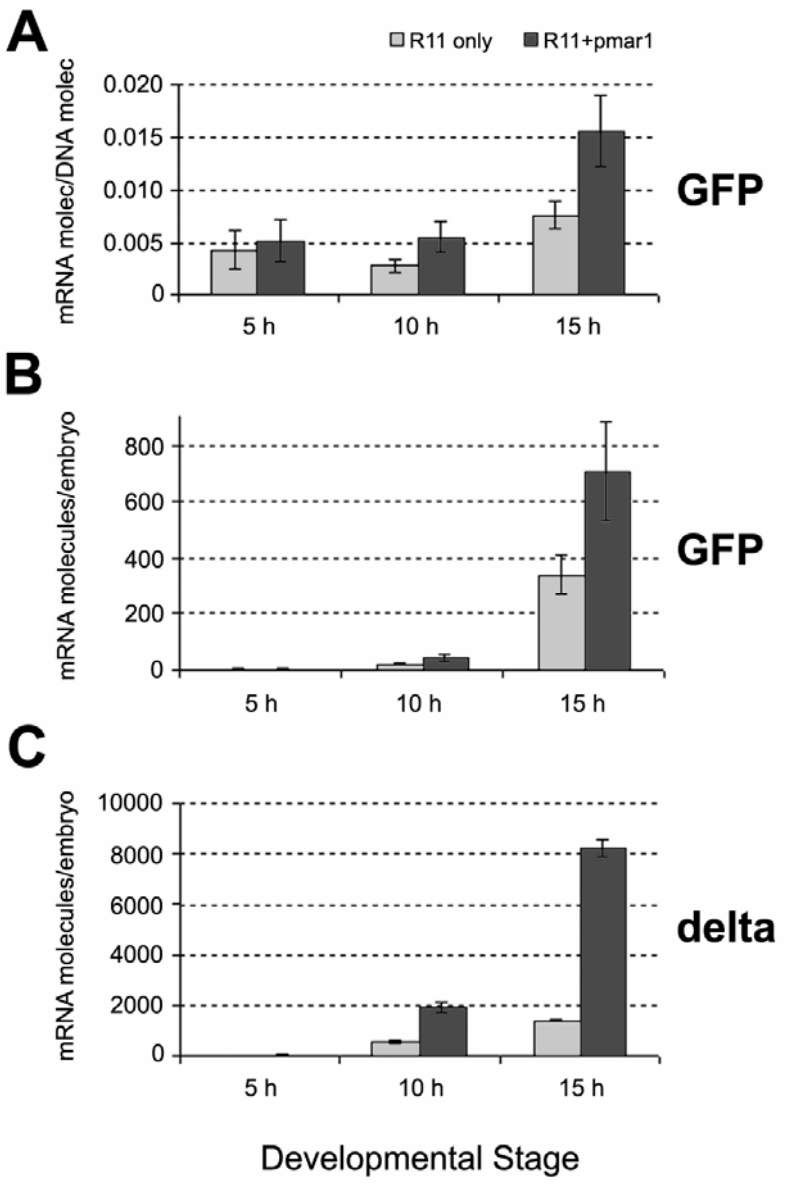


Fig. 2.6. Effect of global *pmar1* mRNA expression on the normalized activity of R11-GFP, and on the level of *delta* mRNA. Results from embryos microinjected with R11-GFP reporter construct are shown as light grey bars and from embryos simultaneously microinjected with the R11-GFP reporter construct plus *pmar1* mRNA as dark grey bars. (A) QPCR data indicating normalized levels of GFP mRNA (i.e., GFP mRNA/R11-GFP DNA) at different developmental stages. (B) Same QPCR data as in (A) multiplied by the number of R11-GFP DNA molecules incorporated in the control sample expressing the endogenous *pmar1*gene normally. This representation still reflects normalized activities and it can be directly compared to (C) (see Results). (C) QPCR data indicating amount of *delta* mRNA at different developmental stages.

Up to 5 h after fertilization, the *delta* gene is silent, and ectopic expression of *pmar1* mRNA has no effect on the very low observed levels of *delta* mRNA. But at the 10 h and 15 h stages, it results, respectively, in greater than 3- and 5-fold increases in endogenous *delta* mRNA (Fig. 2.6C). Similarly, the activity of R11-GFP is significantly increased (more than 2-fold) by ectopic *pmar1* expression at the 10 h and 15 h stages, but it is not affected at the 5 h stage (Figs. 2.6A and B). Even though Fig. 2.6A seems to indicate that R11-GFP is active at 5 h after fertilization, Fig. 2.6B clearly shows that the amount of GFP mRNA at this stage is insignificant; less than 5 molecules per embryo are detected at the 5 h stage. These results confirm that the derepression of the R11 regulatory element caused by ectopic *pmar1* mRNA can be detected as a quantitative increase in the activity of R11-GFP; and they also indicate that this happens at the same stages at which expression of endogenous *delta* is observed to increase in the same embryos.

It is important to note that in experiments in which the amount of incorporated R11-GFP DNA was ~10 times larger than in the experiment of Fig. 2.6, the amount of measured GFP mRNA was also ~10 times larger (data not shown). Therefore, in the experiment of Fig. 2.6 the amount of R11-GFP DNA incorporated was far from the amount of DNA required

DISCUSSION

We show here that the R11 DNA fragment contains *cis*-regulatory information sufficient to recreate the exact spatial and temporal pattern of the *delta* gene in its initial

phase of expression, when it is transcribed exclusively in the micromere lineage early in development. As we shall report elsewhere, a different *cis*-regulatory module of the *delta* gene reproduces the next phase of its expression in the secondary mesenchyme precursors. The properties of R11 bear directly on the GRN model for endomesoderm specification, as we discuss briefly below. But before this there are two methodological aspects of this work that bear consideration. These are the means by which R11 was found, and the means by which its response to experimental perturbation was quantitatively determined.

Identification of R11 by interspecific genomic sequence comparison

In the comparison of the orthologous genomic sequences of *S. purpuratus* and *L. variegatus* surrounding the *delta* gene, the R11 *cis*-regulatory element appears as a 3 kb-long block of very well conserved sequence, surrounded by very poorly conserved sequence (Fig. 2.3B). Conservation at the level of 90%-100% identity covers almost the entire block. Previous studies from this laboratory have already shown that the evolutionary distance between *S. purpuratus* and *L. variegatus* is very useful for identification of functional *cis*-regulatory elements (Brown *et al.*, 2002; Yuh *et al.*, 2002, 2004). The immediate identification of R11 by the same method adds further supporting evidence. R11 is located more than 13 kb downstream of the termination of the *delta* gene coding sequence (Fig. 2.3A), and finding this element by conventional mapping or deletion methods would have been extremely laborious. The FamilyRelations software (Brown *et al.*, 2002) was used for this interspecific sequence comparison, and other more or less equivalent sequence comparison methodologies have also been successful in

identifying *cis*-regulatory elements in many different genes and species pairs (e.g., Aparicio *et al*., 1995; Nonchev et al., 1996; Oeltjen et al., 1997; Brickner et al., 1999; Hardison, 2000; Loots et al., 2000; Manzanares et al., 2000; Muller et al., 2002). Many additional examples could be cited. Given the appropriate species distance for the gene in question, interspecific sequence comparison can be an extremely effective method for locating the control machinery of the genome; at the right distance, *cis*-regulatory elements stand out very clearly as conserved sequence patches. In the case we illustrate in Fig. 2.3, the signal to noise ratio is so high that the element is unmistakably distinguished from the surrounding sequence.

What is most impressive is how sharply defined are the boundaries of the element. These boundaries are revealed explicitly by the dot plot of Fig. 2.3B at the 9 out of 10 identity criterion here applied. This represents in principle a significant augmentation of methodologies for *cis*-regulatory analysis: experimental procedures generally provide either a convenient but much larger fragment than the actually functional regulatory module, or a "minimal element" that gives some function. We see that there is available an additional independent criterion, the computational definition of the natural boundaries of the conserved regulatory sequence patch.

Quantification of exogenous incorporated DNA and reporter mRNA

This work has included an augmentation of experimental *cis*-regulatory analysis methods as well. There are many applications when it is necessary to measure the output of an exogenous *cis*-regulatory expression construct in quantitative terms. Chief among these is to determine the effects of various mutations; and to determine the response of

the element to perturbation of a *trans* input that affects its activity, positively or negatively. So far, quantification of the level of expression of exogenous constructs in sea urchin embryos has been achieved by use of a reporter gene encoding chloramphenicol acetyltransferase (CAT), the enzymatic activity of which can be measured in lysates (e.g., Flytzanis et al., 1987; Livant et al., 1988; Kirchhamer et al., 1996; Yuh and Davidson, 1996; Yuh et al., 1996, 1998, 2001, 2004).

In general the amount of transcribed reporter mRNA depends on the number of molecules of the expression construct that are incorporated into the genomes of the microinjected embryos (Livant *et al.*, 1988). Flytzanis et al. (1987) used filter hybridization with radioactively-labeled probes to measure incorporated DNA, and concluded that if enough construct DNA is incorporated, the levels of transcribed reporter mRNA are independent of the amount of this DNA; in other words, the amounts of expression describe a saturation function with respect to the number of incorporated DNA molecules (cf. Livant et al., 1988, 1991). This fact has been exploited in a number of studies in order to analyze the quantitative effects of mutations on the kinetics of *cis*-regulatory expression (e.g., Yuh et al., 1998, 2001).

Here we describe a new method, based on QPCR measurements, for the simultaneous quantification of transcribed reporter mRNA and incorporated reporter DNA. This method provides certain advantages with respect to measurement of CAT activity. First, since the level of transcription is obtained by directly measuring the amount of reporter mRNA at given times, the result depends only on the rates of construct transcription and of reporter mRNA turnover, rather than on these rates plus the rates of reporter protein synthesis and protein turnover. The last is particularly difficult to

measure or estimate. Second, the QPCR method is compatible with the use of any reporter gene, including GFP, rather than limited to the use of the CAT reporter. Thus, for example, measurements can be carried out on samples of embryos that have previously also been scored for spatial GFP fluorescence. Third, the amount of incorporated reporter DNA is very easily quantified at the same time, and in the same sample of embryos in which the reporter mRNA is measured. This provides a very efficient way of normalizing the levels of reporter mRNA to the amount of incorporated reporter DNA, which is a major source of variation in the activity of different batches of embryos. The major advantage of this normalization is that it is no longer required to microinject enough DNA so that transcription of the reporter gene reaches saturation. Finally, the QPCR method allows for measurement of the expression of any endogenous gene(s) in the same sample of embryos in which the levels of reporter mRNA and DNA are quantified. This can be particularly useful for analysis of the effects of perturbations on an incorporated *cis*-regulatory element.

cis-*Regulatory analysis of R11 expression and the network model for endomesoderm specification*

The GRN model predicts genomically encoded *cis*-regulatory interactions that would explain the expression of its constituent genes at the right places and times to serve their developmental functions in the specification process (Davidson et al., 2002a,b; Oliveri and Davidson, 2004; for current version of this model, see http://sugp.caltech.edu/endomes/). The *cis*-regulatory element controlling early *delta* gene expression in the micromere lineage is a particularly important node of the GRN: it

accounts for transcriptional expression of the spatial information that sets in train the specification of the secondary mesenchyme domain of the embryo. The specific prediction is that the expression of *delta* in the micromere lineage under control of this *cis*-regulatory element depends on activating factors that are ubiquitously present, and on a repressor ("Repressor of mic" in Fig. 2.1) that is in turn repressed exclusively in the cells of the micromere lineage in consequence of *pmar1* expression (Fig. 2.1; Oliveri et al., 2002). The isolation and experimental analysis of the R11 *delta cis*-regulatory element reported here proves that there indeed exists a genomic DNA fragment that executes exactly the predicted interactions.

In untreated embryos, R11 accurately drives expression of the reporter construct, exclusively in the micromere lineage, while in embryos globally expressing *pmar1* mRNA, R11 becomes capable of causing expression in any cell of the embryo. This behavior perfectly reproduces the response of the endogenous *delta* gene to the same perturbation. R11 may contain target sites for activating factors that are ubiquitously present, and it may also contain the sites for the repressor controlled by the *pmar1* gene product. However, until such sites are identified by mutation it remains possible that this repression is mediated indirectly, and that R11 (and the *delta* gene) are controlled by a localized activator which is under *pmar1* system control. But the kinetics of *delta* gene expression, which very shortly follows *pmar1* activation (Oliveri *et al.*, 2002; Fig. 2.2 of this paper), suggest that the repression is likely to be exerted directly.

This work illustrates one of the major useful aspects of the GRN model, *viz.,* that the model specifies experimentally testable candidate inputs into each of its *cis*-regulatory elements. In turn experimental *cis*-regulatory analysis feeds back into the

network model by validating these predictions. As such analysis is extended to the key nodes of the GRN there emerges an explanatory structure that will directly represent the genomic regulatory code underlying specification and development.

ACKNOWLEDGMENTS

We are very grateful to Dr. Ellen Rothenberg and Dr. Paola Oliveri for very useful criticism and fruitful discussions. We also thank Pat Leahy for taking care of the sea urchins used for this work, and Jane Wyllie and Dr. Veronica Hinman for technical guidance. The first author also thanks Dr. Albert Erives for his instruction on molecular biology, and he would like to apologize to the sea urchins sacrificed for this work. This research was supported by NIH Grants HD37105 and GM61005 and by NASA/Ames Grant NAG2-1587.

REFERENCES

Aparicio, S., Morrison, A., Gould, A., Gilthorpe, J., Chaudhuri, C., Rigby, P., Krumlauf, R., Brenner, S. 1995. Detecting conserved regulatory elements with the model genome of the Japanese puffer fish, *Fugu rubripes*. Proc. Natl. Acad. Sci. USA 92, 1684-1688.

Arnone, M.I., Davidson, E.H. 1997. The hardwiring of development: organization and function of genomic regulatory systems. Development 124, 1851-1864.

Brickner, A.G., Koop, B.F., Aronow, B.J., Wiginton, D.A. 1999. Genomic sequence comparison of the human and mouse adenosine deaminase gene regions. Mammalian Genome 10, 95-101.

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

Cameron, R.A., Hough-Evans, B.R., Britten, R.J., Davidson, E.H. 1987. Lineage and fate of each blastomere of the eight-cell sea urchin embryo. Genes Dev. 1, 75-84.

Cameron, R.A., Mahairas, G., Rast, J.P., Martinez, P., Biondi, T.R., Swartzell, S., Wallace, J.C., Poustka, A.J., Livingston, B.T., Wray, G.A., Ettensohn, C.A., Lehrach, H., Britten, R.J., Davidson, E.H., Hood, L. 2000. A sea urchin genome project: Sequence scan, virtual map, and additional resources. Proc. Natl. Acad. Sci. USA 97, 9514-9518.

Cameron, R.A., Oliveri, P., Wyllie, J., Davidson, E.H. 2004. *cis*-Regulatory activity of randomly chosen genomic fragments from the sea urchin. Gene Exp. Patt. 4, 205-213.

Davidson, E.H. 1986. Gene Activity in Early Development, Second Edition. Academic Press, New York.

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. j., Schilstra, M.J., Clarke, P.J.C., Arnone, M. I., Rowen, L., Cameron, R.A., McClay, D.R., Hood, L., Bolouri, H. 2002a. A genomic regulatory network for development. Science 295, 1669-1678.

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.C., 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, 162-190.

Ettensohn, C.A., Illies, MR., Oliveri, P., De Jong, D.L. 2003. Alx1, a member of the Cart1/Alx3/Alx4 subfamily of Paired-class homeodomain proteins, is an essential component of the gene network controlling skeletogenic fate specification in the sea urchin embryo. Development 130, 2917-2928.

Flytzanis, C.N., McMahon, A.P., Hough-Evans, B.R., Katula, K.S., Britten, R.J., Davidson, E.H. 1985. Persistence and integration of cloned DNA in postembryonic sea urchins. Dev. Biol. 108, 431-442.

Flytzanis, C.N., Britten, R.J., Davidson, E.H. 1987. Ontogenic activation of a fusion gene introduced into sea urchin eggs. Proc. Natl. Acad. Sci. USA 84, 151-155.

Franks, R.R., Hough-Evans, B.R., Britten, R.J., Davidson, E.H. 1988. Direct introduction of cloned DNA into the sea urchin zygote nucleus, and fate of injected DNA. Development 102, 287-299.

Franks, R.R., Anderson, R., Moore, J.G., Hough-Evans, B.R., Britten, R.J., Davidson, E.H. 1990. Competitive titration in living sea urchin embryos of regulatory factors required for expression of the *CyIIIa* actin gene. Development 110, 31-40.

Fuchikami, T., Mitsunaga-Nakatsubo, K., Amemiya, S., Hosomi, T., Watanabe, T., Kurokawa, D., Kataoka, M., Harada, Y., Satoh, N., Kusunoki, S., Takata, K., Shimotori, T., Yamamoto, T., Sakamoto, N., Shimada, H., Akasaka, K. 2002. T-brain homologue (HpTb) is involved in the archenteron induction signals of micromere descendant cells in the sea urchin embryo. Development 129, 5205-5216.

Hardison, R.C. 2000. Conserved noncoding sequences are reliable guides to regulatory elements. Trends Genet. 16, 369-372.

Hough-Evans, B.R., Britten, R.J., Davidson, E.H. 1988. Mosaic incorporation and regulated expression of an exogenous gene in the sea urchin embryo. Dev. Biol. 129, 198-208.

Kurokawa, D., Kitajima, T., Mitsunaga-Nakatsubo, K., Amemiya, S., Shimada, H., Akasaka, K. 1999. HpEts, an *ets*-related transcription factor implicated in primary mesenchyme cell differentiation in the sea urchin embryo. Mech. Dev. 80, 41-52.

Livant, D.L., Cutting, A.E., Britten, R.J., Davidson, E.H. 1988. An in vitro titration of regulatory factors required for expression of a fusion gene in transgenic sea urchin embryos. Proc. Natl. Acad. Sci. USA 85, 7607-7611.

Livant, D.L., Hough-Evans, B.R., Moore, J.G., Britten, R.J., Davidson, E.H. 1991. Differential stability of expression of similarly specified endogenous and exogenous genes in the sea urchin embryo. Development 113, 385-398.

Loots, G.G., Locksley, R.M., Blankspoor, C.M., Wang, Z.E., Miller, W., Rubin, E.M., Frazer, K.A. 2000. Identification of a coordinate regulator of interleukins 4, 13, and 5 by cross-species sequence comparisons. Science 288, 136-140.

Luke, N.H., Killian, C.E., Livingston, B.T. 1997. Spfkh1 encodes a transcription factor implicated in gut formation during sea urchin development. Dev. Growth Differ. 39, 285-294.

Makabe, K.W., Kirchhamer, C.V., Britten, R.J., Davidson, E.H. 1995. *cis*-Regulatory control of the *SM50* gene, an early marker of skeletogenic lineage specification in the sea urchin embryo. *Development* 121, 1957-1970.

Manzanares, M., Wada, H., Itasaki, N., Trainor, P.A., Krumlauf, R., Holland, P.W.H. 2000. Conservation and elaboration of *Hox* gene regulation during evolution of the vertebrate head. Nature 408, 854-857.

McClay, D.R., Peterson, R.E., Range, R.C., Winter-Vann, A.M., Ferkowicz, M.J. 2000. A micromere induction signal is activated by beta-catenin and acts through notch to initiate specification of secondary mesenchyme cells in the sea urchin embryo. Development 127, 5113-5122.

McMahon, A.P., Flytzanis, C.N., Hough-Evans, B.R., Katula, K.S., Britten, R.J., Davidson, E.H. 1985. Introduction of cloned DNA into sea urchin egg cytoplasm: Replication and persistence during embryogenesis. Dev. Biol. 108, 420-430.

Muller, F., Blader, P., Strahle, U. 2002. Search for enhancers: teleost models in comparative genomic and transgenic analysis of *cis*-regulatory elements. Funct. Genomics Bioinfo**.** 24, 564-572.

Nemer, M., Rondinelli, E., Infante, D., Infante, A.A. 1991. Polyubiquitin RNA characteristics and conditional introduction in sea urchin embryos. Dev. Biol. 145, 255-265.

Nonchev, S., Maconochie, M., Vesque, C., Aparicio, S., Ariza-McNaughton, L., Manzanares, M., Maruthainar, K., Kuroiwa, A., Brenner, S., Charnay, P., Krumlauf, R. 1996. The conserved role of Krox-20 in directing *Hox* gene expression during vertebrate hindbrain segmentation. Proc. Natl. Acad. Sci. USA 93, 9339-9345.

Oeltjen, J.C., Malley, T.M., Muzny, D.M., Miller, W., Gibbs, R.A., Belmont, J.W. 1997. Large-scale comparative sequence analysis of the human and murine Bruton's tyrosine kinase loci reveals conserved regulatory domains. Genome Res. 7, 315-329.

Oliveri, P., Carrick, D M., Davidson, E.H. 2002. A regulatory gene network that directs micromere specification in the sea urchin embryo. Dev. Biol. 246, 209-228.

Oliveri, P., Davidson, E.H., McClay, D.R. 2003. Activation of *pmar1* controls specification of micromeres in the sea urchin embryo. Dev. Biol. 258, 32-43.

Ransick, A., Rast, J.P., Minokawa, T., Calestani, C., Davidson, E.H. 2002. New early zygotic regulators expressed in endomesoderm of sea urchin embryos discovered by differential array hybridization. Dev. Biol. 246, 132-147.

Rast, J.P. 2000. Transgenic manipulation of the sea urchin embryo. In: Tuan, R. S. Lo, C.W. (Eds.) Methods in Molecular Biology, Vol 136: Developmental Biology Protocols, Vol II. Human Press Inc., Totowa, NJ, pp. 365-373.

Sweet, H.C., Hodor, P.G., Ettensohn, C.A. 1999. The role of micromere signaling in Notch activation and mesoderm specification during sea urchin embryogenesis. Development 126, 5255-5265.

Sweet, H.C., Gehring, M., Ettensohn, C.A. 2002. LvDelta is a mesoderm-inducing signal in the sea urchin embryo and can endow blastomeres with organizer-like properties. Development 129, 1945-1955.

Wang, D.G.-W., Britten, R.J., Davidson, E.H. 1995. Maternal and embryonic provenance of a sea urchin embryo transcription factor, SpZ12-1. Mol. Marine Biol. Biotechnol. 4, 148-153.

Yuh, C.-H., Davidson, E.H. 1996. Modular *cis*-regulatory organization of *Endo16*, a gut-specific gene of the sea urchin embryo. Development 122, 1069-1082.

Yuh, C.-H., Moore, J.G., Davidson, E.H. 1996. Quantitative functional interrelations within the *cis*-regulatory system of the *S. purpuratus Endo16* gene. Development 122, 4045-4056.

Yuh, C.-H., Bolouri, H., Davidson, E.H. 1998. Genomic *cis*-Regulatory logic: Experimental and computational analysis of a sea urchin gene. Science 279, 1896-1902.

Yuh, C.-H., Bolouri, H., Davidson, E.H. 2001. *cis*-Regulatory logic in the *endo16* gene: Switching from a specification to a differentiation mode of control. Development 128, 617-629.

Yuh, C.-H., Brown, C.T., Livi, C.B., Rowen, L., Clarke, P.J.C., Davidson, E.H. 2002. Patchy interspecific sequence similarities efficiently identify positive *cis*-regulatory elements in the sea urchin. Dev. Biol. 246, 148-161.

Yuh, C.-H., Dorman, E.R., Howard, M.L., Davidson, E.H. 2004. An *otx cis*-regulatory module: A key node in the sea urchin endomesoderm gene regulatory network. Dev. Biol. 269, 536-551.

Zhu, X., Mahairas, G., Illies, M., Cameron, R.A., Davidson, E.H., Ettensohn, C.A. 2001. A large-scale analysis of mRNAs expressed by primary mesenchyme cells of the sea urchin embryo. Development 128, 2615-2627.

CHAPTER 3

A "Missing Link" in the Sea Urchin Embryo Gene Regulatory Network: *hesC* and the Double Negative Specification of Micromeres

Roger Revilla-i-Domingo, Paola Oliveri, and Eric H. Davidson

Submitted for publication to *Proceedings of the National Academy of Science USA*

ABSTRACT

Specification of sea urchin embryo micromeres occurs early in cleavage, with the establishment of a well defined regulatory state. The architecture of the gene regulatory network controlling the specification process indicates that transcription of the initial tier of control genes depends on a double negative gate. A gene encoding a transcriptional repressor, *pmar1*, is activated specifically in micromeres where it represses transcription of a second repressor that is otherwise active globally. Thus the micromere specific control genes which are the target of the second repressor are expressed exclusively in this lineage. The double negative specification gate was logically required from the results of numerous prior experiments, but the identity of the gene encoding the second repressor remained elusive. Here we show that *hesC* is this gene, and demonstrate experimentally all of its predicted functions, including global repression of micromere specific regulatory genes. As logically required, blockade of *hesC* mRNA translation and

global overexpression of *pmar1* mRNA have the same effect, which is to cause all the cells of the embryo to express micromere-specific genes.

KEY WORDS: *Gene regulatatory networks / skeletogenic micromeres / transcriptional repression*

INTRODUCTION

The genomic regulatory code for specification of endomesoderm in the sea urchin embryo is represented as a gene regulatory network (GRN), which explains the mechanism by which distinct regulatory states are deployed in different territories of the developing embryo (for reviews, (1-4); for current version see http://sugp.caltech.edu/endomes/). One portion of this GRN pertains to the specification of the micromeres, which arise at the unequal fourth cleavage at which the four micromeres are segregated off from the vegetal pole of the egg. The large daughter cells of the micromeres arising at the next cleavage are the founder cells of the skeletogenic micromere lineage. This lineage is the sole normal source of the embryonic biomineral skeleton, a distinct synapomorphic feature of echinoid embryos and larvae, and it also produces essential short range signals required for other aspects of endomesoderm specification (5-7). Three particular developmental events that are relevant for what follows are the expression of the Delta signaling ligand on the surfaces of the micromere descendants during the early blastula stage (Fig. 3.1A); their ingression into the blastocoel at late blastula stage (Fig. 3.1B), after which they are known as primary

mesenchyme cells (pmc's); and their expression of the biomineralization and cytoskeletal genes which enable them to generate the skeleton (Fig. 3.1C; (8)).

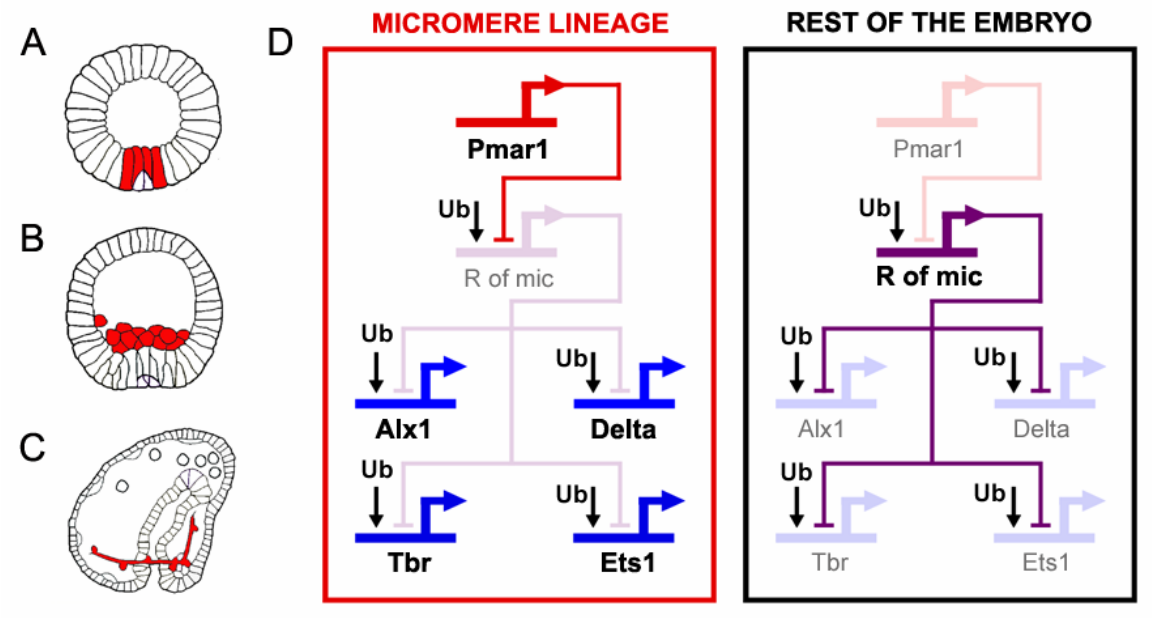


Figure 3.1: Key elements of the GRN model for the early specification of the skeletogenic micromere lineage. The model is based on (9), with subsequent updates (10), as reviewed in (1, 4). **(A-C)** Sea urchin embryo drawings (adapted from (2) at early blastula stage (12–15 h after fertilization; A), mesenchyme blastula stage (24 h; B), and late gastrula stage (48 h; C). The cells of the skeletogenic micromere lineage at each stage are depicted in red. **(D)** GRN model (corresponding to cleavage and blastula stages). Active genes are represented in strong color and bold font. Inactive genes are represented in dim color. Within the micromere lineage *pmar1* is active, and it represses the predicted gene *r of mic*. The *delta, alx1, ets* and *tbr* genes are allowed to be zygotically expressed in this domain. In the rest of the embryo *r of mic* keeps *delta, alx1, ets* and *tbr* silent.

Immediately after the fourth cleavage micromeres are born they express a gene, *pmar1*, in response to maternally localized factors (9). This gene encodes a transcriptional repressor of the paired homeodomain family. In the GRN *pmar1* serves as the linchpin of a double negative gate controlling the institution of the micromere

regulatory state. The second component of this gate is a *pmar1* target gene which encodes another transcriptional repressor. This gene is also zygotically expressed but it is transcribed everywhere in the embryo except in the micromere lineage, where it is subject to repression by Pmar1. It has eight known targets included in the GRN, of which the most important for present purposes are the genes encoding the Delta ligand, and three regulatory genes, *tbr*, *ets*, and *alx1*. These three genes lie upstream of all the rest of the micromere regulatory apparatus. Thus the double negative gate ensures expression of this apparatus exclusively in the micromere lineage. Because its identity was unknown, the second repressor has been referred to in the GRN as "Repressor of Micromeres" or "R of mic." Its existence and its properties are specifically implied by the two following perturbation experiments (9, 10): first, if expression of *pmar1* is forced to occur globally (by injection into the egg of the mRNA), then the *delta*, *tbr*, *ets*, *alx1* and downstream genes are transcribed in all cells of the embryo, and all cells thereby adopt skeletogenic micromere lineage fate; second, exactly the same outcome follows if an mRNA encoding a dominantly repressive Engrailed fusion of the Pmar1 protein is injected. It follows that the *pmar1* gene product naturally acts as a repressor (also indicated by its sequence); that *delta*, *tbr*, *ets* and *alx1* are controlled by ubiquitous activators; and that localization of expression of these genes to the micromere lineage in normal embryos depends on their repression by R of mic everywhere else in the embryo (Fig. 3.1D).

To prove the existence of the double negative gate for micromere lineage specification in the GRN model, it is necessary to find the gene playing the role of the predicted R of mic, and to establish that its expression and its functions are also as predicted. The *r of mic* gene should encode a transcriptional repressor, and it should have

three very distinct characteristics: (a) Its zygotic expression should be trancriptionally repressed by Pmar1; (b) it should be zygotically expressed everywhere except in the micromere lineage by the time zygotic expression of *delta, alx1, ets1 and tbr* starts; (c) the outcome of knocking down its expression should be similar to forcing global Pmar1 expression, i.e., all cells of the embryo should adopt micromere lineage specification, and express *delta*, *alx1, ets1* and *tbr*.

RESULTS

Genomic screen for candidate *r of mic* genes

The *S. purpuratus* genome sequence enabled consideration of all sea urchin genes encoding transcription factors in our search for *r of mic*. The total number of annotated transcription factors in this genome excluding C2H2 Zinc Finger genes is 283 (11), and the total number of predicted C2H2 Zinc Fingers (some of which encode transcription factors) is 377 (12). The levels of mRNA expression at several developmental time points were measured for all of these 660 genes (12-16). We selected as *r of mic* candidates all putative regulatory genes for which at least 200 transcripts were detected per embryo at 12 h after fertilization, when *delta*, *alx1, ets1* and *tbr* are all zygotically transcribed. This is a conservative (low) threshold, given that *r of mic* must be expressed in most of the embryo, or in 100-150 cells, at this time. This resulted in a list of about 100 candidate genes. We excluded those which are maternally and not zygotically expressed up to 12 h after fertilization, and all previously studied transcription factors for which enough

information was available to confirm that they could not be *r of mic*. The surviving list now contained 46 candidates (Supplementary Table 1 in appendix 3).

Since *r of mic* should be transcriptionally repressed by Pmar1, we screened the 46 candidate genes for down-regulation upon forced expression of Pmar1 in the whole embryo (Fig. 3.2; Supplementary Figure 1 in appendix 3; mRNA overexpression, MOE). The effect of *pmar1* mRNA MOE on the level of transcripts of each R of mic candidate gene was measured at 9 h and 12 h after fertilization by using Quantitative PCR (QPCR). The *delta* gene was included in the screen as a control. As expected, in the two experiments performed, delta was significantly up-regulated (3-fold or greater changes in transcript levels were considered significant) both at 9 h and 12 h after fertilization (Fig. 3.2). This indicated that *r of mic* must have been down-regulated at both time points in these two experiments. Five of the 46 regulatory genes tested were found to be significantly down-regulated at both time points in the two experiments performed. These were *six3*, *smadIP*, *awh*, *hesC* and *foxJ1* (Fig. 3.2).

Among these five transcriptional regulatory genes, hesC particularly caught our attention. Its level of mRNA expression at 9 h and 12 h is highest of all five (data not shown). In addition, HesC is a bHLH transcription factor belonging to the HES (Hairy/E(spl)) family, and almost all transcription factors of this family are known to function as repressors (17). That HesC belongs to this family is supported by a phylogenetic analysis (16), and by the fact that it contains the two characteristic domains of the family: the C-terminus WRPW motif (used to recruit TLE/Grg/Groucho and mediate transcriptional repression (18, 19), and the Orange domain. We therefore focused on *hesC*.

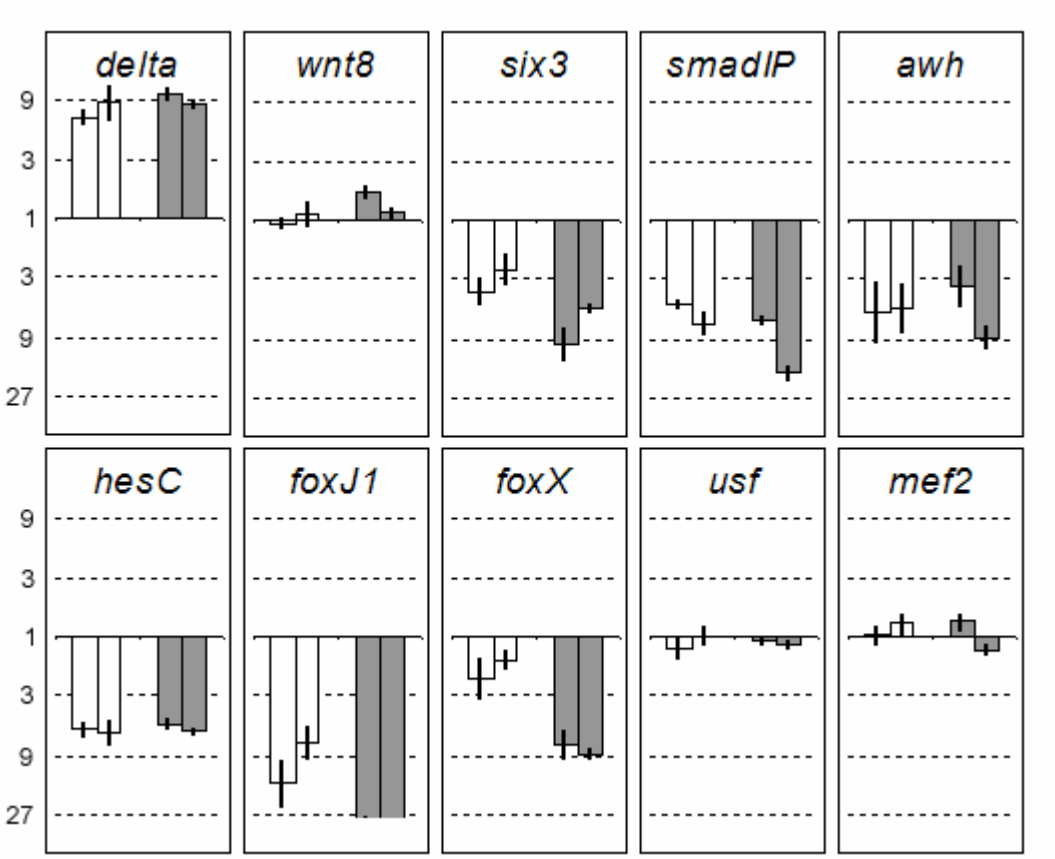


Figure 3.2: Pmar1 MOE screen. Graphs showing fold change in mRNA expression for *r of mic* candidate genes upon overexpression of Pmar1 mRNA. *delta* and *wnt8* were used as positive and negative controls, respectively. A fold change of "1" (solid line) indicates "no change." The numbers situated above "1" indicate fold increase, and the numbers situated below "1," fold decrease (in logarithmic scale). A 3-fold or greater change was considered to be significant. White bars and grey bars represent data from samples at 9 h and 12 h of development, respectively. For each color, the two bars correspond to two independent batches of embryos. Error bars represent the standard deviation from three independent measurements on the same sample. Results for only 8 of the 46 *r of mic* candidate genes are shown here (see Supplementary Figure 1 in appendix 3 for data on the remaining 38 candidates).

Temporal and spatial expression of HesC

The spatial and temporal patterns of expression predicted for the *r of mic* gene are unique.

The time course of *hesC* expression was determined at 1-2 hour intervals by means of

QPCR (Fig. 3.3A). This showed that *hesC* is maternally expressed, but only at very low levels. The level of *hesC* transcript then increases steeply between 8 h and 12 h after fertilization, indicating zygotic transcription. To compare the temporal expression of *hesC* to that of upstream and downstream genes in the double negative gate, we measured the levels of *pmar1* and *delta* mRNA in the same embryo samples (Fig. 3.3A). As would be expected for *r of mic*, the zygotic expression of *hesC* starts before *delta* is detected, and it occurs while *pmar1* mRNA is present.

Whole Mount In Situ Hybridization (WMISH) provided strong evidence. At 8 h the steep zygotic expression of *hesC* has just started and at this time, *hesC* mRNA is found essentially everywhere in the embryo, including the micromere lineage (Fig. 3.3F; compare to control in Fig. 3.3B). The two small cells at the vegetal pole of the embryo, which show weaker staining than the rest of the embryo, are the "small micromeres," which do not belong to the skeletogenic micromere lineage which is the subject of this article. At 12 h the steep zygotic increase in HesC expression has attained its plateau value (Fig. 3.3A), and *delta* mRNA is already present. There is now a dramatic change in *hesC* expression, in that this transcript has disappeared from a set of 12 cells at the vegetal pole (Fig. 3.3C,G), while it continues to be expressed everywhere else. Exactly 12 cells express *delta* mRNA at this time (Fig. 3.3D, H), and this is the number of cells now in the micromere lineage. To confirm that the 12 cells lacking *hesC* mRNA expression correspond to the 12 micromere lineage cells, we performed double-WMISH. As shown in Fig. 3.3(E,I), every cell of the embryo expresses either *hesC* (purple) or *delta* (orange), but no cell expresses both genes. Zygotic expression of *hesC*, therefore,

occurs everywhere in the embryo except the micromere lineage, precisely as predicted for *r of mic*.

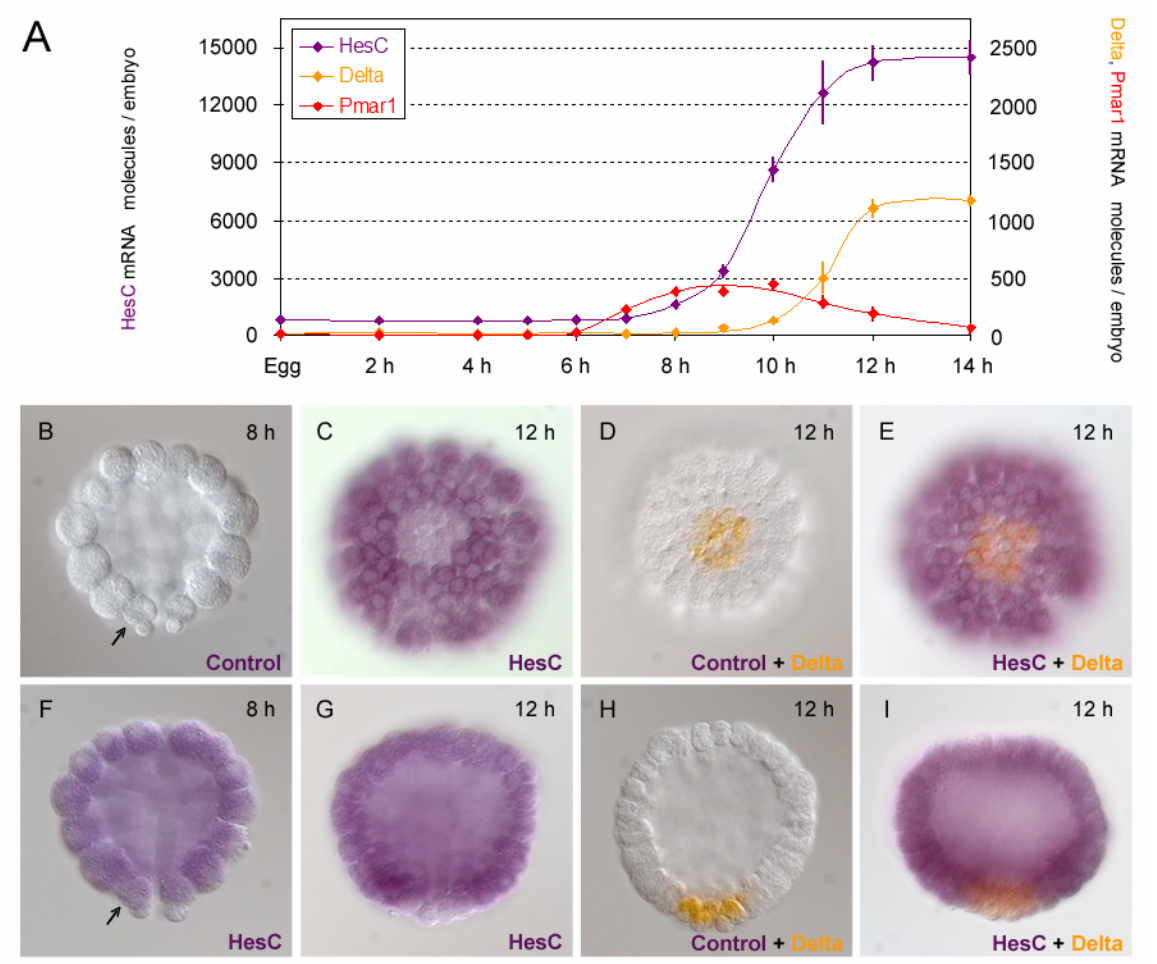


Figure 3.3: HesC temporal and spatial expression pattern. (A) Measurements of *hesC* mRNA molecules per embryo (purple) are compared to those of *pmar1* (red) and *delta* (orange) at the indicated developmental time points. Error bars represent the standard deviation from three individual measurements on the same sample. **(B-I)** Images of embryos on which WMISH (B, C, F and G) or double-WMISH (D, E, H and I) was performed. The developmental stage of each embryo is indicated at the upper right corner. Panels B, F, G, H and I are side views, with vegetal side at the bottom. Panels C, D and E are vegetal views. The arrows in B and F point at one of the two visible skeletogenic micromere cells. Probe(s) used are indicated at the lower right corner of each panel. Control: Probe used to control for nonspecific staining.

Functional analysis of *hesC*

The predicted function of *r of mic* is to repress micromere lineage specification. Thus if *hesC* is *r of mic*, blocking its translation should result in all cells of the embryo becoming specified similarly to skeletogenic micromeres, the same as when global *pmar1* expression is forced to occur (9). We used a morpholino antisense oligonucleotide (MASO) targeting *hesC* mRNA for this experiment. The striking effect of this perturbation on the morphology of the developing embryos is shown in Fig. 3.4. Up to blastula stage, *hesC* MASO embryos were indistinguishable from unperturbed embryos (Fig. 3.4A,C). In both, ingression of pmcs into the blastocoel started ~20 h after fertilization (not shown). However, while in unperturbed embryos, pmc ingression had been completed by 24 h after fertilization (Fig. 3.4B), in *hesC* MASO embryos ingression of cells continued until the blastocoel was essentially full (Fig. 3.4D). All, or almost all, cells of *hesC* MASO embryos thus behave in a way normally unique to the micromere lineage. Importantly, at all three stages, *hesC* MASO embryos look strikingly similar to *pmar1* MOE embryos (Figs. 3.4C,D,E and F; data not shown for 20 h stage).

We next assessed the effect of HesC MASO perturbation on the levels of mRNA of *delta*, *alx1*, *ets* and *tbr*. If *hesC* is *r of mic* the prediction (Fig. 3.1D) is that these genes will now be allowed to be expressed in all cells, and their level of transcript should therefore increase, as occurs in *pmar1* MOE embryos (9, 10). Fig. 3.5 and Supplementary Table 2 (appendix 3) show this result. By 12 h after fertilization, the amount of transcript of *delta* and *alx1*had increased 4- to 7-fold above normal in the two experiments performed (Fig. 3.5A), and by 24 h that of *ets* and *tbr* had similarly increased (Supplementary Table 2 in appendix 3). The level of expression of *pmar1* was not

affected, indicating that the up-regulation of these genes was not caused by any change in *pmar1* (Fig. 3.5A).

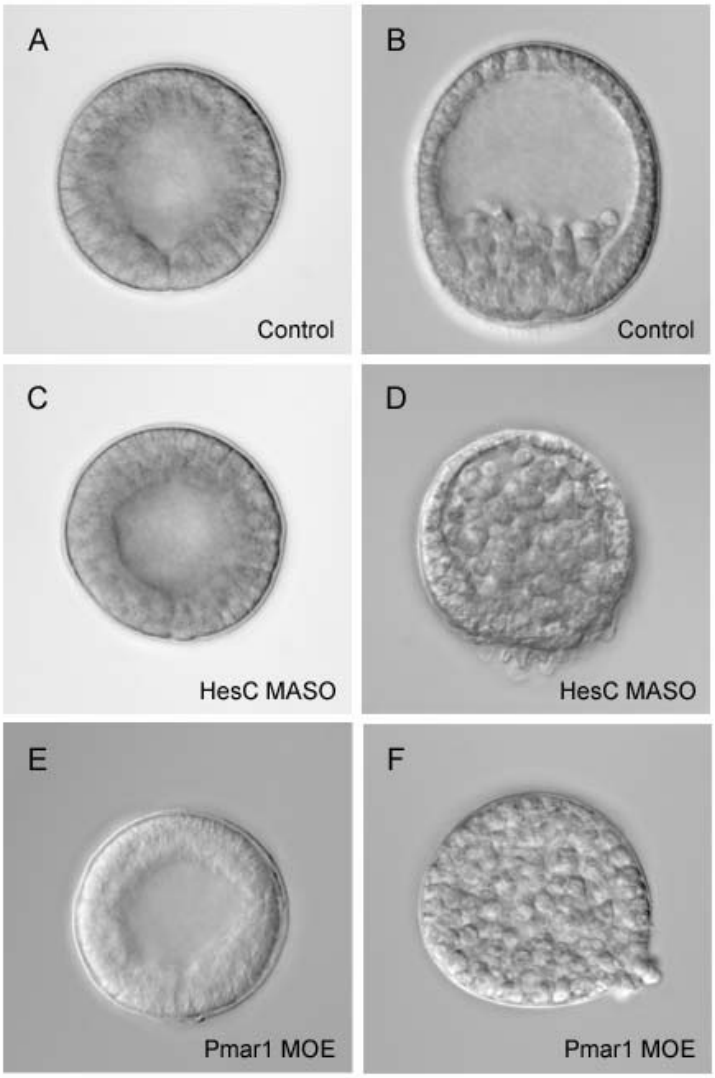


Figure 3.4: Morphology of HesC MASO embryos. Images of embryos that were either unperturbed (A and B), or that had been perturbed by HesC MASO (C and D). Pmar1 MOE embryos from a different batch are also shown for comparison (E and F). **(A, C and E)** Blastula stage embryos, 16 h after fertilization. **(B, D and F)** Late mesenchyme blastula stage embryos, 24-26 h after fertilization.

The derepression of micromere lineage specification occurs in all cells of *hesC* MASO embryos. This is illustrated for the *delta* marker as shown in Figs. 3.5B-E. While in unperturbed embryos *delta* mRNA is localized to the micromere lineage (Fig. 3.5B,D), in *hesC* MASO embryos it is detected throughout the whole embryo (Fig. 3.5C,E). HesC thus functions to repress micromere lineage specification in all cells other than the micromere lineage, the defining characteristic of the predicted *r of mic*.

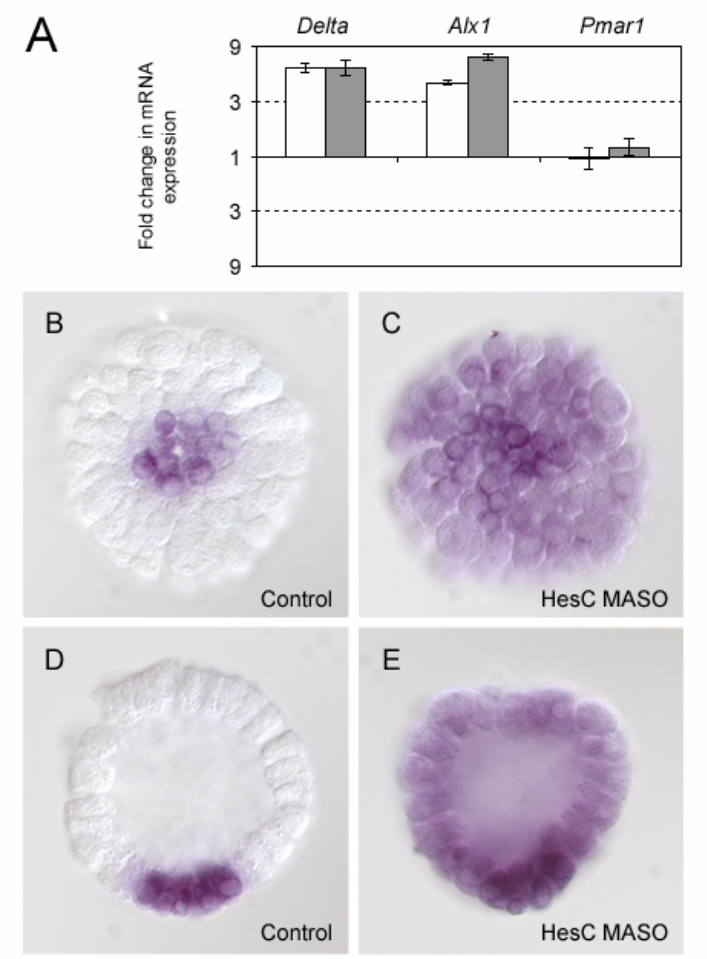


Figure 3.5: Effect of HesC MASO on micromere lineage genes. (A) The graph shows the fold change in *delta*, *alx1* and *pmar1* mRNA expression in HesC MASO embryos relative to unperturbed embryos (12 h after fertilization). Fold change representation is as in Fig. 3.2. White and grey bars indicate two independent batches of embryos. **(B-E)** Images of embryos (12 h after

fertilization) on which WMISH was performed using *delta* probe. **(B and D)** Vegetal view (B) and side view (D) of an unperturbed embryo. **(C and E)** Vegetal view (C) and side view (E) of a HesC MASO embryo.

DISCUSSION

The GRN model prediction and the evidence

As regions of a GRN approach completion, the levers of logic can be used to generate precise predictions of missing components. As will be described elsewhere, that portion of the sea urchin endomesoderm GRN which pertains to specification and initial differentiation of the skeletogenic micromere domain is now nearly complete, in that it incorporates all regulatory genes expressed specifically in these cells up to the onset of gastrulation (cf. http://sugp.caltech.edu/endomes/). From the GRN analysis came the prediction of the double negative gate shown in Fig. 3.1D (9), and we have now identified the predicted missing component of this gate: the "*repressor of micromere*" gene of the GRN model is *hesC*. In retrospect the exact match between the predicted behavior of *r of mic* and the observed behavior of *hesC* is remarkable. Both the unique pattern of expression of *hesC*, which is not reproduced by any other known gene in this embryo, and the unique effects of preventing its expression, are those required by the double negative gate model in Fig. 3.1D. No additional players are likely to be inserted in the specification gate of Fig. 3.1D since manipulation of either component, *pmar1* overexpression or *hesC* underexpression, suffices to transform the whole embryo into cells specified as skeletogenic mesenchyme: In either perturbation all cells express the regulatory state of the skeletogenic micromere lineage, i.e., they transcribe the *delta*,

alx1, *tbr* and *ets* genes, normally at this stage specific to the skeletogenic micromere lineage (in the pmar1 overexpression they even express terminal differentiation genes, such as sm50, not examined here); they ingress into the blastocoel; and they assume mesenchymal form (Fig. 3.4, 3.5; Supplementary Table 2 in appendix 3; (9, 10).

Though this remains to be authenticated at the *cis*-regulatory level, HesC interactions with the target genes of Fig. 3.1D are likely to be direct, as is the interaction of Pmar1 with the *hesC* regulatory apparatus. First, both genes encode proteins that contain transcriptional repression domains (see Results above for HesC and (9) for Pmar1). Second, the kinetics with which the gate operates almost precludes any intervening steps. In sea urchin embryos at 15 °C it requires about 2-3 h for a regulatory gene to be activated, its product to be translated and transported to the nucleus, and a target gene to respond (20). We show here (Fig. 3.3A) that zygotic expression of *hesC* starts only about two hours after that of *pmar1*, and the zygotic expression of *delta* starts only about 2 h after that of *hesC.*

The double negative gate

The main feature of this mechanism is the use of two repressors in regulatory tandem, and nonlocalized, here ubiquitous, activators to produce a highly confined spatial pattern of gene expression. This is not uncommon; for example, the dorsal-ventral GRN for the early *Drosophila* embryo (21) affords several examples that are in essence similar. The alternative first step is of course highly localized expression of activators. This is a common mechanism of later development, but in the early embryo the boundaries of expression domains are very often controlled negatively, by activation of repressors (1).

In the sea urchin embryo the known maternal regulatory transcripts are all globally distributed (e.g., (12, 15, 16)). Early on, before territorial regulatory states have been established, regional activation of repressors in response to initial anisometric cues is as parsimonious a strategy as regional activation of activators (1). An additional advantage of the double negative gate is that it provides de facto, the active repression of regulatory states outside the correct domain of their expression. Thus it acts as an "exclusion effect" (22), actively ensuring silence of target genes in ectopic locations while at the same time ensuring their expression in correct locations.

Evolutionary implications

The sea urchins are the only echinoderm class which produces an embryo/larva skeleton from a precociously specified micromere lineage. Thus the regulatory apparatus for skeletogenic micromere specification, including the double negative gate, arose in this lineage. An idea proposed earlier is that generation of the larval skeleton evolved as a cooption of the gene regulatory program for the production of the adult calcite skeleton (9, 23). The *hesC-pmar1* double negative gate provides in principle a particularly economical means for highjacking the downstream skeletogenic regulatory machinery. Part of the circuitry is likely to have been already available. The Hes family factors are utilized to repress the *delta* gene across the Bilateria, e.g., in both insect and vertebrate nervous system development (24, 25). Sea urchin HesC repression of *delta* may indicate the inclusion in the co-option process of an ancient widespread "plug-in," i.e., a conserved GRN linkage that is used in multiple, entirely unrelated, developmental contexts (26). Now that the regulatory players are all in hand, and their roles known, it

should be possible to experimentally explore the evolution of the sea urchin skeletogenic specification, by synthetically recreating the regulatory steps that led to its existence.

MATERIALS AND METHODS

Animals, pmar1 mRNA Over-Expression (MOE) and HesC Morpholino Antisense Oligonucleotide (MASO)

Pmar1 was overexpressed, i.e., its expression was forced in all cells of the embryos, by microinjecting *pmar1* mRNA into fertilized eggs. Microinjection solutions were prepared containing 25 ng/µl of pmar1 mRNA and 0.12M KCl.

Translation of HesC transcripts was blocked by microinjection of HesC MASO into fertilized eggs. MASO was synthesized (Gene Tools, Philomath, OR) complementary to the sequence of the first 25 bp of the coding region of *hesC*. The sequence of the oligonucleotide is: 5'-GTTGGTATCCAGATGAAGTAAGCAT–3'. Microinjection solutions were prepared containing 0.12M KCl and 100µM, 250µM or 500µM HesC MASO.

Gametes from *S. purpuratus* were microinjected as described by (27). We aimed at microinjecting a volume of approximately 10 pl. Unperturbed embryos from the same batch were used as control. Living embryos were visualized at chosen developmental time points on an Axioscope 2 Plus microscope (Zeiss, Hallbergmoos, Germany) equipped with the recording device AxioCam MRm (Zeiss).

Quantification of mRNA

RNeasy Micro Kit (74004, Qiagen, Valencia, CA) was used to isolate RNA from samples of ~100 embryos as described in the manufacturer's manual. cDNA was prepared from these samples by reverse transcription-PCR (RT-PCR). The "iScript cDNA Synthesis Kit" (170-8891, Bio-Rad, Hercules, CA) was used for this purpose.

Quantitative PCR (QPCR) was conducted as described by (27), using primer sets designed to produce amplicons of 125-150 bp (for primer sequences see http://sugp.caltech.edu/resources/methods/q-pcr.psp). Amplification reactions were analyzed on an ABI 7900HT Fast Real-Time PCR System using SYBR Green chemistry (iScript One-Step RT-PCR Kit; Bio-Rad, Hercules, CA). Levels of Ubiquitin mRNA are known to remain relatively constant (~220000 molecules per embryo) during the relevant developmental stages (28, 29), and were used as internal standard to determine the levels of mRNA per embryo of all other genes.

Whole Mount In Situ Hybridization (WMISH)

DIG-labeled RNA probes were prepared as described (30). DIG-labeled HesC probe was transcribed from the HesC cDNA clone yde51c06 (CX199264; from a *Strongylocentrotus purpuratus* EST library), kindly provided by Dr. James Coffman. A sense DIG-labeled "Control probe" was transcribed from the same clone, which does not recognize any known or predicted transcript. Dinitrophenol (DNP)-labeled RNA Delta probe was prepared as described by (31) using the same plasmid as used for the DIG-labeled Delta probe of (9).

WMISH was performed using a standard method, as described (32, 33), with minor modifications (Sagar Damle and E.H.D., unpublished data). Hybridization reaction and washes were carried out at 65 ºC. Concentration of probe in hybridization reaction was 1 ng/µl. Antibody incubation was carried out containing a 1000-fold dilution of Anti-Digoxigenin antibody (Fab fragments; Roche; for DIG-labeled probes) or Anti-DNP antiboy-AP (alkaline phosphatase; Mirus; for DNP-labeled probes).

Double-WMISH protocol (from Sagar Damle and E.H.D., unpublished data) was based on the above protocol for WMISH and the double-WMISH protocol described by (34). Steps prior to the hybridization reaction were as described above for WMISH. Hybridization reaction was carried out containing two probes (1 ng/µl each): A DNP-labeled probe and a DIG-labeled probe. Anti-Digoxigenin antibody was used for the first antibody incubation. The first staining reaction (purple) was then carried out as described above for WMISH protocol, with NBT (N-6876, Sigma-Aldrich)/BCIP (B-8503, Sigma-Aldrich). The staining reaction and the antibody activity were stopped as in (34). Anti-DNP antibody-AP was used in the second antibody incubation. The second staining reaction (orange) was similar to the first one, except that INT (1-8377, Sigma-Aldrich)/BCIP was used instead of NBT/BCIP.

ACKNOWLEDGMENTS

We are very grateful to Jina Yun for her enormous help in the Pmar1 screen; to Mary Wohl for carrying out the experiment of Supplementary Table 2 in appendix 3; to Dr. Qiang Tu for help in alignments and phylogenetic analyses; and to Dr. Joel Smith for

very helpful comments on the manuscript. We also thank Dr. Jim Coffman for providing a HesC cDNA clone and Pat Leahy for taking care of the sea urchins used for this work. The first author also thanks Sagar Damle for his extremely useful instruction in performing WMISHs and double-WMISHs, and he would like to apologize to the sea urchins sacrificed for this work.

REFERENCES

1. Davidson, E. (2006) *The Regulatory Genome: Gene Regulatory Networks In Development and Evolution* (Academic Press.
2. 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. j., Schilstra, M. J., Clarke, P. J. C., Arnone, M. I., Rowen, L., Cameron, R. A., McClay, D. R., Hood, L. & Bolouri, H. (2002) *Science* **295,** 1669-1678.
3. Oliveri, P. & Davidson, E. H. (2004) *Current Opinion in Genetics & Development* **14,** 351-360.
4. Levine, M. & Davidson, E. H. (2005) *PNAS* **102,** 4936-42.
5. McClay, D., Peterson, R., Range, R., Winter-Vann, A. & Ferkowicz, M. (2000) *Development* **127,** 5113-5122.
6. Sweet, H., Hodor, P. & Ettensohn, C. (1999) *Development* **126,** 5255-5265.
7. Sweet, H. C., Gehring, M. & Ettensohn, C. A. (2002) *Development* **129,** 1945-1955.
8. Livingston, B. T., Killian, C., Wilt, F., Cameron, A., Landrum, M., Ermolaeva, O., Sapojnikov, V., Maglott, D., Buchanan, A. & Ettensohn, C. (2006) *Developmental Biology* **300,** 335-48.
9. Oliveri, P., Carrick, D. M. & Davidson, E. H. (2002) *Developmental Biology* **246,** 209-228.
10. Ettensohn, C. A., Illies, M. R., Oliveri, P. & De Jong, D. L. (2003) *Development* **130,** 2917-2928.
11. Howard-Ashby, M., Materna, S. C., Brown, C. T., Tu, Q., Oliveri, P., Cameron, R. A. & Davidson, E. H. (2006) *Developmental Biology* **300,** 27-34.
12. Materna, S. C., Howard-Ashby, M., Gray, R. F. & Davidson, E. H. (2006) *Developmental Biology* **300,** 108-120.
13. Rizzo, F., Fernandez-Serra, M., Squarzoni, P., Archimandritis, A. & Arnone, M. I. (2006) *Developmental Biology* **300,** 35-48.
14. Tu, Q., Brown, C. T., Davidson, E. H. & Oliveri, P. (2006) *Developmental Biology* **300,** 49-62.

15. Howard-Ashby, M., Materna, S. C., Brown, C. T., Chen, L., Cameron, R. A. & Davidson, E. H. (2006) *Developmental Biology* **300,** 74-89.
16. Howard-Ashby, M., Materna, S. C., Brown, C. T., Chen, L., Cameron, R. A. & Davidson, E. H. (2006) *Developmental Biology* **300,** 90-107.
17. Kageyama, R., Ohtsuka, T., Hatakeyama, J. & Ohsawa, R. (2005) *Experimental Cell Research* **306,** 343-348.
18. Paroush, Z., Finley, R. L., Kidd, T., Wainwright, S. M., Ingham, P. W., Brent, R. & Ish-Horowicz, D. (1994) *Cell* **79,** 805-815.
19. Grbavec, D. & Stifani, S. (1996) *Biochemical and Biophysical Research Communications* **223,** 701-705.
20. Bolouri, H. & Davidson, E. (2003) *PNAS* **100,** 9371-9376.
21. Stathopoulos, A. & Levine, M. (2005) *Developmental Biology* **280,** 482-93.
22. Oliveri, P. & Davidson, E. H. (2007) *Science* **315,** 1510-1511.
23. Hinman, V., Nguyen, A., Cameron, R. A. & Davidson, E. H. (2003) *Proceedings National Academy of Science USA* **100,** 13356-61.
24. Beatus, P. & Lendahl, U. (1998) *Journal of Neuroscience Research* **54,** 125-136.
25. Gibert, J. & Simpson, P. (2003) *Int J Dev Biol* **47,** 643-651.
26. Davidson, E. H. & Erwin, D. H. (2006) *Science* **311,** 796-800.
27. Rast, J. P., Amore, G., Calestani, C., Livi, C. B., Ransick, A. & Davidson, E. H. (2000) *Developmental Biology* **228,** 270-286.
28. Nemer, M., Rondinelli, E., Infante, D. & Infante, A. (1991) *Developmental Biology* **145,** 255-265.
29. Ransick, A., Rast, J. P., Minokawa, T., Calestani, C. & Davidson, E. H. (2002) *Developmental Biology* **246,** 132-147.
30. Minokawa, T., Rast, J. P., Arenas-Mena, C., Franco, C. B. & Davidson, E. H. (2004) *Gene Expression Patterns* **4,** 449-456.
31. Long, S. & Rebagliati, M. (2002) *Biotechniques* **32,** 494-498.
32. Croce, J., Lhomond, G. & Gache, C. (2003) *Mechanisms of Development* **120,** 561-572.
33. Bradham, C. A. & McClay, D. R. (2006) *Development* **133,** 21-32.
34. Minokawa, T., Wikramanayake, A. & Davidson, E. (2005) *Developmental Biology* **288,** 545-558.

CHAPTER 4

Regulatory Functions in the *delta* R11 *cis*-Regulatory Element

Roger Revilla-i-Domingo and Eric H. Davidson

In preparation for publication – Ongoing research

INTRODUCTION

As described in detail in chapter 2 of this thesis, *delta* is expressed exclusively in the micromere lineage from late cleavage and during blastula stage. Expression of the Delta signal in these cells serves as the spatial cue that triggers the segregation between mesodermal and endodermal cell types from common progenitors (McClay et al., 2000; Sweet et al., 1999; Sweet et al., 2002). In agreement with numerous experiments, the endomesoderm gene regulatory network (GRN) model (Davidson, 2006; Oliveri et al., 2002) makes predictions about the mechanism by which the expression of *delta* is localized to the micromere lineage (Fig. 4.1A; see chapter 3). Oliveri *et al.* (2002) demonstrated that the program of gene expression specific to this lineage is set in train by the *pmar1* gene, a paired homeodomain transcriptional repressor. This gene is expressed in the micromeres as soon as these cells are born, and as we have shown in chapter 3, it acts as a repressor of *hesC*, a transcription factor of the HES family. *hesC* is maternally present in the egg, and it is zygotically expressed everywhere except in the cells of the

micromere lineage, where it is repressed by Pmar1 (see chapter 3). Preventing expression of HesC, either by repressing its zigotic expression through forced global expression of Pmar1, or by blocking translation of its transcripts through a Morpholino-subsituted AntiSense Oligo (MASO), results in *delta* being expressed in all cells of the embryo (chapter 3). Thus, the prediction of the GRN model is that expression of *delta* is activated by factors that are ubiquitously present, and its localization in the micromere lineage depends on repression by HesC in all other cells (Fig. 4.1A). An alternative model, consistent with the same results, is that localization of *delta* in the micromere lineage depends on activation by some gene the expression of which has already been localized in these cells through repression by HesC (Fig. 4.1B). These two mechanisms represent very different architectures of the GRN. Ultimately, the distinction between them can only be made by investigating the regulatory functions executed by the relevant *cis*-regulatory element that controls the expression of *delta*.

In chapter 2 we showed that a 3 kb-long genomic DNA sequence, named R11, is responsible for controlling the expression of *delta* in the micromere lineage. R11 was shown to drive expression of a reporter gene in these cells at the right time. In addition, repressing *hesC* by forced global expression of Pmar1 resulted in R11 driving expression of the reporter everywhere in the embryo, mimicking the effect of the same perturbation on the expression of *delta*. In this work we have dissected R11 and analyzed its regulatory logic. In agreement with the GRN model of Fig. 4.1A, we demonstrate that expression pattern driven by R11 depends on activators that are present in all domains of the embryo, and HesC-dependent repression everywhere except the micromere lineage.

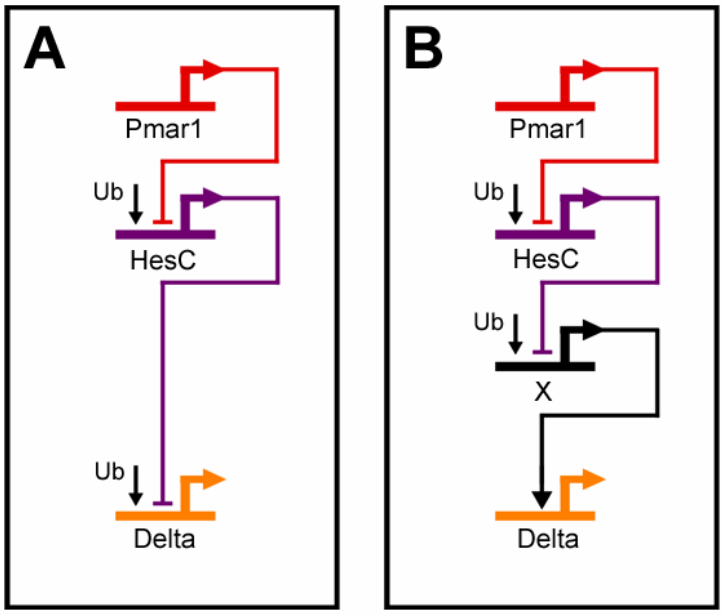


Figure 4.1: Competitive GRN models for control of *delta* expression in the micromere lineage. (A) Model proposed by Oliveri et al. (2002) and chapter 3 of this thesis. *pmar1* is expressed in the micromere descendants, and its product represses zygotic expression of *hesC*. Zygotic expression of *hesC* is global, except in the micromere descendants, where it is repressed by Pmar1. **(B)** Alternative model. An unknown gene "X" is activated by ubiquitous factors, and repressed in non-micromere lineage cells by HesC. Gene "X" in turn activates expression of *delta*.

MATERIALS AND METHODS

Preparation of reporter constructs

R11-GFP reporter construct was prepared described in chapter 2. DNA regions A1, A2, C1 and C2 of R11 were amplified by PCR from the R11-GFP reporter construct. The sequence of the primers used were: A1, left primer, 5'-catgccaacatgaagatgc-3', right primer, 5'-aatacgatggaagagcgtgc-3'; A2, left primer, 5'-ttcaagcagcgtgcaatcac-3', right primer, 5'-aattgaagtccagattagcatgcac-3'; C1, left primer, 5'- gtcattcgtccatctcaggaa-3', right primer, 5'-cacgtctcgtctcgtttaatca-3'; C2, left primer, 5'-tggtttgcattcatgctcata-3', right

primer, 5'-tagcacgcgttttgtgagtg-3'. Amplicons were then fused to the universal *S. purpuratus* expression vector EpGFPII (Cameron et al., 2004) by means of a PCR fusion-based approach (Hobert, 2002). The vector EpGFPII contains the region around the start of transcription of the *endo16* gene (from -117 to +20). The activity of this basal promoter element has been described in detail elsewhere (Yuh and Davidson, 1996; Yuh et al., 1996; Yuh et al., 1998). The EpGFPII expression vector also contains the coding sequence of the GFP protein. All reporter constructs cloned into the pGEM-T Easy vector. All constructs were amplified by PCR and cleans with PCR purification kit (QIAGEN) for microinjection. A construct containing only the basal promoter and the GFP protein sequence was also prepared. This construct was called BP (Basal Promoter), and was used to estimate the level of background expression.

Animals and microinjection of reporter constructs

Microinjection solutions were prepared as in chapter 2. Gametes from *S. purpuratus* were obtained and microinjected as described (Rast et al., 2000). This protocol is essentially based on the original protocol (McMahon et al., 1985). Modifications were as described in chapter 2.

***pmar1* mRNA overexpression (MOE) and HesC Morpholino-substituted AntiSense Oligonucleotide (MASO)**

Pmar1 was overexpressed, i.e., its expression was forced in all cells of the embryos, by microinjecting *pmar1* mRNA into fertilized eggs. Microinjection solutions were prepared as described in chapter 3.

Translation of *hesC* transcripts was blocked by microinjection of *hesC* MASO into fertilized eggs. This was done as in chapter 3.

Quantification of the normalized activity of reporter constructs

The normalized activity of reporter constructs was measured as described in chapter 2, with some modifications. For each batch of eggs, two or three samples were cultured independently. Measurements from each sample were obtained independently, and then averaged. Genomic DNA and total mRNA was extracted from samples of ~60-100 microinjected embryos using the "AllPrep DNA/RNA Mini Kit" (QIAGEN) as described in the manufacturer's manual. "iScript cDNA Synthesis Kit" (170-8891, Bio-Rad, Hercules, CA) was used for the reverse transcription-PCR reaction (RT-PCR). Quantitative PCR (QPCR) was conducted as described (Rast et al., 2000), using primer sets designed to produce amplicons of 125-150 bp (for primer sequences see http://sugp.caltech.edu/resources/methods/q-pcr.psp). Amplification reactions were analyzed on an ABI 7900HT Fast Real-Time PCR System using SYBR Green chemistry (iScript One-Step RT-PCR Kit; Bio-Rad, Hercules, CA). *foxA* and *nodal* primers were used as genomic DNA standards. Data from each cDNA sample were normalized against *18s* rRNA levels, which are known to remain relatively constant during the developmental stages used (Ransick et al., 2002).

RESULTS

Dissection of R11 *cis*-regulatory element

If, as predicted by the GRN model of Fig. 4.1A, *delta* expression is activated by factors that are ubiquitously present in the embryo, portions of R11 sequence should exist that drive expression of a reporter everywhere in the embryo, and other portions should also exist that repress expression everywhere except in the micromere lineage. If instead, the alternative model of Fig. 4.1B were true it should not be possible to find, within R11, subelements that drive expression of the reporter outside of the micromere lineage. We first dissected R11 into several overlapping subelements of 600-1500 bp in length (not shown). GFP reporter constructs were prepared from these subelements, and they were microinjected into embryos. GFP expression was then observed at mesenchyme blastula stage. At this stage GFP driven by R11 is still clearly detectable (see chapter 2), and the cells of the micromere lineage are morphologically easily distinguishable from the cells of the rest of the embryo. All descendants of the micromeres have ingressed into the blastocoel as Primary Mesenchyme Cells (PMCs), while the rest of the cells of the embryo form the blastula wall at this time (Fig. 4.2 C-E). R11-GFP construct was used for comparison to all other constructs, and the GFP reporter without any *cis*-regulatory element (named Basal Promoter, or BP) was used as a control for background expression. Fig. 4.2A indicates the locations where GFP expression was observed for the relevant constructs. In interpreting these data we have to bear in mind that exogenous DNA is incorporated in mosaic fashion in microinjected sea urchin embryos (McMahon et al., 1985). As a consequence, each of the microinjected embryos will have one or a few

clones of cells that contain exogenous DNA, and that therefore have the possibility to express the reporter gene.

A

	TOTAL	Expressing		In PMCs/Ingressed cells			In Blastula Wall Cells		
			%Total		% Expr	% Total		%Expr	% Total
R11	438	239	**55**	209	**87**	**48**	41	**17**	**9**
A1	151	74	**49**	55	**74**	**36**	23	**31**	**15**
C1	136	70	**51**	60	**86**	**44**	14	**20**	**10**
A2	150	111	**74**	28	**25**	**19**	94	**85**	**63**
C2	107	81	**76**	45	**56**	**42**	51	**63**	**48**
R11+Control MASO	74	44	**59**	40	**91**	**54**	7	**16**	**9**
R11+HesC MASO	128	113	**88**	100	**88**	**78**	56	**50**	**44**
BP	142	28	**20**	7	**25**	**5**	22	**79**	**15**

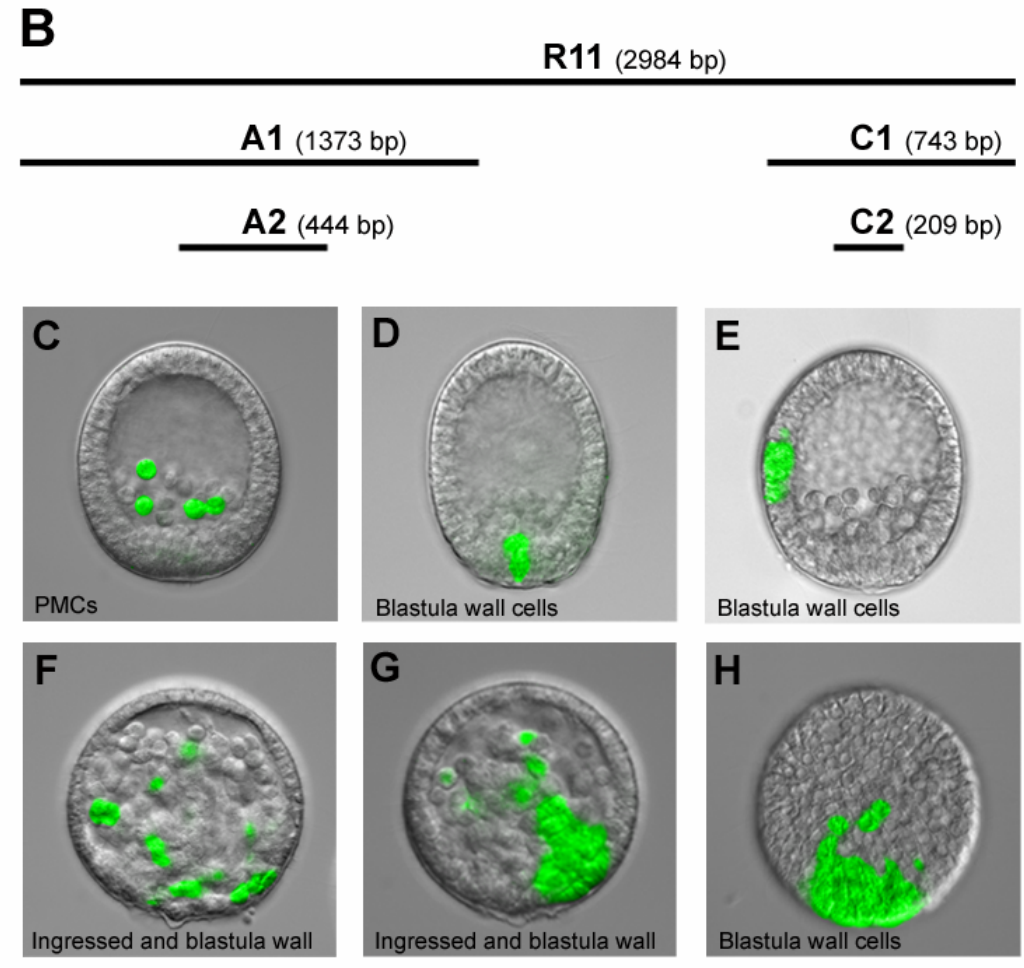


Figure 4.2: (A) Table showing locations of GFP expression in embryos microinjected with GFP reporter constructs. '% Total' means percentage of embryos with respect to the 'TOTAL' number

of embryos observed. '% Expr' means percentage of embryos with respect to the number of 'Expressing' embryos. 'Ingressed cells' are cells that have ingressed into the blastocoel at the time of observation. A1, C1, A2, C2 are subelements of R11 as indicated in (B). BP means 'Basal promoter' and indicates background expression by a GFP reporter construct that contains no regulatory DNA. **(B)** Map of the subelements of R11 from which GFP reporter activity is indicated in (A). **(C)** Example of a normal embryo expressing GFP in PMCs. **(D and E)** Two examples of normal embryos expressing GFP in blastula wall cells. **(F and G)** Two examples of HesC MASO embryos expressing GFP in ingressed cells and in blastuala wall cells. **(H)** An example of a HesC MASO embryo expressing GFP in blastula wall cells.

Two separate subelements, named A1 and C1 (Fig. 4.2B), were found to drive expression of GFP mainly in the PMCs, similarly to R11. As shown in Fig. 4.2A, about 50% of the embryos microinjected with R11-GFP, A1-GFP or C1-GFP showed expression. Of these, most showed expression in the PMCs (Fig. 4.2C; 87%, 74% and 86% for R11-GFP, A1-GFP and C1-GFP, respectively). Expression of GFP in cells other than the PMCs (Fig. 4.2D, E) was observed in 17%, 31% and 20% of the expressing embryos microinjected with R11-GFP, A1-GFP and C1-GFP, respectively, which represented 9%, 15% and 10% of the total number of embryos observed. These numbers of embryos expressing in cells other than PMCs were not considered significant, because approximately the same percentage of the total number of embryos microinjected with BP-GFP showed GFP expression in these cells (15%; Fig. 4.2A).

The sequence between A1 and C1 was not found to drive any GFP expression, or to significantly affect the expression of either A1 or C1 (data not shown). Its *cis*-regulatory function was therefore not further investigated.

Subelements A1 and C1 were then each dissected into overlapping fragments that ranged in size between 100 and 700 bp (not shown). Two of these subelements, named

A2 and C2 (Fig. 4.2B), were found to activate GFP expression in a nonlocalized manner. Both drove GFP expression in a significantly higher percentage of embryos than A1 or C1 (A2 and C2 drove expression in 74% and 76% of the embryos, respectively, while A1 and C1 only in 49% and 51%, respectively). This was clearly due to a significantly higher number of embryos expressing GFP in cells other than the PMCs: For A2, 25% of the embryos expressing GFP did so in PMCs and 85% elsewhere in the embryo; for C2, 56% of the embryos expressing GFP did so in PMCs and 63% elsewhere in the embryo. The expression of GFP in cells other than PMCs was clearly above background. When calculated relative to the total number of embryos observed, the number of embryos expressing GFP in non-PMCs in the samples microinjected with A2-GFP (63%) or C2-GFP (48%) was much higher than in the samples microinjected with BP-GFP (15%). It is important to add that among the embryos expressing GFP in cells other than PMCs, no bias was seen between vegetal plate (Fig. 4.2D) or any particular position in the ectoderm (Fig. 4.2E; data not shown). It is worth noting also that any examined fragments of A1 that did not contain A2, or fragments of C1 that did not contain C2 did not drive GFP expression in any higher number of embryos than background (data not shown).

These results indicated that the factors that activate expression of the reporter gene through A2 and C2 are not localized to the micromere lineage, and they are consistent with these factors being present in all cells of the embryo. These results also indicate that a repressor must operate through the sequence of A1 outside of A2, and through the sequence of C1 outside of C2, to repress expression of the reporter gene in cells that are not in the micromere lineage. This evidence is consistent with the GRN model of Fig. 4.1A and rules out the alternative model of Fig. 4.1B.

In what follows, we provide compelling evidence that the repressor that operates through the sequence of A1 and C1 is HesC.

Effect of perturbing *hesC* expression on the expression driven by R11 and its subelements A1 and A2

In chapter 2 we showed that when *hesC* is repressed by forcing global expression of Pmar1, R11 drives expression of a reporter gene everywhere in the embryo. In the following we confirm that this effect was indeed due to the elimination of *hesC* expression and not due to any other effect of the *pmar1* mRNA overexpression (MOE) perturbation. To do this we blocked translation of *hesC* transcripts through microinjection of a Morpholino-substituted AntiSense Oligo (MASO). This perturbation has been shown not to affect the expression of *pmar1* transcripts (see chapter 3).

The construct R11-GFP was microinjected into embryos either with a Control MASO or with a HesC MASO. The location of GFP expression was then examined at mesenchyme blastula stage. In analyzing the results of this experiment it is important to note that in HesC MASO embryos the descendants of the micromere lineage are not the only cells that ingress into the blastocoel. As a consequence, the cells of the micromere lineage are morphologically indistinguishable from all other cells that have ingressed at this time (see chapter 3). However, they can indeed be distinguished from the cells that still remain at the blastula wall of the embryo (Fig. 4.2 F-H).

As shown in Fig. 4.2A, the expression of GFP in embryos microinjected with R11-GFP together with Control MASO was very similar to that in embryos microinjected with R11-GFP alone. Over half of these embryos expressed GFP, and among these,

almost all did so in PMCs (about 90%). Less than 15% of the total number of embryos observed expressed GFP in cells of the blastula wall, which as discussed above, was not considered to be significant.

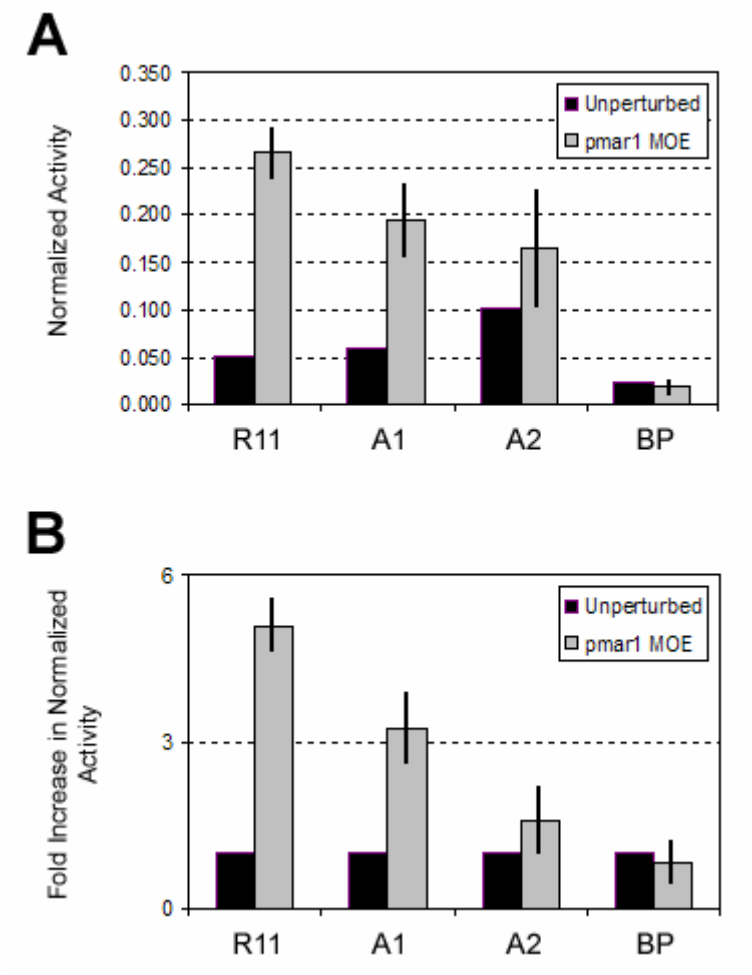


Figure 4.3: Effect of *pmar1* MOE on the normalized activity of R11 and its subelements. (A) Normalized activity (*gfp* mRNA molecules per *gfp* DNA molecule) of GFP reporter constructs in unperturbed embryos (black columns) and in *pmar1* MOE embryos (grey columns). Each column represents an average from four to six independent cultures of embryos. The *pmar1* MOE measurements were normalized to the average of the unperturbed samples (therefore unperturbed samples have no error bars). Error bars represent the standard deviation from the average. **(B)** Average fold increase in the normalized activity of the GFP reporter constructs in *pmar1* MOE

embryos with respect to unperturbed embryos. Columns representing unperturbed samples are here set to "1" by definition. Error bars represent the standard deviation from the average.

The expression of GFP in embryos microinjected with R11-GFP together with HesC MASO was very different. Almost 90% of these embryo expressed GFP, and large patches of expressing cells were now observed (Fig. 4.2 G and H). These observations are of course consistent with GFP being expressed in any cell that receives the exogenous DNA. Many of the embryos expressing GFP did so in ingressed cells (88%), but also a significantly high number did so in cells of the blastula wall (50%). As implied by the numbers, many of the embryos expressed GFP both in ingressed cells and in cells of the blastula wall (Fig. 4.2 F and G). Since many of the ingressed cells are not micromere descendants, the number of embryos expressing GFP in cells of the blastula wall is a clear underestimate of the number of embryos expressing GFP in cells that are not micromere descendants. Overall, these results indicate that in HesC MASO embryos GFP expression driven by R11 is not localized to the cells of the micromere lineage, similarly to what was reported for *pmar1* MOE embryos. This means that HesC is required for R11 to be able to localize the expression of the reporter in the micromere lineage. If expression of HesC is prevented, either by HesC MASO or *pmar1* MOE, repression in non-micromere lineage cells is disrupted.

If, as predicted by the GRN model (Fig. 4.1A), HesC is indeed the repressor of R11, we should expect it to operate through A1 and C1. Should this be the case, repressing *hesC* expression by *pmar1* MOE should result in A1 and C1 driving expression of the reporter gene everywhere in the embryo, but should not affect the ubiquitous expression of A2 or C2. We tested this for A1 and A2. Due to the morphology

of *pmar1* MOE embryos, GFP fluorescence could not be used to show the absence of an effect of this perturbation on the expression driven by A2. However, quantification of the normalized activity (*gfp* mRNA molecules per *gfp* DNA molecule) driven by each subelement in perturbed and unperturbed embryos should reveal any effect of the perturbation or the absence of it (see chapter 2).

Fig. 4.3A shows the average normalized activity measured for A1-GFP and A2-GFP in unperturbed embryos and in *pmar1* MOE embryos. The normalized activities for R11-GFP and BP-GFP are also shown for comparison. Fig. 4.3B shows the average fold increase in the normalized activity, calculated from Fig. 4.3A. As expected, the normalized activity driven by R11 and A1 significantly increased upon repression of *hesC* by *pmar1* MOE. Instead, the normalized activity driven by A2 and BP did not change significantly, if at all. This indicates that the repression function of A1 requires HesC for its operation, and provides strong support for the GRN model of Fig. 4.1A.

DISCUSSION

The endomesoderm GRN model makes predictions about the *cis*-regulatory interactions that are responsible for the expression of the relevant genes at the right places and times to serve their role in the specification process. The *cis*-regulatory element that drives the expression of the *delta* gene in the micromere lineage is particularly important, since it is responsible for the transcriptional expression of the spatial information that sets in action the specification of several mesodermal fates. The prediction is that the expression of

delta in the micromere lineage depends on activating factors that are ubiquitously present and on repression by HesC (Fig. 4.1A) (Davidson, 2006; Oliveri et al., 2002) (chapter 3).

This work demonstrates that indeed activators of *delta* are present everywhere in the embryo. This is implied by the fact that two subelements of R11, A2 and C2, drive expression of a reporter gene in every domain of the embryo, in addition to the micromere lineage (Fig. 4.2A). Importantly, this result rules out an alternative model in which expression of *delta* in the micromere lineage is driven by a gene the expression of which is localized in these cells (Fig. 4.1B).

Our work also demonstrates that R11 mediates HesC-dependent repression in non-micromere lineage cells. While R11 normally drives expression of a reporter exclusively in the micromere lineage (see chapter 2; also reproduced in Fig. 4.2A), we have shown that blocking translation of *hesC* results in expansion of its expression domain to other cells of the embryo. Giving support to this result, repression of *hesC* by *pmar1* MOE also results in R11 driving expression of the reporter in all cells of the embryo, and in a significant increase in its normalized activity (chapter 2; Fig. 4.3).

Our work does not demonstrate, but very strong supports, that HesC directly operates through the sequence of R11. We have shown that there are subelements within R11 that repress activation of expression in non-micromere lineage cells: The sequence of A1 outside of A2, and the sequence of C1 outside of C2 (Fig. 4.2) operate this function. The fact that *pmar1* MOE increases the normalized activity of A1 while it does not affect that of A2 (Fig. 4.3), strongly supports that HesC directly operates through the sequence of A1 outside of A2 to repress expression in non-micromere lineage cells. Alternatives to this explanation are either that HesC activates the repressor that operates

through the sequence of A1, or that HesC is a repressor of a repressor of the repressor that operates through this sequence. Both are extremely unlikely. First, HesC does not contain any known transactivating domain (data not shown), and it contains the WRPW domain at its terminal end (see chapter 3; which is used by many known transcription factors to recruit TLE/Grg/Groucho to mediate transcriptional repression (Grbavec and Stifani, 1996; Paroush et al., 1994)). Second, almost all transcription factors of the HES family are known to operate as transcriptional repressors (Kageyama et al., 2005). And third, the time lag between the start of zygotic expression of *hesC* and that of *delta* is about 2 h (see chapter 3), which makes essentially impossible the presence of two intervening steps between *hesC* and *delta* (Bolouri and Davidson, 2003).

The means by which HesC represses R11 remain to be elucidated, and this is the subject of ongoing work. Transcription factors of the HES family have been reported to bind to N boxes (CACNAG or CACG[A/C/T]G) to mediate transcriptional repression in several organisms, including vertabrates and insects (Kageyama et al., 2005; Ledent and Vervoort, 2001). However, HesC is not likely to bind to these sequences to mediate repression of R11. No N boxes are found in the sequence of A1 outside of A2, and mutation of all the N boxes in the sequence of C1 has no effect on the expression driven by this subelement (data not shown). A very likely possibility is that HesC recognizes sequences other than N boxes. Most HES factors have a Proline in the middle of their basic domain, which has been suggested to be important for their binding affinity to N boxes (Iso et al., 2003; Kageyama et al., 2005; Tietze et al., 1992; Wainwright and Ish-Horowicz, 1992). HesC has a Histidine in the position of this characteristic Proline. Notably, no other HES factor known in humans, mouse, *Drosophila*, *C. elegans* or sea

urchin contains a Histidine in the same position of the basic domain (data not shown). Another possibility is that the interaction between HesC and R11 is mediated by a partner or a complex, which directly binds to R11 and recruits HesC.

The architecture of a GRN is determined by the regulatory functions executed at the participating CRMs. This work shows that dissection of a CRM can be used to reveal its regulatory logic and thereby test competitive GRN architectures that cannot be distinguished by perturbation analysis alone.

AKNOWLEDGEMENTS

We are very thankful to Julie Hahn for her help with making mutation constructs. We also thank Dr. Jongmin Nam for extremely useful discussions about this work and constructive criticism, and Dr. Qiang Tu for help with alignments and phylogenetic analyses of HesC.

REFERENCES

Bolouri, H., and Davidson, E. (2003). Transcriptional regulatory cascades in development: Initial rates, not steady states, determine expression kinetics. *PNAS* **100,** 9371-9376.

Cameron, R. A., Oliveri, P., Wyllie, J., and Davidson, E. H. (2004). cis-Regulatory activity of randomly chosen genomic fragments from the sea urchin. *Gene Expression Patterns* **4,** 205-213.

Davidson, E. (2006). "The Regulatory Genome: Gene Regulatory Networks In Development and Evolution." Academic Press,

Grbavec, D., and Stifani, S. (1996). Molecular Interaction between TLE1 and the Carboxyl-Terminal Domain of HES-1 Containing the WRPW Motif. *Biochemical and Biophysical Research Communications* **223,** 701-705.

Hobert, O. (2002). PCR fusion-based approach to create reporter gene constructs for expression analysis in transgenic C. elegans. *BioTechniques* **32,** 1-3.

Iso, T., Kedes, L., and Hamamori, Y. (2003). HES and HERP families: Multiple effectors of the notch signaling pathway. *Journal of Cellular Physiology* **194,** 237-255.

Kageyama, R., Ohtsuka, T., Hatakeyama, J., and Ohsawa, R. (2005). Roles of bHLH genes in neural stem cell differentiation. *Experimental Cell Research* **306,** 343-348.

Ledent, V., and Vervoort, M. (2001). The Basic Helix-Loop-Helix Protein Family: Comparative Genomics and Phylogenetic Analysis. *Genome Res.* **11,** 754-770.

McClay, D., Peterson, R., Range, R., Winter-Vann, A., and Ferkowicz, M. (2000). A micromere induction signal is activated by beta-catenin and acts through notch to initiate specification of secondary mesenchyme cells in the sea urchin embryo. *Development* **127,** 5113-5122.

McMahon, A., Flytzanis, C., Hough-Evans, B., Katula, K., Britten, R., and Davidson, E. H. (1985). Introduction of cloned DNA into sea urchin egg cytoplasm: replication and persistence during embryogenesis. *Developmental Biology* **108,** 420-430.

Oliveri, P., Carrick, D. M., and Davidson, E. H. (2002). A Regulatory Gene Network That Directs Micromere Specification in the Sea Urchin Embryo. *Developmental Biology* **246,** 209-228.

Paroush, Z., Finley, R. L., Kidd, T., Wainwright, S. M., Ingham, P. W., Brent, R., and Ish-Horowicz, D. (1994). Groucho is required for Drosophila neurogenesis, segmentation, and sex determination and interacts directly with hairy-related bHLH proteins. *Cell* **79,** 805-815.

Ransick, A., Rast, J. P., Minokawa, T., Calestani, C., and Davidson, E. H. (2002). New Early Zygotic Regulators Expressed in Endomesoderm of Sea Urchin Embryos Discovered by Differential Array Hybridization. *Developmental Biology* **246,** 132-147.

Rast, J. P., Amore, G., Calestani, C., Livi, C. B., Ransick, A., and Davidson, E. H. (2000). Recovery of Developmentally Defined Gene Sets from High-Density cDNA Macroarrays. *Developmental Biology* **228,** 270-286.

Sweet, H., Hodor, P., and Ettensohn, C. (1999). The role of micromere signaling in Notch activation and mesoderm specification during sea urchin embryogenesis. *Development* **126,** 5255-5265.

Sweet, H. C., Gehring, M., and Ettensohn, C. A. (2002). LvDelta is a mesoderm-inducing signal in the sea urchin embryo and can endow blastomeres with organizer-like properties. *Development* **129,** 1945-1955.

Tietze, K., Oellers, N., and Knust, E. (1992). Enhancer of SplitD, a Dominant Mutation of Drosophila, and its Use in the Study of Functional Domains of a Helix-Loop-Helix Protein. *PNAS* **89,** 6152-6156.

Wainwright, S. M., and Ish-Horowicz, D. (1992). Point mutations in the Drosophila hairy gene demonstrate in vivo requirements for basic, helix-loop-helix, and WRPW domains. *Mol. Cell. Biol.* **12,** 2475-2483.

Yuh, C., and Davidson, E. (1996). Modular cis-regulatory organization of Endo16, a gut-specific gene of the sea urchin embryo. *Development* **122,** 1069-1082.

Yuh, C., Moore, J., and Davidson, E. (1996). Quantitative functional interrelations within the cis-regulatory system of the S. purpuratus Endo16 gene. *Development* **122,** 4045-4056.

Yuh, C.-H., Bolouri, H., and Davidson, E. H. (1998). Genomic Cis-Regulatory Logic: Experimental and Computational Analysis of a Sea Urchin Gene. *Science* **279,** 1896-1902.

CONCLUSIONS

The first published version of the endomesoderm gene regulatory network (GRN) model was based on expression data for many genes, and on an extensive perturbation analysis (see appendices 1 and 2). Implicit in the model were predictions about the genomic regulatory functions that control the expression of the participating genes. The model for the genomic control mechanism for the localization of *delta* expression in the micromere lineage is particularly important, since the Delta signal serves as the spatial cue for a critical specification event: the segregation of mesodermal and endodermal cell types from common progenitors. Two important predictions were implicit in the model: 1) That there exist a gene, named *r of mic*, the function of which is to repress micromere lineage specification everywhere except in the micromere lineage (where this gene is repressed by Pmar1); 2) that expression of *delta* in the micromere lineage depends on ubiquitous activators and on repression by R of mic. The results presented in this thesis validate these two predictions of the GRN model (see Figure 5.1).

The work described in chapter 2 recovered the *cis*-regulatory module (CRM) that is responsible for the localization of *delta* expression in the micromere lineage. This CRM was named R11. It consists of a 3 kb-long piece of genomic DNA located 28 kb downstream of the *delta* translation start site. In normal embryos, R11 accurately drives expression of a reporter construct exclusively in the micromere lineage, while in embryos globally expressing *pmar1* mRNA, it becomes capable of causing expression in any cells of the embryo. This behavior perfectly reproduces the response of the endogenous *delta* gene to the same perturbation. This work confirmed that localization of *delta* expression

in the micromere lineage, and expansion of its expression to every cell in response to global Pmar1 expression are transcriptionally controlled. This indicated that the predicted *r of mic* indeed had to exist and that it had to be a transcription factor.

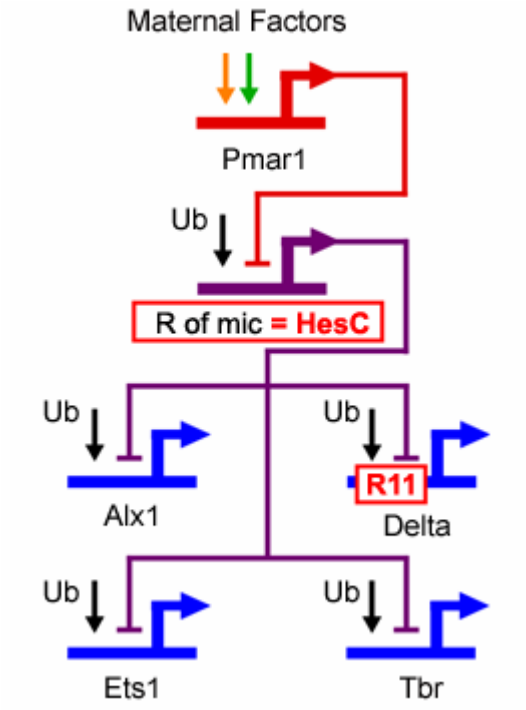


Figure 5.1: Validation of predictions of the GRN model. The diagram shows the GRN model for early micromere lineage specification. Red boxes represent predictions of the model. The predicted R of mic is HesC. The piece of genomic DNA that mediates the function of HesC to repress expression of *delta* in the micromere lineage is R11. Note that the interaction between HesC and R11 has not yet been demonstrated to be direct.

The work described in chapter 3 identified the transcription factor that plays the role of *r of mic*. This gene is *hesC*. In all aspects tested, this gene precisely conforms to the characteristics predicted for *r of mic*. It is transcriptionally repressed by Pmar1, and it is zygotically expressed at the right time and place. Most importantly, functional analyses indicate that HesC is responsible for repressing micromere lineage specification in all

cells that are not in the micromere lineage. Blocking translation of *hesC* transcripts transforms the whole embryo into cells specified as skeletogenic mesenchyme: All cells express the regulatory state of the skeletogenic micromere lineage, i.e., they transcribe the *delta, alx1*, *tbr* and *ets* genes, normally at this stage specific to the skeletogenic micromere lineage; they ingress into the blastocoel; and they assume mesenchymal form.

The work described in chapter 4 shows that the predicted regulatory functions are executed by the R11 delta CRM: The activators of R11 are ubiquitously present, and repression in non-micromere lineage cells depends on HesC. Subelements within R11 exist that can drive the expression of a reporter in all domains of the embryo, while other subelements repress the ectopic expression. That the repression function of R11 depends on HesC is shown by the fact that blocking translation of *hesC* transcripts disrupts the repression of ectopic expression.

The work presented in chapter 4 strongly supports, but does not demonstrate, that the interaction between HesC and *delta* is direct. If HesC acts as a repressor (which is implied by its sequence) it can only interact with R11 either directly or through at least two intervening steps. The latter is extremely unlikely, because zygotic expression of *delta* starts only about 2 h after that of *hesC* (see chapter 3). An ultimate demonstration for a direct interaction between HesC and *delta* would require to identify binding sites for HesC on the sequence of R11, and to show that mutation of these sites results in R11 driving the expression of a reporter outside the micromere lineage domain. This is the subject of ongoing work.

Overall, the results of this thesis demonstrate that the predicted *r of mic* exists, it is *hesC*, and that there exists a piece of genomic DNA, namely R11, that mediates the

function of HesC by executing the regulatory functions predicted by the GRN model. The significance of these findings is remarkable. The identification of *hesC* confirms the use of a double negative gate for the translation of the spatial information in the egg into the institution of the micromere lineage domain. The main aspect of this mechanism is the use of two repressors in regulatory tandem, here *pmar1* and *hesC,* and non-localized activators to generate a highly localized spatial pattern of gene expression. The use of this mechanism for early specification events is not uncommon: for example, a similar GRN architecture is used during the specification of the dorsal-ventral axis in the early *Drosophila* embryo (Stathopoulos and Levine, 2005). The alternative is of course highly localized activation of activators. This is a more parsimonious mechanism later in development, but during the first steps of specification, before territorial regulatory states have been established, regional activation of repressors in response to initial anisotropic cues is as parsimonious a strategy as regional activation of activators (Davidson, 2006). An additional advantage of the double negative mechanism is that it provides active repression of regulatory states outside the correct domain of their expression, thereby ensuring silence of target genes in ectopic locations. The identification of R11 and its functional dissection illustrate that the double negative architecture of the GRN is determined by the regulatory functions executed by the relevant pieces of genomic DNA. It is this set of regulatory functions that ultimately determines how the spatial information in the egg is translated into the institution of the micromere lineage domain, and the correct localization of *delta* expression.

In retrospect, these results illustrate that the approach undertaken in unraveling the endomesoderm GRN is remarkably powerful. Expression data and perturbation

analyses, when applied at the system level, provide enough logic constraints to make predictions about the regulatory functions executed by the participating CRMs. In turn, *cis*-regulatory analysis feeds back into the model by validating these predictions. As such analysis is extended to all portions of the GRN, there emerges an explanatory structure that will explicitly show how the genome processes information and thereby controls specification and development.

REFERENCES

Davidson, E. (2006). "The Regulatory Genome: Gene Regulatory Networks In Development and Evolution." Academic Press,

Stathopoulos, A., and Levine, M. (2005). Localized repressors delineate the neurogenic ectoderm in the early Drosophila embryo. *Developmental Biology* **280,** 482-93.

APPENDIX 1

Developmental Biology **246,** 162–190 (2002)
doi:10.1006/dbio.2002.0635, available online at http://www.idealibrary.com on IDEAL®

A Provisional Regulatory Gene Network for Specification of Endomesoderm in the Sea Urchin Embryo

Eric H. Davidson,*,[1] Jonathan P. Rast,* Paola Oliveri,* Andrew Ransick,* Cristina Calestani,* Chiou-Hwa Yuh,* Takuya Minokawa,* Gabriele Amore,* Veronica Hinman,* César Arenas-Mena,* Ochan Otim,* C. Titus Brown,* Carolina B. Livi,* Pei Yun Lee,* Roger Revilla,* Maria J. Schilstra,† Peter J. C. Clarke,† Alistair G. Rust,†,[2] Zhengjun Pan,†,[3] Maria I. Arnone,‡ Lee Rowen,§ R. Andrew Cameron,* David R. McClay,¶ Leroy Hood,§ and Hamid Bolouri†

**Division of Biology, California Institute of Technology, Pasadena, California 91125; †Science and Technology Research Centre, University of Hertfordshire, Hertfordshire AL10 9AB, United Kingdom; ‡Stazione Zoologica, Anton Dohrn 80121 Naples, Italy; §Institute for Systems Biology, Seattle, Washington 98105; and ¶Department of Zoology, Duke University, Durham, North Carolina 27708-0325*

We present the current form of a provisional DNA sequence-based regulatory gene network that explains in outline how endomesodermal specification in the sea urchin embryo is controlled. The model of the network is in a continuous process of revision and growth as new genes are added and new experimental results become available; see http://www.its.caltech.edu/~mirsky/endomeso.htm (End-mes Gene Network Update) for the latest version. The network contains over 40 genes at present, many newly uncovered in the course of this work, and most encoding DNA-binding transcriptional regulatory factors. The architecture of the network was approached initially by construction of a logic model that integrated the extensive experimental evidence now available on endomesoderm specification. The internal linkages between genes in the network have been determined functionally, by measurement of the effects of regulatory perturbations on the expression of all relevant genes in the network. Five kinds of perturbation have been applied: (1) use of morpholino antisense oligonucleotides targeted to many of the key regulatory genes in the network; (2) transformation of other regulatory factors into dominant repressors by construction of Engrailed repressor domain fusions; (3) ectopic expression of given regulatory factors, from genetic expression constructs and from injected mRNAs; (4) blockade of the β-catenin/Tcf pathway by introduction of mRNA encoding the intracellular domain of cadherin; and (5) blockade of the Notch signaling pathway by introduction of mRNA encoding the extracellular domain of the Notch receptor. The network model predicts the *cis*-regulatory inputs that link each gene into the network. Therefore, its architecture is testable by *cis*-regulatory analysis. *Strongylocentrotus purpuratus* and *Lytechinus variegatus* genomic BAC recombinants that include a large number of the genes in the network have been sequenced and annotated. Tests of the *cis*-regulatory predictions of the model are greatly facilitated by interspecific computational sequence comparison, which affords a rapid identification of likely *cis*-regulatory elements in advance of experimental analysis. The network specifies genomically encoded regulatory processes between early cleavage and gastrula stages. These control the specification of the micromere lineage and of the initial veg_2 endomesodermal domain; the blastula-stage separation of the central veg_2 mesodermal domain (i.e., the secondary mesenchyme progenitor field) from the peripheral veg_2 endodermal domain; the stabilization of specification state within these domains; and activation of some downstream differentiation genes. Each of the temporal–spatial phases of specification is represented in a subelement of the network model, that treats regulatory events within the relevant embryonic nuclei at particular stages.

Key Words: **gene network; sea urchin embryo; gene regulation.**

---

[1] To whom correspondence should be addressed. Fax: (626) 793-3047. E-mail: davidson@caltech.edu.
[2] Present address: European Bioinformatics Institute, Wellcome Trust Genome Campus, Hinxton, Cambridgeshire CB101 1SD, United Kingdom.
[3] Present address: Altera European Technology Centre, Holmers Farm Way, High Wycombe, Buckinghamshire HP12 4XF, United Kingdom.

0012-1606/02 $35.00

INTRODUCTION

Almost a century has passed since Theodor Boveri's realization that development of an embryo is controlled by the genomes of the embryonic cells (Boveri, 1902, 1918). The experiments that led to his controversial conclusion were carried out on sea urchin embryos. The genomic control network, which is the subject of this paper, underlies the process of endomesoderm formation in the embryo of *Strongylocentrotus purpuratus*. Its foundations are built out of the rich store of knowledge now available about how endomesoderm specification works in sea urchin embryos. Most of this knowledge has accumulated within the last decade from research in molecular and experimental embryology.

Specification is the process by which cells in a given spatial and temporal domain of an embryo obtain their developmental identity. Once specified, they contribute to a particular part of the developing embryo and express a particular set of genes. In mechanistic terms, specification consists of the set of events leading to the installation and ultimately the stabilization of given gene regulatory states. That is, the result of specification is the expression of unique sets of genes encoding sequence-specific transcription factors which directly control the program of gene expression that the cells end up executing. As the regulatory bases of developmental processes are revealed, results from many different systems have converged on an alarming fact of genomic life: specification never depends only on one or a few "master genes," but always requires large networks of functionally linked regulatory genes. Some networks are deeper and more extensive than others, and some genes are more important than others (at least under given experimental paradigms), but a system that includes many interacting regulatory genes underlies every developmental specification event (reviewed by Davidson, 2001). Solving such large gene regulatory systems requires a cross between the methods, technologies, attitudes, and instrumentation developed in genomics and computational molecular biology, and the paradigms of hard-core experimental regulatory developmental biology. We have constructed a strategy of this mixed nature by which to assemble the relevant embryonic regulatory relationships into an understandable network. The main form of evidence through which the connections of the network have been revealed is a large-scale perturbation analysis, in which the expression of many different genes was interfered with and the effect on many other genes measured. This paper is our first progress report on the outcome of this analysis. It is the product of a large collaborative effort which has many authors, all of whom contributed importantly, and which has involved several laboratories. The reward is that we can now understand the encoded control logic of endomesoderm specification much more deeply and explicitly than we could have envisioned earlier.

Endomesoderm Specification

By early in the sixth cleavage cycle, the cell lineages of typical sea urchin embryos have been segregated into a canonical set of territories, each of which is destined to give rise to certain distinct cell types (Hörstadius, 1939; Cameron *et al.*, 1987, 1991), and in each of which a distinct set of genes is already running (reviewed by Davidson *et al.*, 1998; Davidson, 2001). The upper or animal pole half of the embryo now consists of 32 blastomeres that produce only the cell types ultimately found in the oral and aboral ectoderm. The lower half consists of the 8 cells of the veg_1 ring bounded on top by the equatorial third cleavage plane and below by their 8 sister cells of the veg_2 ring; plus 4 large and 4 small micromeres at the vegetal pole. In the undisturbed embryo, all of the endomesoderm derives from these vegetal components, i.e., the progeny of veg_1, veg_2, and the micromeres. During blastulation, the tall, columnar veg_2 progeny and the micromere lineages form a thickened disc at the vegetal end of the embryo, the "vegetal plate." The micromere progeny are located at the center of this disc. A very important point for what follows was established by Ruffins and Ettensohn (1993, 1996) by use of dil lineage tracing: this is that the veg_2 mesoderm cell types derive from the more central region of the vegetal plate, and the endodermal veg_2 cell types from the surrounding, more peripheral region. Thus, viewed from the vegetal pole, the blastula of these embryos has a radial organization. For example, in *S. purpuratus*, at the ciliated swimming-blastula stage, there are, at the very center of the vegetal plate, 8 small micromeres; surrounding them are 16 skeletogenic mesenchyme precursors; surrounding these is a ring of around 30 prospective veg_2 mesodermal cells; surrounding these is the outer ring of about an equal number of prospective veg_2 endodermal cells; and surrounding the vegetal plate as a whole are the veg_1 descendants which form the subequatorial part of the blastula wall. Some relevant stages of embryogenesis are illustrated in Fig. 1, for those less familiar with the morphology of this embryo.

A "Process Diagram" for Endomesoderm Specification

The role of the gene network that is the chief object of this work is to control the transcriptional functions on which this concentric pattern of specification depends. We now have a fair understanding of the initial maternally organized localizations, and the interblastomere signaling events that are required for endomesoderm specification. But this does not tell us how or why the process works: for that we need to discover the zygotic gene regulatory program. Our initial problem is to extract from the experimental embryology an image of the progressive specification process that will provide a guide to the underlying regulatory functions.

Figure 1 shows an interpretation which serves this purpose, based on an immense amount of work done in many different laboratories (see legend for a brief compilation of

evidence and citations). The key steps can be summarized as follows:

(1) Initial specification of the veg$_2$ domain. The experimental evidence indicates that, under normal conditions, two inputs are required for the specification of veg$_2$ as a field of cells that will execute endomesodermal fates. The first of these is an intercellular signal passed from the micromeres to the adjacent cells, the grandparents and parents of the sixth cleavage veg$_2$ ring. This very early signaling function implies that, at least in some measure, the micromeres are already specified when they are born at fourth cleavage. New experimental insights into the mechanism by which micromere functions are confined to their lineage are reported by Oliveri *et al.* (2002). The second input required for specification of the veg$_2$ lineage is the nuclearization of β-catenin, a cofactor of the Tcf transcription regulator that is required for it to function as a gene activator. This takes place by a cell-autonomous mechanism for which intercellular contact is not necessary: remarkably, every cell, the progeny of which will express an endodermal or a mesodermal fate, displays elevated nuclear β-catenin at seventh cleavage, compared to any other cells in the embryo. Furthermore, interference with the β-catenin nuclearization process by any of several different means completely cancels endomesoderm specification.

(2) The endomesodermal Wnt8 loop. A gene encoding Wnt8, a ligand which activates the β-catenin/Tcf system, is expressed in the same prospective endomesodermal cells in which the autonomous maternal system initially causes β-catenin nuclearization (A. Wikramanayake, unpublished data). This observation implies an autoreinforcing Tcf control loop, which is set up within the endomesodermal domain once this is defined (Fig. 1). This loop is necessary, for if it is blocked by introduction of a negatively acting form of the Wnt8 ligand, so is endomesoderm specification. We note that the inferred Wnt8 loop conforms to the "community effect" concept (Gurdon, 1988; Gurdon *et al.*, 1993), i.e., a requirement for intercellular signaling within a field of cells in a given state of specification that is necessary for the maintenance and the further developmental progression of that state.

(3) The micromere Delta signal. During the seventh to ninth cleavage interval, a second signal is transmitted from the micromeres to the adjacent surrounding cells, i.e., now the inner ring of veg$_2$ lineage blastomeres. The result is the specification of these cells as mesodermal precursors. The signaling ligand is Delta, which activates the Notch (N) receptor (Sweet *et al.*, 2002; McClay *et al.*, 2000; Oliveri *et al.*, 2002). In response, the N receptor is activated specifically in the progenitors of the future veg$_2$ mesoderm (i.e., the mesoderm formed from progeny of the 8, sixth cleavage veg$_2$ cells). This event is specifically required for veg$_2$ mesodermal specification (see legend to Fig. 1).

(4) Late specification of veg1 endoderm. After midblastula stage, the elevated level of nuclear β-catenin progressively disappears from the micromere and veg$_2$ progeny, but at late blastula stage (after 24 h), β-catenin reappears in the nuclei of a ring of cells just outside the veg$_2$ domain. Thereupon, these veg$_1$ progeny begin to express endodermal markers, such as the *endo16* gene (Ransick and Davidson, 1998); the *evenskipped* gene (Ransick *et al.*, 2002); and the *krox1* gene (C.B.L. and E.H.D., unpublished data). These veg$_1$ progeny invaginate and will contribute mainly to the hindgut.

The diagram in Fig. 1 suggests that regulatory genes carrying out several different classes of function are likely to be required for endomesoderm specification. These include genes required for micromere functions; genes required for endomesodermal specification that are dependent for activation on the Tcf system; mesodermal genes that are activated downstream of the N system; regulatory genes required for endoderm or for mesoderm cell type specification; and also batteries of downstream genes that encode skeletogenic, mesodermal, and endodermal differentiation products. A miscellaneous collection of genes was already known that fell into one or more of these functional categories. An initial challenge was to search more systematically for additional members of the endomesodermal gene set so that we would not be dependent on prior accidents of discovery.

METHODS

Overview

A broad strategy has emerged from this project by which the regulatory gene network underlying endomesoderm specification can be solved. We began by constructing an *a priori* "logic model," now of course obsolete, which proposed a minimum set of interrelations between regulatory genes on the basis of the interpretation summarized in Fig. 1. Known genes that might be involved in endomesoderm specification were then placed on this model. This produced a series of predicted inputs and outputs among these genes, i.e., an initial proposal of how they might be functionally linked to one another. Table 1 provides an overview of the methodological components of the strategy used to arrive at this network. In addition, this project has relied on several newly devised computational procedures. These are discussed in a separate paper (Brown *et al.*, 2002). Figure 1 of Brown *et al.* (2002) is a flow diagram which shows how the specific computational aids were used.

Perturbation Analysis

We applied the same three kinds of regulatory perturbation to analysis of the effects on network genes as used in the gene discovery screens of Table 1, *viz* introduction of cadherin mRNA; introduction of mRNA encoding the N extracellular domain; and alteration of the location and level of the Brachyury transcription factor. In addition to these, we made extensive use of three other methods. The first of these was injection of antisense morpholino oligonucleotides (Howard *et al.*, 2001) in order to block translation of specific messages. The efficiency was checked in each case by use of a fusion mRNA in which the sequence encoding GFP follows the target mRNA sequence to which the antisense oligonucleotide was designed to bind. No nonspecific phenotypic effects of morpholino oligonucleotide injection other than a slight delay at the cleavage stage of development were observed at the concentrations

TABLE 1
Experimental Approaches to the Endomesoderm Network

Requirement	Method	Reference
Gene discovery	Differential macroarray screen[a]	Rast *et al.*, 2000
	cadherin vs. LiCl	Ransick *et al.*, 2002
	DnN vs. LiCl	Calestani, unpublished data
	Ectopic *bra* vs. Bra MASO[b]	Rast *et al.*, 2002
	Homology screens using probes from other species	Table 3
Finding *cis*- regulatory elements	Interspecific sequence comparison	Brown *et al.*, 2002
	BAC libraries from *S. purpuratus* and *L. variegatus*	Cameron *et al.*, 2000
	Isolate BACs containing relevant genes[c]	
	Obtain and annotate genomic sequences[d]	Table 2
	Determine and test conserved elements[e]	Yuh *et al.*, 2002
Gene expression characterization	WMISH	Table 3
	QPCR	Table 3
Perturbation analysis	Cadherin MOE	Ransick *et al.*, 2002;
	DnN MOE	Table 3; Appendix 1
	MASO (many genes)	Table 3; Appendix 1
	Engrailed fusions (5 genes)	Table 3; Appendix 1
	Transcription factor MOE (one gene)	Oliveri *et al.*, 2002

Note. Abbreviations: MASO, morpholino antisense oligonucleotide; WMISH, whole-mount *in situ* hybridization; QPCR real time quantitative fluorescence PCR; DnN, dominant negative Notch (Sherwood and McClay, 1997); *bra*, *brachyury*; MOE, mRNA overexpression or ectopic expression.

[a] Subtractive hybridization, in which single-stranded driver RNA lacking a given class of endomesodermal sequences is reacted to high driver C_0t with a complementary single-stranded population of DNA fragments that include endomesodermal sequences. The products are separated by hydroxyapatite chromatography, and the selectate sequences remaining single stranded are linearly amplified to produce several micrograms of asymmetric RNA probe. Large nylon filters that contain arrayed cDNA libraries prepared with a Genetix QBot robot are screened with this probe and with control probe representing the unselected RNA population in order to identify and recover differentially expressed (i.e., endomesodermal) clones. Complete libraries containing $\sim 10^5$ clones and representing each of several different embryonic stages were screened. The screens are analyzed with the aid of the BioArray software developed for this purpose in the course of this project (Brown *et al.*, 2002). The screening sensitivity of this method is such as to permit recovery of sequences present at less than five molecules per average cell, in one case for example leading to the identification of a regulatory gene expressed at low levels in only four cells in the whole embryo (Ransick *et al.*, 2002).

[b] Expression of *brachyury* was forced to occur ectopically in clones of cells outside of the vegetal plate, and the cells expressing the *brachyury* gene were isolated by fluorescence-activated cell sorting, using a coexpressed GFP marker. Translation of *brachyury* mRNA was also interrupted by use of α-*bra* MASO.

[c] Five or so candidate BACs from each library were crudely mapped to ensure that the gene is not near the end of the insert, and after confirming the presence of the gene, the best of these candidates were selected on the basis of gene position and insert length. Mapping was done on partial or complete restriction enzyme digests displayed by pulse field gel electrophoresis, and then blot hybridized to separate probes representing the right and left ends of the BAC vector, and the gene itself.

[d] BACs were sequenced at DOE's Joint Genome Institute (JGI) in Walnut Creek, CA, or at the Institute for Systems Biology (ISB) in Seattle, WA, as indicated in Table 2. The BACs were sequenced to $\geq 8\times$ coverage, so that a completed scaffold was obtained, with <10 small gaps per BAC, sometimes none. The assembled sequences were then analyzed for predicted and recognized genes, and for homologies to other sea urchin sequences, including ESTs, cDNAs, and repetitive elements, using a custom-designed annotator, SUGAR (Brown *et al.*, 2002).

[e] The FamilyRelations algorithm uses a small window set by the operator (usually 20 or 50 bp) which is slid along the *S. purpuratus* BAC sequence. The *L. variegatus* sequence is searched for similarity to the *S. purpuratus* sequence within each window at a set level (usually 70-100%) as it moves along the sequence. No assumptions or constraints with respect to alignment are imposed.

of morpholino oligonucleotide at which data were extracted. Nonspecific effects were monitored directly by use of control morpholino oligonucleotides. A powerful control on the specificity of those consequences of the perturbation that were observed is provided by the large number of other genes, the activity of which was unaffected in each experiment. A second form of perturbation that we used for a few genes encoding transcription factors is the use of Engrailed repressor domain fusions, usually with the DNA-binding domain of the factor. *In vivo*, the effect of mRNA encoding such fusions is to silence the target genes of the transcription factor. The third method, used only for one gene, is injection of the native mRNA, leading to its ectopic expression.

The effects of these perturbations on other genes in the network were determined by quantitative PCR (QPCR). This method affords simultaneous measurement of the real-time build up of fluorescent PCR product in 96 samples simultaneously. From the kinetics of product accumulation, the prevalence of the transcript recognized by the primers in each well can be calculated directly. For example, if a gene is a direct target of a positively acting transcription factor and an Engrailed domain fusion to this factor has been introduced,

the transcript level for that gene will be decreased compared to normal, sometimes by factors of greater than 10-fold. In our analyses, we ignored all changes that were less than 3-fold, i.e., we generally required that at one or more of the stages tested, less than 30% of the control level of the transcript or more than 300% result from the perturbation. Examples of perturbation experiments and analysis of data obtained therefrom can be seen in several other papers, *viz* those of Oliveri *et al.* (2002), Ransick *et al.* (2002), and Rast *et al.* (2002). The QPCR measurements obtained in these perturbation experiments, from which the possible inputs in Table 2 were deduced, are posted on our Web site (http://www.its.caltech.edu/~mirsky/qpcr.htm (End-mes Gene Network Update, Network QPCR Data).

A good number of the perturbation results underlying the network, though not all, have been confirmed on multiple batches of cDNA made from independent embryo cultures at each relevant stage. Other results have been taken on single batches of cDNA per time point, but are strongly substantiated by similar outcomes at successive time points. A fraction of the results represent only a single batch of cDNA at a single time point. These details and all of the quantitative results of the perturbation experiments are posted on the QPCR Web site given above and are included here as an Appendix. In our experience, such QPCR measurements are inherently reliable and reproducible: the data listed on the Web site are the averages of duplicate or triplicate samples, and usually the variance among these samples is very small (~10% of the minimum level of difference between control and experimentally perturbed samples that is taken here as significant, i.e., ≥3-fold difference). There is much more variance between different cDNA batches made from different embryo cultures, particularly if these are not precisely at the same stage and the measurement concerns a gene the expression of which is changing rapidly at that point in development. Despite all this, we do not often see data sets in which ≥3-fold differences in transcript level between perturbations and control samples fail to reproduce in independent cDNA batches. Indeed, it is possible that some of the weaker results that we excluded from the model in fact indicate true relationships.

The 42 genes currently included in the network model that we discuss below are listed in Table 3. Here are given the functional nature of the protein encoded by each gene; the source of the gene if published earlier, or the screen from which it derived; the perturbations used to establish its linkages into the network; and the implied inputs, direct or indirect, into its *cis*-regulatory system.

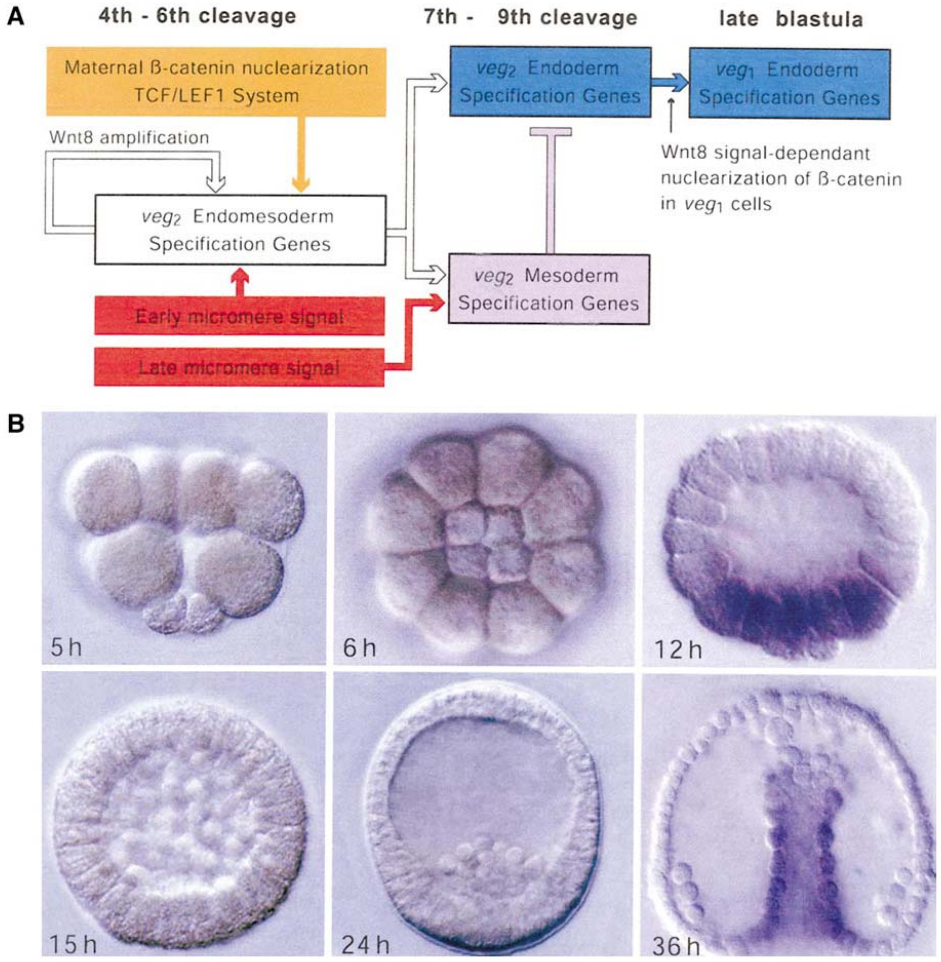
RESULTS

The Network Model

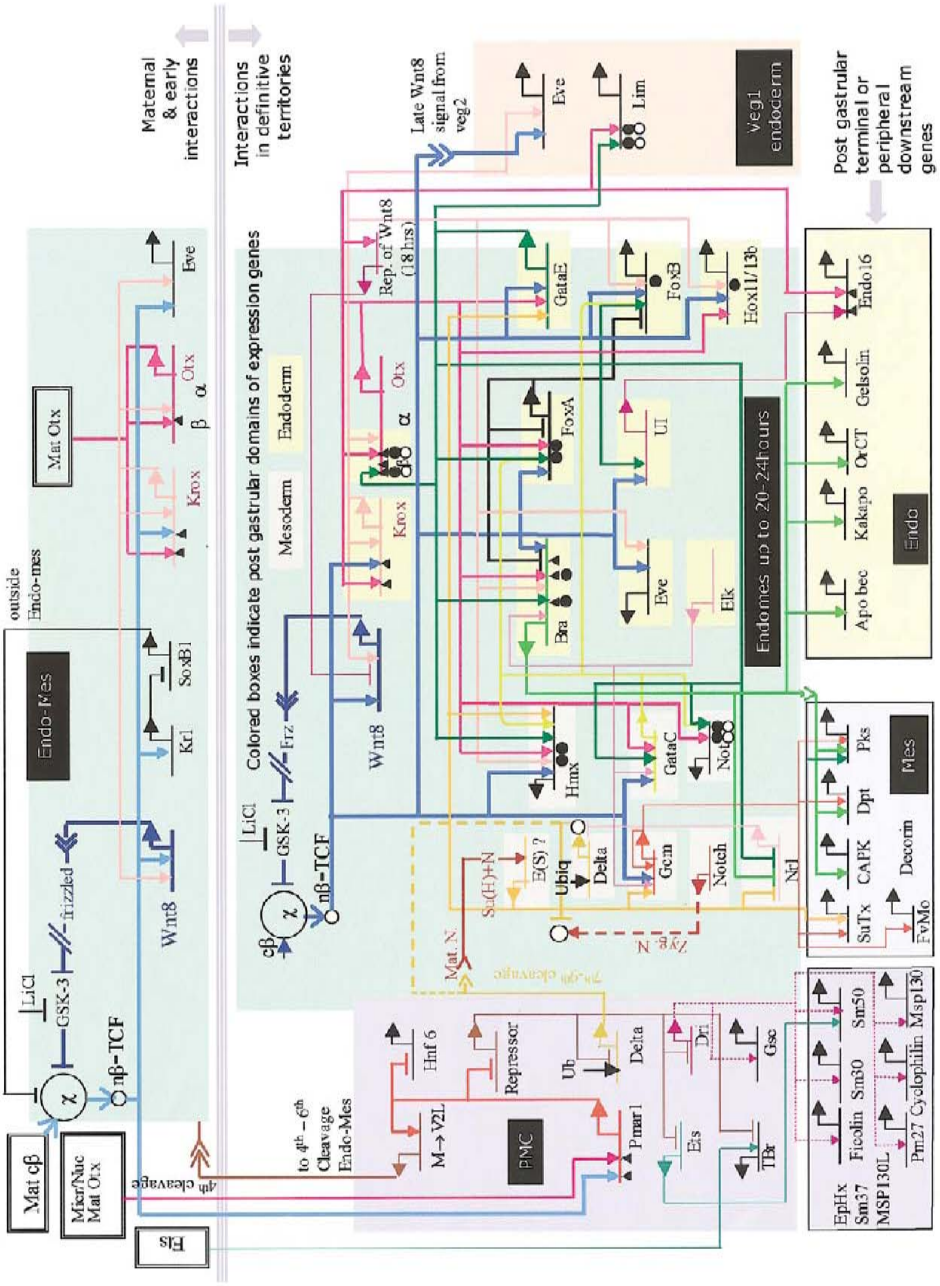
A model of the endomesoderm specification network as of this writing is given in Fig. 2 (the network is continuously updated as further information accumulates; see our Web site: http://www.its.caltech.edu/~mirsky/endomes.htm (End-mes Gene Network Update). Figure 2 presents the network in the form of a "view from the genome" (Arnone and Davidson, 1997; Bolouri and Davidson, 2002). This means that all of the presumed interactions that occur among the genes of the network throughout the process of endomesoderm specification, in all relevant cell types, at all stages, are shown at once. The import of the view from the genome is that it displays the structure of the network architecture, as this would be perceived in the genomic DNA sequence if the interactions are direct, and if we knew all the relevant *cis*-regulatory target sites. Every node of the view from the genome is subject to proof or disproof by appropriate *cis*-regulatory analysis. The developmental workings of the model are shown in Figs. 3–6, which provide instead, "views from the nucleus." These indicate those subsets of the interactions shown in the view from the genome that operate at different developmental stages, specifically in the individual nuclei of the micromere lineage, the veg_2 endomesoderm, and the resolving veg_2 mesoderm and endodermal domains.

The diagram in Fig. 2 does include in its organization some low-resolution spatial and temporal information. The earliest developmental events to occur in the future veg_2

FIG. 1. Process diagram for endomesoderm specification in regularly developing sea urchin embryos and stages of development. (A) Process diagram. Signaling functions expressed in the large (skeletogenic) micromeres are in red boxes, endodermal functions of veg_2 and veg_1 progeny are in blue boxes; veg_2 mesodermal functions are in lavender. All boxes surrounded by black lines imply sets of regulatory genes required for execution of these respective fates. The tan box indicates the maternal β-catenin nuclearization system, discovered by Logan *et al.* (1999). If this and the Wnt8 loop are blocked by injection of mRNA encoding the intracellular domain of cadherin, the embryo develops without endomesoderm, as a hollow ball of ectoderm, and it fails to express many endomesodermal regulatory and other genes (Logan *et al.*, 1999; Wikramanayake *et al.*, 1998; Gross and McClay, 2001; Ransick *et al.*, 2002; this work). The same exclusively ectodermal embryoids are produced if the loop is blocked in other ways, i.e., overexpression of Gsk3 (Emily-Fenoille *et al.*, 1998) or of negative form of Tcf (Huang *et al.*, 2000; Vonica *et al.*, 2000). The early micromere signal denoted by the upper red box was demonstrated by transplantation of micromeres, which results in induction of an ectopic gut (Ransick and Davidson, 1993) and by ablation of micromeres, which prevents normal gastrulation or normal expression of *endo16*, a blastula stage marker of endomesodermal fate (Ransick and Davidson, 1995). This signal is probably transmitted in the fourth to sixth cleavage interval, for after that the micromeres can be removed without affecting aspects of veg_2 specification that do not work properly if they are removed earlier (Ransick and Davidson, 1995). The Wnt8 loop shown is based on the expression pattern of the *wnt8* gene (see text). Injection of mRNA encoding a negatively acting form of Wnt8 also causes failure of endomesoderm specification. Note, however, that Wnt8 is not the early micromere signaling ligand because micromeres bearing the negatively acting form of Wnt8 can still induce secondary gut on transplantation (D.R.M., unpublished data). The late micromere signal indicated in the lower red box is Delta, a ligand for Notch (N) (Sweet *et al.*, 2002; Oliveri *et al.*, 2002). The role of N activation is inferred from the studies of Sherwood and McClay (1997, 1999) and Sweet *et al.* (1999): these experiments demonstrate the mobilization of N exactly in the veg_2 mesodermal domain; the production of excess mesodermal cell types on introduction of constitutively active N receptor; and the failure of veg_2 mesodermal specification on introduction of excess negatively acting N receptor (i.e., extracellular



domain). Further evidence is the spatial distribution of *delta* mRNA (Sweet *et al.*, 2002; Oliveri *et al.*, 2002). The negative interaction shown between mesodermal and endodermal genes is based on the observation that in LiCl-treated embryos the mesodermal domain is enlarged at the expense of cells that would otherwise have become endoderm (the endodermal domain is also enlarged at the expense of prospective ectoderm cells; Cameron and Davidson, 1997; Hörstadius, 1939). Furthermore, the expansion of the mesodermal domain in response to constitutively active N also occurs at the expense of endoderm (Sherwood and McClay, 1999, 2001). For the late specification of veg_1 endoderm, see Logan and McClay (1997) and Ransick and Davidson (1998)). The late endodermal specification of veg_1 cells depends on a signal from veg_2, which is likely to be Wnt8 (McClay and Logan, 1996; Davidson *et al.*, 1998; D.R.M., unpublished data). (B) Some representative developmental stages: 5 h, fourth cleavage 16-cell stage embryo viewed from side (only half of the cells are visible); note micromeres at pole of embryo. The veg_1 and veg_2 cell lineages derive from the four large cells (macromeres) at sixth cleavage. Six hour, fifth cleavage, 28-cell stage, vegetal view. The micromeres in the center directly above the six veg_2 cells. Twelve hour, eighth cleavage, early blastula stage, approximately 200 cells; the WMISH shows expression of the *evenskipped* gene (Ransick *et al.*, 2002) in the descendants of the veg_2 lineage, which now form the vegetal plate. Fifteen hour, early ninth cleavage, approximately 300 cells, viewed from the side. There are a total of nine cleavages in *S. purpuratus*, but these are increasingly asynchronous after sixth cleavage. Twenty-four hour, mesenchyme blastula stage, approximately 650 cells, viewed from the side. Thirty-six hour, midgastrula stage, lateral view from the oral side, about 800 cells. The endodermal constituents of the archenteron are expressing *endo16*, though the gene is beginning to be down regulated in the foregut region (Ransick *et al.*, 1993). Note the mesenchymal cells delaminating from the tip of the archenteron. Skeletogenic mesenchyme cells can be seen at the base of the ectodermal wall on either side.



and veg_1 endomesoderm are indicated above the double horizontal line at the top, and in the polar micromere domain in the lavender field at the left. This field represents, in particular, the skeletogenic lineage descendant from the four fifth-cleavage large micromeres. Most of the *cis*-regulatory elements shown in the light green veg_2 endomesodermal region eventually become specified as either endodermal or mesodermal, and these destinations are indicated in Fig. 2 by the color of their backgrounds. In Table 3 is indicated the domains of expression of all of the genes in the model at several different stages, as established by *in situ* hybridization. As can be seen there, many of the regulatory genes in the network are initially activated throughout the endomesodermal veg_2 domain ("EM" in Table 3). The process by which the concentric mesodermal and endodermal specification domains are established within the veg_2 region of the embryo can be observed in the pattern of transcription of certain regulatory genes as early as 15 h (see below). But the patterns generated by others do not become clearly endodermal or mesodermal until about 24 h. This indicates completion of the specification process in *S. purpuratus* all across the vegetal plate, as also shown by the cell marking experiments of Ruffins and Ettensohn (1993, 1996).

The rectangular areas at the bottom of Fig. 2 contain downstream differentiation genes of each domain, i.e., genes expressed in differentiated skeletogenic cells, in veg_2 mesodermal cells (mainly pigment cells), and in gut endoderm cells. The network in Fig. 2 can be said to terminate with these genes: its most peripheral linkages are those which lead from the regulatory apparatus that is portrayed in all the rest of the model into the *cis*-regulatory elements of these differentiation genes. Note that the large majority of the genes in the network are genes encoding regulatory proteins, or signaling components that ultimately affect genes encoding regulatory proteins.

Technical Aspects

The regulatory interrelationships proposed in the model of Fig. 2 are derived directly from the perturbation analyses summarized in Table 3, but for many of the key players, there are other data as well. The time and place of expres-

FIG. 2. Regulatory gene network for endomesoderm specification: the view from the genome. The architecture of the network is based on perturbation and expression data listed in Table 3 and compiled in the Appendix, on data from *cis*-regulatory analyses for several genes, and on other experiments discussed in text. See http://www.its.caltech.edu/~mirsky/endomes.htm (End-mes Gene Network Update) for the current version of the model in this figure, and http://www.its.caltech.edu/~mirsky/qpcr.htm (End-mes Network QPCR Data) for a current list of quantitative results of perturbation experiments and temporal details. Each short horizontal line from which bent arrows extend represents the *cis*-regulatory elements responsible for expression of the genes named, in the spatial domain shown. Genes are indicated by the names of the proteins they encode. The arrows and barred lines indicate the normal function of the input (activation or repression), as deduced from changes in transcript levels due to the perturbations. The relationships shown may in some cases be indirect, though as indicated in text, all known or suspected indirect relationships excepting those mediated by intercellular signaling have been excluded from the model (see Notes to Appendix for details). For linkages that are direct, each input arrow constitutes a prediction of specific transcription factor target site sequence(s) in the relevant *cis*-regulatory control element. In some cases, the predicted target sites have been identified in experimentally defined *cis*-regulatory elements that generate the correct spatial pattern of expression (solid triangles). At the upper left, the light blue arrow represents the maternal β-catenin (cβ) nuclearization system (χ), which autonomously causes accumulation of β-catenin in the nuclei of all future endomesodermal cells. This transcriptional system (nβ/Tcf) is soon accelerated and then taken over by zygotic Wnt8 (dark blue lines); its initial activation, of mixed zygotic and maternal origins, is shown in light blue. Data for the roles of SoxB1 and Krüppel-like (Krl) are from Kenny *et al.* (1999) and Howard *et al.* (2001). Data for the role of Ets are from Kurokawa *et al.* (1999) and K. Akasaka (unpublished data). "Micr/Nuc Mat Otx" refers to the early localization of maternal Otx in micromere nuclei at fourth cleavage (Chuang *et al.,* 1996). Genes labeled "Repressor" are inferred; all other genes shown are being studied at the DNA sequence level and by multiplexed QPCR. "Ub" indicates a ubiquitously active positive input inferred on the basis of ubiquitous expression seen by whole-mount *in situ* hybridization, under conditions in which a spatial repression system that normally confines expression has been disarmed. At the top, above the triple gray line are the earliest interactions; in the middle tier the spatial domains of the endomesoderm are color coded, and genes are placed therein according to their final loci of expression. As indicated (black background labels), the lavender area to the left represents the skeletogenic micromere (pmc) domain prior to ingression; the light green area indicates the veg_2 endomesoderm domain, with genes eventually expressed in endoderm on yellow backgrounds, and genes eventually expressed in mesoderm on blue backgrounds; the tan box at right represents the veg_1 endoderm domain. Many genes are initially expressed over broader ranges, and their expression later resolves to the definitive domains. The rectangles in the lower tier of the diagram show downstream differentiation genes. Dotted lines indicate inferred but indirect relationships. Arrows inserted in arrow tails indicate intercellular signaling interactions. Small circles indicate perturbation effects that resist rescue by introduction of mRNA encoding another input into the same *cis*-regulatory element: i.e., both inputs are required and one cannot substitute for the other. In the case of the β-Otx transcription control element, the experiment was done both by introduction of Krox mRNA in the presence of Otx-En mRNA and vice versa (open and closed pairs of circles). Large open ovals represent cytoplasmic biochemical interactions at the protein level, e.g., those responsible for nuclearization of β-catenin, for the effect of Delta on Notch (Jacobsen *et al.,* 1998); or the effect of Neuralized, an E3 Ubiquitin ligase with specificity for Delta (Yeh *et al.,* 2000, 2001). The diagram displays what we term "the view from the genome" (Arnone and Davidson, 1997; Bolouri and Davidson, 2002), i.e., it purports to illustrate the sum of linkages that are functional in different places and at different stages of the endomesodermal specification process. This is the form of the model that is required for prediction of genomic target site sequences.

sion (Table 3) are in some cases immediately revealing. In general, the patterns of expression provide reality checks throughout: a prime object of the network analysis is ultimately an explanation, in terms of *cis*-regulatory interactions, of why each gene goes on and when and where it does.

Some of the interactions shown in Fig. 2 may be true in the sense that they accurately represent the results of a particular experimental perturbation, but are not useful because they do not predict a *cis*-regulatory input. This will be the case where the functional linkage between given genes is actually indirect. If gene A activates gene B, and gene B activates gene C, a knockout of gene A expression will affect expression of gene C, but the target sites for the transcription factor encoded by gene A are only to be found in the *cis*-regulatory element of gene B, not in gene C. Many possible linkages initially suggested by the perturbation analyses have been excluded from the model on grounds of probable and sometimes demonstrated indirectness. These exclusions are indicated explicitly by the footnotes in the Appendix, keyed to the specific measurements. For positive interactions, indications of indirect relationships are the expression of the apparent target gene in different cells or at different times than expression of its apparent regulator; or the rescue of expression of the apparent target gene by introduction of mRNA encoding a different transcription factor than produced by the apparent regulator. For example, a strong effect on *delta* expression was observed on introduction of Cadherin mRNA, such that the gene is expressed at only about 10% the normal level. This suggested a possible input of the β-catenin/Tcf system into *delta* (Appendix, *delta*/Cad MOE data). But *delta* expression is controlled via expression of *pmar1*, also a β-catenin/Tcf target gene (Fig. 2). In fact, *delta* expression is rescued in embryos into which both Pmar1 and Cadherin mRNA were introduced. This rescue experiment shows that the cadherin effect on *delta* is in fact indirect (Appendix, note 15).

Rescue experiments have been attempted on several of the genes in the model, and direct *cis*-regulatory observations have been carried out on some others (see legend to Fig. 2). So far, the predicted *cis*-regulatory relationships have been substantiated where tested, but the *cis*-regulatory level of demonstration is yet available for only a minority of the genes in the model. These are indicated by the initials ECRA (experimental *cis*-regulatory analysis) in Table 1. So, at present, the provisional network of Fig. 2 is mainly based on the results of the large matrix of perturbation analyses, summarized in the Appendix, applied to the endomesodermal specification process visualized in Fig. 1; and on the time and place of expression of the network genes, and the nature of the products that they encode.

A potential problem exists in any perturbation experiment where an mRNA encoding a transcription factor is overexpressed. If the result is to raise the level of the mRNA per cell by more than about an order of magnitude, the higher concentration of the factor could result in binding to weak target sites that are not normally engaged by it in that given context. This of course refers either to natural mRNAs or to Engrailed domain fusions. We were at pains in this work to measure the natural concentrations of the mRNAs per cell (by QPCR) and to introduce synthetic mRNAs at levels within an order of magnitude per cell of the natural level. Furthermore, these mRNAs decay, and the actual concentration by blastula stage is significantly lower than that introduced into the egg. The *pmar1* mRNA used by Oliveri *et al.* (2002) encodes a repressor, and produces almost the same results as mRNA encoding a Pmar1-Engrailed domain fusion, introduced at the same very low levels. The Otx-Engrailed fusion was created and its specificity shown by Li *et al.* (1999), and also Yuh *et al.* (2001). Of particular importance, because of the early role in the network of the *krox* gene, is the Krox-Engrailed fusion. This was again used at a level (100 fg/egg) that would produce per cell concentrations close to the level of natural *krox* mRNA. In all of these cases, the large majority of genes tested showed no response to the introduced mRNAs (see Appendix).

Initial Specification of the Micromere and the veg$_2$ Endomesodermal Lineages

The initial zygotic phase of endomesoderm specification is completed during cleavage. As summarized above, the endomesoderm derives from the micromeres, veg$_2$, and part of the veg$_1$ lineages: the micromere lineage is born at fourth cleavage and their skeletogenic daughters at fifth cleavage; and the veg$_2$ lineage, all progeny of which contribute to endomesoderm, is born at sixth cleavage, together with their veg$_1$ sister cells. The early micromere signal to the grandparents and parents of the veg$_2$ lineage is passed between fourth and sixth cleavage (Ransick and Davidson, 1995); β-catenin nuclearization arrives at its maximum extent in the micromere, the veg$_2$ and part of the veg$_1$ lineages at sixth cleavage (Logan *et al.*, 1999); the Delta signal from the skeletogenic micromeres to the adjacent veg$_2$ blastomeres is passed between seventh and about ninth cleavage (McClay *et al.*, 2000). In Fig. 3 are shown those aspects of the overall network model in Fig. 2, the particular "view from the nuclei" that refers to the initial specification of the micromere and veg$_2$ lineages. This diagram includes *cis*-regulatory interactions occurring between fourth and ninth cleavages, approximately.

The interactions included in Fig. 3 result in three major steps forward. The notes keyed to the red numerals in Fig. 3 summarize individual events of regulatory significance. The first of these is the installation of a state of specification specific to the micromeres (Notes 1–5). The second is the initiation of zygotic regulatory functions that begin to lock in a state of endomesodermal specification in the veg$_2$ domain (Notes 6–9). The third is the establishment of the regulatory system that underlies the developmentally essential signaling functions executed specifically by the micromere lineage (Notes 10 and 11).

From the time of their birth, the nuclei of the micromeres are in some ways unique with respect to the remainder of the embryo. They are the first to contain nuclearized β-catenin (Note 1; Logan *et al.,* 1999); and maternal Otx transcription factors accumulate in them (Note 2; Chuang *et al.,* 1996). In addition, they lack specific transcription factors found elsewhere, such as SoxB1 (Kenny *et al.,* 1999). The *pmar1* gene is activated by the β-catenin/Tcf and Otx inputs (Fig. 3, Note 3; Table 1). These are most probably direct inputs, as the respective target sites are present in an active *pmar1 cis*-regulatory element (P.O. and E.H.D., unpublished data). As described in detail elsewhere (Oliveri *et al.,* 2002), the Pmar1 homeodomain regulator acts as a repressor of a gene encoding another, unknown repressor which is ubiquitously active ("Repressor" gene in Fig. 3). This repressor keeps off a series of other genes encoding skeletogenic lineage-specific transcription factors, except in the micromeres and their descendants, where *pmar1* is expressed. The known target genes of the *pmar1* repression system include *tbrain* (*tbr*), *deadringer* (*dri*), and an *ets* class gene, all of which encode transcription factors that are required for the skeletogenic functions of the large micromere lineage (Notes 4 and 5; Kurokawa *et al.,* 1999; Oliveri *et al.,* 2002; G.A. and E.H.D., unpublished data). Ectopic expression of *pmar1* mRNA causes global derepression of these genes and a general, irreversible conversion of the embryonic blastomeres to cell types expressing skeletogenic functions (Oliveri *et al.,* 2002). Note that *pmar1* expression is ephemeral, in that the mRNA normally disappears during early blastula stage (Oliveri *et al.,* 2002). Much later events, such as the expression of *dri* in the oral ectoderm after mesenchyme blastula stage (G.A. and E.H.D., unpublished data), are not affected by the early network of repression, which it is the role of *pmar1* to relieve in the large micromere lineage. The regulatory interactions summarized in Notes 1–5 of Fig. 3 at least partially explain the specificity of the micromere lineage transcriptional program.

In the veg_2 endomesoderm, two regulatory subcircuits execute the process by which the zygotic transcriptional apparatus interprets the initial cues with which it is confronted, and by which it establishes an endomesodermal state of specification. As reviewed above, the initial cues are the early micromere signal to the blastomeres from which the veg_2 endomesoderm lineage derives, and the nuclearization of β-catenin in the veg_2 lineage, i.e., activation in these cells of a positive Tcf transcriptional input. Current studies have revealed a number of putative targets of the Tcf control system. Among the earliest are those indicated by Notes 6 and 7 in Fig. 3. The *krüppel-like* (*krl*) gene and its dependence on the β-catenin/Tcf system were discovered by Howard *et al.* (2001). Interaction 6 of Fig. 3 shows that Krl acts as a transcriptional repressor of the *soxb1* gene in the endomesodermal domain in which the *krl* gene is active following β-catenin nuclearization (Howard *et al.,* 2001). It has been shown that some Sox factors physically bind to β-catenin, thereby suppressing the β-catenin/Tcf signal transduction pathway (Zorn *et al.,* 1999). So the significance of the repression of *soxb1* expression in the endomesoderm, and its expression elsewhere, is that it sets up a reinforcing function by confining β-catenin nuclearization to the endomesodermal cells where *soxb1* is not expressed. The β-catenin/Tcf input also contributes to activation of the *krox1* gene in the endomesoderm (Table 1; interaction 7). This gene in turn locks itself on, and also provides an input to the Otx gene, which eventually locks itself on as well. The maternal Otx gene product may in turn positively regulate the *krox1* gene, though in quantitative terms the evidence for this is not as strong as for all the other interactions shown for the *krox1* and *β1/2otx* systems (see Web site QPCR data). Zygotic activation of the *β1/2otx cis*-regulatory system occurs only toward the end of the fourth to ninth cleavage period here considered (see Yuh *et al.,* 2002, for the *cis*-regulatory system of the *otx* gene), but the *krox1* gene is clearly a very early regulatory player in endomesoderm specification. In any case, the autostimulatory lock-on of both genes, combined with their cross-regulation, produces a transcription-level stabilization of the endomesodermal regulatory state (Note 8).

And this is not all: we discovered that cleavage-stage expression of the *wnt8* gene in the endomesoderm is sharply downregulated by introduction of mRNA encoding a Krox1-Engrailed fusion protein (Table 1). The level of *wnt8* transcript drops to only a few percent of normal at 6 and 12 h in these embryos, strongly suggesting that there is direct control of *wnt8* expression by the Krox1 transcription factor, causing its expression in the cells of the endomesoderm since that is where the *krox1* gene is active (interaction 9 in Fig. 3). The Wnt8 ligand stimulates the β-catenin/Tcf system in the cells receiving the signal. So the result is to transfer control of this system from the autonomous cytoplasmic mechanism by which its activity was initiated to a zygotically controlled, intercellular signaling mechanism operating among the cells of the endomesoderm.

Finally, returning to the micromere domain, we see that *pmar1* transcription is required for the localized expression of both of the developmentally essential signals that these cells produce. The expression of the *pmar1* gene is upstream of the early signal (Note 10) to the immediate ancestors of the veg_2 founder cells, in that cells expressing *pmar1* ectopically cause adjacent cells to express *endo16*, just as do transplanted fourth cleavage micromeres (Ransick and Davidson, 1993; Oliveri *et al.,* 2002). The ultimate transcriptional targets of the early signal in the veg_2 lineages are unknown, but as mentioned above, they is required for normal endomesodermal specification (Ransick and Davidson, 1993, 1995). Expression of the Delta signal in the micromeres is permitted to occur in these cells by the operation of the *pmar1* repression system (Note 11). If the *pmar1* gene is expressed ectopically, the *delta* gene is activated in every cell in the embryo, but since *pmar1* is normally expressed only in the micromeres, it is just these polar cells from which the signal normally emanates (Oliv-

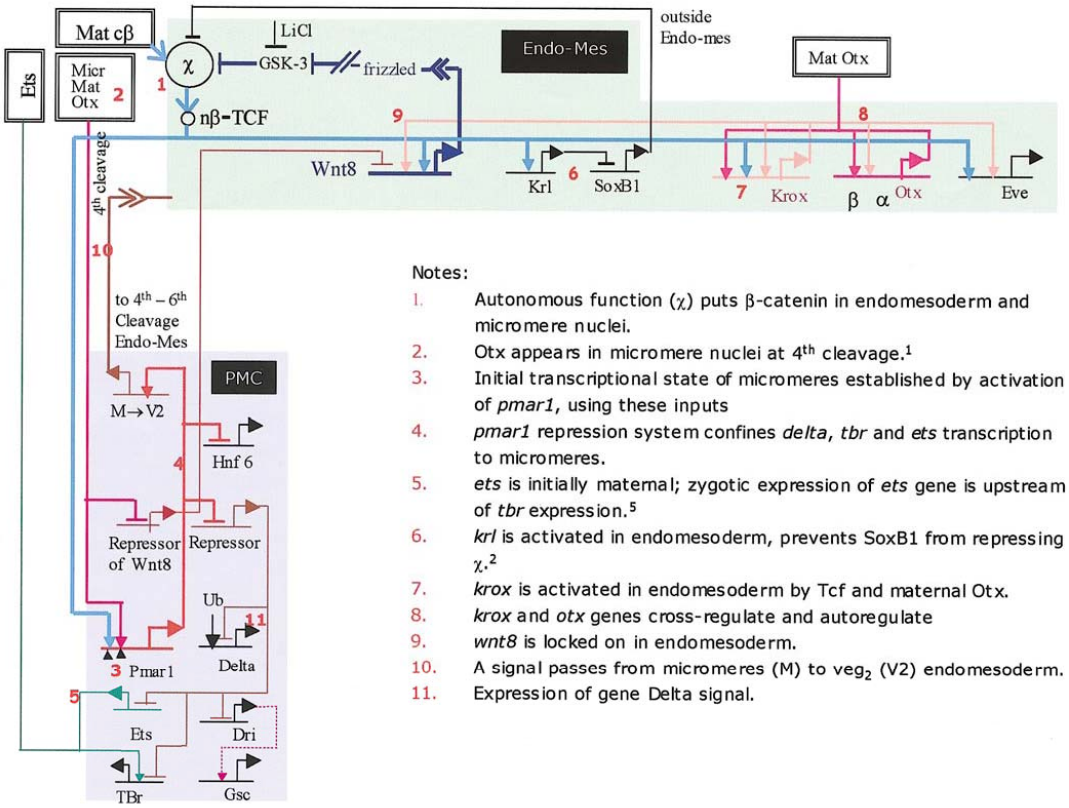


FIG. 3. Initial specification of micromere and veg_2 endomesodermal domains. The diagram includes the portion of the model in Fig. 2 that includes events occurring from about fourth to ninth cleavages within these domains. The "notes" (red numerals) indicate specific functional aspects. See legend to Fig. 2 for symbolism and architectural explanation of the form of the model. References: [1]Chuang *et al.*, (1996); [2]Kenny *et al.*, (1999); [3]Ransick and Davidson (1993; 1995); [4]Oliveri *et al.*, (2002); [5]Kurokawa *et al.*, (1999); unpublished data of K. Akasaka.

eri *et al.*, 2002). The localized expression of Delta in the micromeres provides the crucial spatial cue for mesoderm specification.

In the veg₂ Endoderm from About Eighth Cleavage to Mesenchyme Blastula

As the blastula stage begins after about eighth cleavage (~10 h in *S. purpuratus*), none of the known regulatory genes that later execute definitive endoderm or mesodermal control functions have yet become active, except for *krox1*. By the mesenchyme blastula stage (20–24 h), all of these genes are active. Figure 4A shows that for many of these genes normal expression requires β-catenin/Tcf inputs, and many more are affected indirectly (Cad MOE results in Appendix). These inputs are initially generated by the autonomous β-catenin/Tcf system, but in this time frame are amplified by the Wnt8 intercellular signaling loop just discussed. The widespread importance of this loop as a stable source of activating inputs is explained by this diagram (Notes 1 and 2). The earliest β-catenin/Tcf inputs, which are very probably direct because the other inputs that could provide the pathway for an indirect effect are not yet available, are into the *wnt8*, *krox*, and *pmar1* genes, and very early effects relative to the other known inputs are detected as well for *gatac*, *gatae*, and *foxa* genes. The *eve* gene is apparently affected by two early inputs (Krox and Tcf) and one of these effects could be indirect; however, *eve* expression spreads to the veg_1 endoderm domain exactly as *wnt8* expression does after 24 h, and this plus the strength

TABLE 2
BAC Clones Sequenced for the Endomesoderm Network Project

Gene	Clone name	Original source of *S. purpuratus* cDNA probe	Size (kb)	Sequencing center
Sp apo bec	112F15	Brachyury target screen	50	ISB
Lv apo bec	031L12	HS	55	ISB
Lv β-catenin	48J5	Miller and McClay, 1997; HS	48	ISB
Sp brachyury	117A3	Peterson *et al.*, 1999	146	JGI, ISB
Lv brachyury	187F03	HS	70	ISB
Sp capk	83N11	Brachyury target screen	56	ISB
Lv capk	177F13	HS	49	ISB
Lv decorin	114C22	Dominant negative Notch screen; HS	70	ISB
Sp delta	046A16	Zhu *et al.*, 2001	157	JGI
Lv delta	20B17, 71J9	Cadherin over-expression screen; HS	70	ISB
Sp dpt	188B16	Brachyury target screen	36	JGI
Lv dpt	037C14	HS	60	ISB
Sp eve	079A02	Cadherin over-expression screen	173	JGI
Lv eve	112E15	HS	63	ISB
Sp foxa	041I19	Harada *et al.*, 1996; HS	150	JGI
Lv foxa	004G18	Harada *et al.*, 1996; HS	55	ISB
Lv foxb	92H22	Luke *et al.*, 1997; HS	87	ISB
Sp gatac	081C18	Pancer *et al.*, 1999	140	JGI
Lv gatac	044D13	HS	71	ISB
Sp gatae	091A10	Pancer *et al.*, 1999	184	JGI
Lv gatae	032P20	HS	71	ISB
Sp gcm	033O18	Cadherin over-expression screen	57	ISB
Lv gcm	018J3	HS	70	ISB
Sp gelsolin-like	118E15	Brachyury target screen	50	ISB
Lv hmx	076P20	Martinez and Davidson, 1997	70	ISB
Sp hox11/13b	135O12	Dobias *et al.*, 1996	125	JGI
Lv hox11/13b	235L15	HS	50	ISB
Lv kakapo	229D15	Brachyury target screen; HS	40	ISB
Sp kakapo	12G10	Brachyury target screen	50	ISB
Sp krox	163O19	Wang *et al.*, 1996	114	JGI
Lv krox	060B16	HS	56	ISB
Sp lim	108P4	Kawasaki *et al.*, 1999; HS	158	JGI
Lv lim	097A18	HS	50	ISB
Lv not	219N8	Peterson *et al.*, 1999	55	ISB
Lv notch	97A7	Sherwood and McClay, 1997; HS	71	ISB
Sp notch	191I13	Sherwood and McClay, 1997; HS	178	JGI
Lv nrl1	024M4	Brachyury target screen; HS	70	ISB
Sp orct	095C14	Brachyury target screen	57	ISB
Sp otx	006F13	Li *et al.*, 1997	160	JGI
Lv otx	229L5	HS	62	ISB
Sp pks	080H21	Dominant negative Notch screen	147	JGI
Lv pks	53J24	HS	38	ISB
Lv pmarl	170H13	Oliveri *et al.*, 2001; HS	62	ISB
Lv soxb	208L3	Kenny *et al.*, 1999; HS	70	ISB
Sp soxb	58L24	Kenny *et al.*, 1999; HS	150	ISB
Lv t-brain	192I24	Croce *et al.*, 2001; HS	48	ISB
Sp t-brain	31J8	Croce *et al.*, 2001; HS	130	ISB
Sp wnt8	041A8	Ferkowicz *et al.*, 1998; HS	135	JGI
Lv wnt8	183H12	Ferkowicz *et al.*, 1998	64	ISB

Note. Dominant Negative N screen performed by C. Calestani, A. Ransick, and E. Davidson (unpublished data). For Brachyury target screen, see Rast *et al.*, 2002. For cadherin overexpression screen see Ransick *et al.*, 2002. HS, homology screen of *S. purpuratus* library, using probe from another species. ISB, Institute for Systems Biology, Seattle, WA; JGI, Joint Genome Institute (DOE), Walnut Creek, CA.

TABLE 3
Genes and Gene Interactions in the Endomesoderm Network

		Spatial expression					Input affected by perturbation	
Abbrev.	Protein encoded	15–18 h	20–24 h	30–40 h	Class	Source of gene, information	Gene product	Data type
apobec	Cytidine deaminase	unk	EM	E	RNA editing enzyme	Brachyury target screens	Bra	Bra MASO
bra	Brachyury	EM	E	E, OE	Transcription factor	Peterson *et al.*, 1999; Gross and McClay, 2001	GataE	GataE MASO, CCRA, ECRA
							FoxA	FoxA MASO, CCRA,
							Eve	Eve MASO, ECRA
							Krox	Krox-En
							GataC	GataC MASO, CCRA
							Tcf	Cad MOE, CCRA
							Elk	Elk-En
							Otx[a]	Otx-En
capk	c-AMP protein kinase	unk	M	SMC	Protein-modifying enzyme	Brachyury target screens	Bra[b]	Bra MASO
cyclophilin	Cyclophilin, peptidyl-prolyl *cis-trans* isomerase	PMC	PMC	PMC	Protein-modifying enzyme	Zhu *et al.*, 2001; G. Amore *et al.*, unpublished	unk act	Pmar1-En, Pmar1 MOE
decorin	Decorin	unk	M	unk	Extracellular matrix	Dominant negative Notch screen	unk	
delta	Delta	m (8-18 h)	PMC(20), SMC(24)		Ligand	Zhu *et al.*, 2001	Tcf[b]	Cad MOE, CCRA
							unk rep	Pmar1-En, Pmar1 MOE
							Ub act[c]	Unk
							E(S)[d]	DnN MOE
dpt	d-Dopachrome tautomerase	unk	M	SMC	Enzyme	Brachyury target screens	Bra[b]	Bra MASO,
							Krl	Krl MASO
							E(S)[b,d]	DnN MOE
							Gcm	Gcm MASO
							Elk[b]	Elk-En
dri	DeadRinger	PMC	PMC, OE	OE	Transcription factor	40h-7h subtractive screen	unk rep	Pmar1-En, Pmar1 MOE
endo16	Endo16	EM	EM	E	Cell surface protein	Nocente-McGrath *et al.*, 1989	Otx[a]	Otx-En (Li *et al*, 1991; Yuh *et al*, 2001), ECRA
							Tcf[b]	Cad MOE
							Ui	ECRA
							Krl	Krl MASO
							GataE[b]	GataE MASO
ephx	p33/HEH epoxide hydrolase		PMC	PMC	Enzyme	Brachyury target screens	Bra[b]	Bra MASO
ets	Ets	Ma-Ub	PMC	PMC	Transcription factor	Kurokawa *et al.*, 1999	unk rep	Pmar1 MOE, Pmar1-En
eve	Even-skipped orthologue	E	E	Veg_1(E)	Transcription factor	Cadherin over-expression screen	Tcf	Cad MOE
							Krox	Krox-En
							Otx[b]	Otx-En
ficolin	Ficolin-like	PMC	PMC	PMC	TGFβ membrane BP	Incidental screen	unk act	Pmar1-En, Pmar1 MOE
							Dri	Dri MASO, Dri-En
							Bra[b]	Bra MASO
foxa	Hepatocyte NF 3 orthologue (FoxA)	EM	E	E	Transcription factor	Harada *et al.*, 1996; HS	FoxA	FoxA MASO, CCRA
							Otx	Otx-En
							GataE	GataE MASO
							GataC	GataC MASO
							Gsc[b]	Gsc MASO
							Tcf	Cad MOE
							E(S)[b,d]	DnN MOE
							Dri[b]	Dri MASO, Dri-En
							Eve	Eve MASO
							Elk	Elk-En
foxb	Winged helix factor	ND	E	E	Transcription factor	Luke *et al.*, 1997	Krox	Krox-En
							Tcf[b]	Cad MOE
							Eve	Eve MASO
							GataE	GataE MASO
							GataC	GataC MASO
							FoxA	FoxA MASO, CCRA
fvmo	Flavine mono-oxygenase	unk	M	SMC	Enzyme	Dominant negative Notch screen	E(S)[b,d]	DnN MOE
							Gcm	Gcm MASO

TABLE 3—*Continued*

Abbrev.	Protein encoded	Spatial expression 15–18 h	20–24 h	30–40 h	Class	Source of gene, information	Input affected by perturbation: Gene product	Data type
gatac	GataC	ND	M	ND	Transcription factor	Pancer *et al.*, 1999	Otx[b]	Otx-En
							GataE	GataE MASO
							Elk	Elk-En
							Tcf	Cad MOE
gatae	GataE	EM	EM	E	Transcription factor	Pancer *et al.*, 1999	Otx	Otx-En
							E(S)[d]	DnN MOE
							Tcf	Cad MOE
gcm	Glial cells missing	M	M	SMC	Transcription factor	Cadherin over-expression screen	Gcm	Gcm MASO
							Elk	Elk-En
							Tcf	Cad MOE
							E(s)[d]	DnN MOE
gel	Gelsolin-like	unk	EM	E	Actin-modulating protein	Brachyury target screens	Bra	Bra MASO
hmx	Homeodomain DBP	unk	unk	EM	Transcription factor	Martinez and Davidson, 1997	Otx[a,e]	Otx-En
							Krox	Krox-En
							Tcf	Cad MOE
							GataC	GataC MASO
							E(S)[d]	DnN MOE
							GataE	GataE MASO
hnf6	Hepatocyte NF 6 orthologue	unk	unk	unk	Transcription factor	K. Makabe, unpublished;HS	Pmar1	Pmar1-En, Pmar1 MOE
hox11/13b	Homeobox protein, Hox11/13b	Ub	Ub	E	Transcription factor	Dobias *et al.*, 1996; Martinez *et al.*, 1999	Krox	Krox-En
							Tcf	Cad MOE
							Otx[e]	Otx-En
kakapo	Calponin domain protein	unk	EM	E	Cytoskeletal protein	Brachyury target screens	Bra	Bra MASO
krox1	Krox	E	E	E	Transcription factor	Wang *et al.*, 1996	Otx	Otx-En
							Krox	Krox-En
							Tcf	Cad MOE
							Eve	Eve MASO
krl	Krüppel-like	EM	ND	ND	Transcription factor	Howard *et al.*, 2001	Tcf	Cad-MOE (Howard *et al.*, 2001)
lim1	LIM-1	Veg1, OE	OE	OE	Transcription factor	Kawasaki *et al.*, 1999	Tcf	Cad MOE
							Otx	Otx-En
							GataE[f]	GataE MASO
							Elk[b]	Elk-En
							Krox	Krox-En
msp130	MSP130	PMC	PMC	PMC	Cell surface protein	Parr *et al.*, 1990	Dri	Dri MASO, Dri-En
							unk act	Pmar1-En, Pmar1 MOE
msp130-like	MSP130-like	PMC	PMC	PMC	Cell surface protein	Parr *et al.*, 1990	unk act	Pmar1-En, Pmar1 MOE
							Bra[b]	Bra MASO
not	Not	unk	M	SMC	Transcription factor	Peterson *et al.*, 1999	Otx	Otx-En
							GataE[f]	GataE MASO
							GataC	GataC MASO
notch	Notch	Ma-Ub (protein)	EM	unk	receptor	Sherwood and McClay, 1997		
nrl-1	Neuralized-like-1	ND	M	PMC, M	Signaling intermediate	Brachyury target screens	GataE	GataE MASO
							Bra[b]	Bra MASO
							Otx[b]	Otx-En
							E(S)[d]	DnN MOE
orct	Organic Cation Transporter	unk	EM	EM	Transporter	Brachyury target screens	Bra	Bra MASO
otxβ1/2	Orthodenticle orthologue (β1/2 transcription unit)	unk	OE	E, OE	Transcription factor	Li *et al.*, 1997	Tcf	Cad MOE
							GataE	GataE MASO, CCRA, ECRA
							Otx[a]	Otx-En, CCRA, ECRA
							Krox	Krox-En, CCRA, ECRA
pks	Polyketide synthase	unk	M	SMC	Enzyme	Dominant negative Notch screen	E(S)[b,d]	DnN MOE
							Bra[b]	Bra MASO
							GataE	GataE MASO
							Gcm	Gcm MASO
							Elk[b]	Elk-En
pmar1	Paired-class homeodomain protein	m (6-12h), ND	ND	ND	Transcription factor	Oliveri *et al.*, 2002	Otx	Otx-En
							Tcf	Cad MOE
sm27	Spicule matrix protein-27	ND	PMC	PMC	Structural protein	Harkey *et al.*, 1995	unk act Dri	Pmar1-En, Pmar1 MOE Dri MASO, Dri-En

TABLE 3—*Continued*

Abbrev.	Protein encoded	Spatial expression 15-18 h	20-24 h	30-40 h	Class	Source of gene, information	Input affected by perturbation: Gene product	Data type
Sm30	Spicule matrix protein-30	ND	PMC	PMC	Structural protein	George *et al.*, 1991	Dri	Dri MASO, Dri-En
sm37	Spicule matrix protein-37	unk	PMC	PMC	Structural protein	Lee *et al.*, 1999		
sm50	Spicule matrix protein-50	PMC	PMC	PMC	Structural protein	Sucov *et al.*, 1987; Katoh-Fukui *et al.*, 1991	Ets	neg-Ets (Kurokawa *et al.*, 1999)
							Ets[b]	Pmar1-En, Pmar1 MOE
							Eve	EVE MASO
							Dri	Dri MASO, Dri-En
							Bra[b]	Bra MASO
							Gsc[b]	Gsc MASO
soxb1	Sox protein	Ma-Ub	Ec	Ec	HMG transcription factor	Kenny *et al.*, 1999	Krl	Krl MASO (Howard *et al.*, 2001)
sutx	Sulfotransferase	unk	M	SMC	Enzyme	Dominant negative Notch screen	Soxb1	Soxb1 MASO
							E(S)[d]	DnN MOE
							Gcm	Gcm MASO
tbr	T-brain	Ma-Ub, m (>8), Zy-PMC	PMC	PMC	T-box transcription factor	Croce *et al.*, 2001; HS	Ets	Neg-Ets (Kurokawa *et al.*, 2000)
							unk rep	Pmar1-En, Pmar1 MOE
							Bra[b]	Bra MASO
ui	Unknown	unk	unk	unk	Transcription factor	C.-H. Yuh *et al.*, unpublished	Tcf[b]	Cad MOE
							GataE	GataE MASO
wnt8	Wnt8	m (5), EM (10-15),	E	unk	Wingless family ligand	A. Wikramanayake & W. Klein, unpublished	Tcf	Cad MOE
							GataE[b]	GataE MASO
							Krox	Krox-En
							Otx	Otx-En

Note. If repressors and ubiquitous activators are deduced to be in the network but have not been identified, they are not listed. In the case of known genes, the references listed as a source contain the sequence from which probes were generated to screen for the corresponding BAC clone. When the sequence was obtained from a species related to *S. purpuratus*, the source of the probes used in this work was an *S. purpuratus* cDNA isolated by a homology screen; indicated by HS. Abbreviations: *Spatial Expression*: The expression patterns are indicated by abbreviations for each embryonic territory: Ma, maternal; Ub, ubiquitous; m, micromere; E, definitive endoderm; EM, veg_2 endomesoderm; M, veg_2 mesoderm; OE, oral ectoderm; PMC, primary skeltogenic mesenchyme; SMC, secondary mesenchyme cells of veg_2 origin; unk, unknown; ND, not detected. When the spatial pattern is known at a time not indicated at the head of the columns, the time is shown in parentheses. For example: m (6-12) indicates expression in the micromeres from 6 to 12 h after fertilization. *Data Type*: Experimental manipulations used for quantitative PCR are: antisense morpholino oligonucleotide injection (MASO); engrailed fusion mRNA injection (En); and mRNA overexpression (MOE). In addition, computational *cis*-regulatory analysis or sequence motif search (CCRA) and experimental *cis*-regulatory analysis (ECRA); e.g., gene transfer experiments with reporter constructs containing sequence from the BAC were used to verify input gene relationships. The individual abbreviations in each of these classes is listed: Cad MOE: Cadherin RNA overexpression (Logan *et al.*, 1999); DnN MOE: Dominant negative Notch overexpression (Sherwood and McClay, 1999); Krox-En: Krox-engrailed fusion RNA injection; Otx-En: Otx-engrailed fusion RNA injection (Li *et al.*, 1999); GataE MASO: GataE antisense morpholino oligonucleotide injection; Gcm MASO: Gcm antisense morpholino oligonucleotide injection; FoxA MASO: FoxA antisense morpholino oligonucleotide injection; Pmar1 MOE: Pmar1 RNA overexpression; Pmar1-En: Pmar1-engrailed fusion RNA injection; Krl MASO: Krl antisense morpholino oligonucleotide injection (Howard *et al.*, 1999); H13b MASO: Hox11/13b antisense morpholino oligonucleotide injection; Elk-En: Elk-engrailed fusion RNA injection; GataC MASO: GataC antisense morpholino oligonucleotide injection; Bra-MOE: Brachyury overexpression; Bra MASO: Brachyury antisense morpholino oligonucleotide injection; Dri MASO: DeadRinger antisense morpholino oligonucleotide injection; Neg Ets: DNA binding domain of Ets (Kurakawa *et al.*, 1999).

[a] Otx-En effect cannot be rescued by GataE MOE.

[b] Effect concluded to be indirect or irrelevant: see note regarding specific data in Appendix.

[c] Ubiquitous activating input inferred from ubiquitous expression in presence of Pmar1 MOE and Pmar1-En.

[d] Role of an Enhancer of Split gene [E(S)] inferred from data in other systems; which of several E(S) type genes plays this role in *S. purpuratus* is not yet known.

[e] Otx-En effect cannot be rescued by Krox MOE.

[f] GataE MASO effect cannot be rescued by Otx MOE.

of the effect of cadherin mRNA on *eve* expression (expression of this gene is reduced > 95% throughout) suggests a direct Tcf input. In the case of the *bra* gene, Tcf sites that bind a nuclear factor are found in the relevant *cis*-regulatory element and this gene is also very strongly sensitive to cadherin mRNA injection (Gross and McClay, 2001; Appendix). Irrespective of whether these inputs are in fact all direct, here, we see why introduction of a dominant negative form of Wnt8 (A. Wikramanayake, unpublished data), or interference with β-catenin nuclearization by introduc-

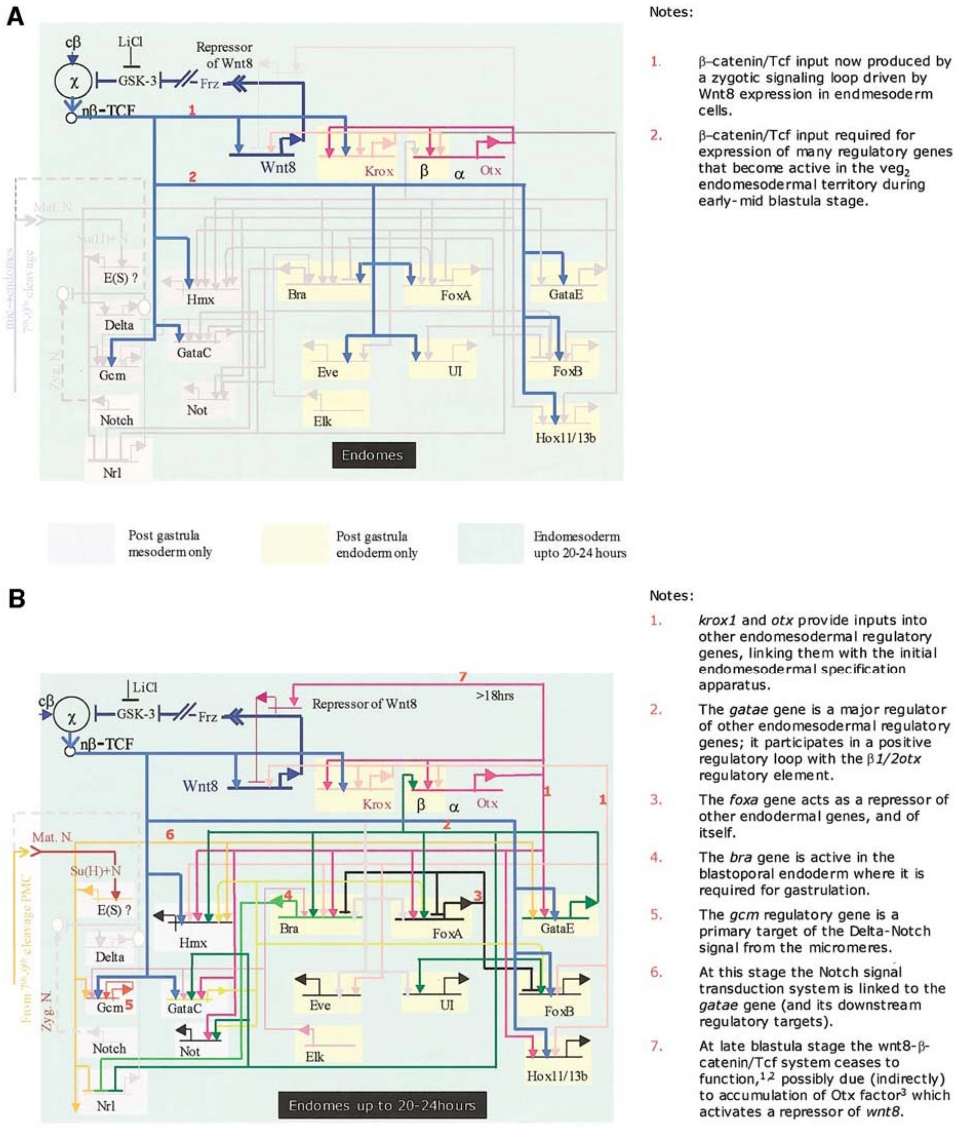


FIG. 4. Views from the nuclei of the veg_2 endomesoderm from about eighth cleavage to the 20- to 24-h mesenchyme blastula stage. (A) β-Catenin/Tcf inputs into regulatory genes of the endomesoderm. Every target gene shown except *wnt8* itself encodes a transcriptional regulator. (B) Network of endomesodermal regulatory genes that become active in this period. References: [1]Logan *et al.*, (1999); [2]A. Wikramanayake *et al.*, (2002); [3]Yuh *et al.*, (2002); C.B.L. and E.H.D., unpublished data.

tion of intracellular cadherin, entirely wipes out endomesoderm specification, producing just a hollow ball of ectoderm-like cells.

Something of the ensuing circuitry is evident in Fig. 4B, although much more remains to be learned. The first point that this portion of the network explains is how the initial

specification functions are linked mechanistically to the regulation of the genes that specify the endomesodermal state: in addition to the β-catenin/Tcf inputs, either the *krox1* gene or the *otx* gene or both, provide inputs to almost all of them (Note 1). Among these genes, *gatae* is especially important. Figure 4B shows not only that the output from the *gatae* gene contributes to activation of six other transcriptional regulatory genes of the endomesoderm, but that its input also feeds back on the *β1/2otx cis*-regulatory system (Table 1; Yuh *et al.*, 2002). This permanently locks in the endomesoderm specification state, for now every regulatory gene in the network of Fig. 4B, except those specific to the N signaling system, has inputs from genes in the *krox1-otx-gatae* feedback loop (Note 2 of Fig. 4B). These genes have diverse kinds of function, as indicated in Notes 2–4 of Fig. 4B, not even considering the downstream differentiation genes that some of them control, as we briefly discuss below.

Those veg_2 cells that lie in direct proximity to the large micromere lineage which gives rise to the skeletogenic mesenchyme receive the additional Delta signaling input. We have not yet been able to determine which of the several *enhancer of split*-like genes in *S. purpuratus* transduce this signal, but we assume one of them does [E(S) in these figures]. Perturbation experiments carried out by introducing an mRNA encoding a dominant negative N form (the N extracellular domain; Sherwood and McClay, 1999) have revealed two regulatory genes that are probably direct targets of the N signal transduction pathway, *hmx* and *gcm* (Table 1; Note 5 of Fig. 4B). A dramatic whole-mount *in situ* hybridization image of *gcm* expression in a one-deep ring of veg_2 cells directly abutting the skeletogenic precursors can be seen in Fig. 2E of Ransick *et al.* (2002). The *gcm* gene also locks itself on, once activated (Appendix). Later it serves as a regulator of the differentiation of pigment cells, one of the veg_2 mesodermal cell types (A.R. and E.H.D., unpublished data).

The N signal transduction system also provides a positive input to the *gatae* gene. This gene is a critical regulator. Its *cis*-regulatory system would appear to integrate inputs from both N/E(S) and Wnt8/Tcf signal transduction systems (Note 6; Fig. 4B). When initially activated, most of the regulatory genes included in Fig. 4B that later serve as dedicated endodermal regulators are expressed across the whole of the veg_2 endomesoderm (see Table 1). At least one reason for this may be that most of these genes are controlled in part by GataE inputs.

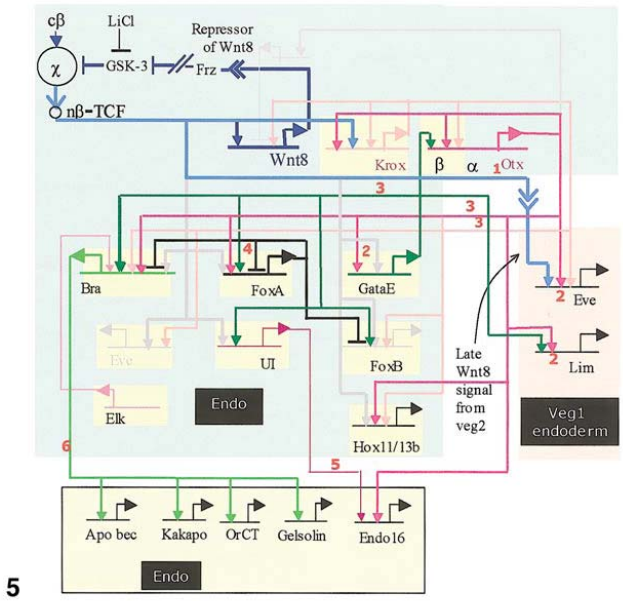
During the late blastula stage, β-catenin disappears stochastically from the nuclei of the skeletogenic micromeres and veg_2 endomesodermal domains (Logan *et al.*, 1999). The mechanism causing this is unknown. However, some measurements listed in Table 1 on embryos expressing an Otx-Engrailed fusion may provide an explanation: at mid-late blastula stage in embryos expressing this obligate repressor of Otx target genes, the level of *wnt8* transcripts increased many fold. This implies that an *otx* gene product must at this time normally activate a repressor of the *wnt8* gene. The *β1/2otx* transcription unit becomes active at the beginning of the period covered by Fig. 4 (C.B.L. and E.H.D., unpublished data). Its activation could thereby account for the transience of the β-catenin/Tcf input, by causing the Wnt8 signal now needed to drive the nuclearization system to diminish. But as we can see in Fig. 4B, by the time of its demise, a network of stable intergenic interactions has been installed, so that the inputs used earlier to initiate transcriptional specification are no longer needed.

Views from the Nuclei of the Definitive Endoderm just before Gastrulation

By 24 h, the concentric mesodermal and endodermal territories have been established within the vegetal plate, the skeletogenic cells originally occupying the center of this domain having now ingressed into the blastocoel. This is only a few hours before the onset of gastrulation, where most of our studies currently end. Our knowledge of this period is less complete than for earlier stages, if only because a large part of the regulatory activity is now shifted to control of differentiation gene batteries. We have only a small set of the target differentiation genes in our hands, and only a glimpse of their immediate regulatory inputs. There may be additional regulatory genes called into play at this stage as well, of which we are yet unaware: the cadherin and dominant negative N screens we used to discover endomesodermal regulatory genes pertain only to earlier specification stages. Only the *brachyury* perturbation screens (Table 1; Rast *et al.*, 2002) illuminate this later period of development.

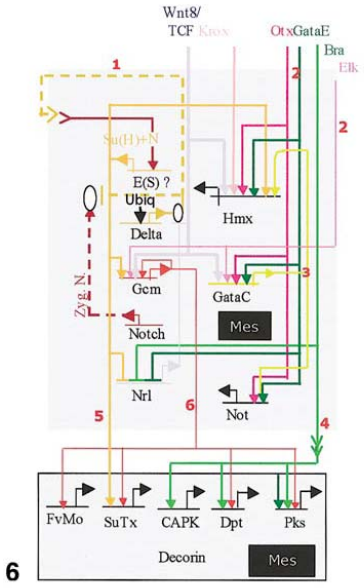
The view from the definitive endoderm nuclei in the period between 24 h and gastrulation is shown in Fig. 5. At right is indicated the late specification of the veg_1 endoderm, now ongoing (Ransick *et al.*, 1998; Logan and McClay, 1997). Here, the same Wnt8-β-catenin/Tcf system as was required much earlier within the veg_2 domain is reactivated, but in different cells (Note 1 in Fig. 5). Two regulatory genes that are active at this time in the prospective veg_1 endoderm are *eve* and *lim* (Note 2); an *in situ* hybridization image of *eve* expression in the veg_1 cells lying immediately above the vegetal plate can be seen in Fig. 2D of Ransick *et al.* (2002). By this time, *eve* is no longer expressed in the veg_2 endoderm; the *lim1* gene is never expressed in the veg_2 domain (Table 3). The *eve* and *lim* genes also receive inputs from the *krox1* and *otx* genes, respectively, a similar role to that the *krox1* and *otx* genes played earlier, in veg_2 endomesoderm specification.

Within the veg_2 endoderm cells, the *gatae* gene continues to provide regulatory input into many other genes encoding endodermal transcription factors (Note 3), according to antisense *gatae* morpholino oligonucleotide results (Table 3; Appendix). The repressive *foxa* interrelations set up earlier probably serve spatial control functions (Note 4). The *foxa* gene is expressed in endomesoderm, and later in the archenteron (Table 3), particularly the anterior regions. But *foxb* and *bra*, which are targets of FoxA repression, are expressed only at the posterior end of the archenteron or in



Notes:

1. β-Catenin/Tcf transcriptional inputs appear now in prospective veg_1 endoderm, probably induced by Wnt8 signals from edge of veg_2 endoderm.
2. Result is activation of regulatory genes in the veg_1 endoderm.
3. Expression of endodermal regulatory genes is maintained by intra-endodermal regulators transcribed from *gatae, krox, β1/2otx* genes.
4. *bra* gene is not expressed in the anterior endoderm, due to *foxa* repression.
5. Otx directly controls at least one differentiation gene, *endo16*.
6. *bra* directly controls a set of differentiation genes



Notes:

1. Delta expression appears in definitive veg_2 mesodermal domain. While the *notch* gene may be expressed at low levels within this domain in response to the Delta signal, its expression is not required for delta expression.
2. GataE and Elk inputs could persist from prior period when these genes are expressed in endomesoderm.
3. *gatac* cross regulates other mesodermal regulatory genes.
4. Several mesodermal differentiation genes are dependent on Bra input for expression; since *bra* is not expressed in mesoderm, this implies an indirect effect mediated by intercellular signaling.
5. Expression of several downstream differentiation genes is depressed by introduction of mRNA encoding dominant negative N.
6. *gcm* controls pigment cell genes.

FIG. 5. View from the nuclei of veg_2 and veg_1 endoderm after 24 h and to the beginning of gastrulation.
FIG. 6. View from the nuclei of the veg_2 mesoderm after 24 h and to the beginning of gastrulation. For open ovals, see legend to Fig. 2.

the blastoporal area (see Table 3 for references). Direct evidence for a causal rather than coincidental relation is that expression of a *bra cis*-regulatory construct spreads upward into the archenteron if introduced into embryos also bearing a *foxa* morpholino antisense oligonucleotide (R.A.C., unpublished data).

Finally, it is important to note that *bra* is likely a direct controller of downstream differentiation genes that are expressed in the same cells as it is (Note 5; for evidence, see Rast *et al.*, 2002). Similarly, the *endo16* gene, which encodes a polyfunctional extracellular protein later secreted into the midgut domain, is a direct target of two other regulators in the network, αOtx and Ui (Note 6; Yuh *et al.*, 1998, 2001). These observations show how control of cell type functionality is linked directly into the regulatory network.

View from the Nuclei of veg$_2$ Mesoderm Cells just before Gastrulation

Figure 6 provides a similar picture for the veg$_2$ mesoderm domain just before gastrulation and delamination of its mesenchymal cell types. At this stage, the prospective mesodermal cells display processed intracellular N (in *L. variegatus*, another species that develops very similarly; Sherwood and McClay, 1999). But right after skeletogenic mesenchyme ingression, all prospective mesoderm cells also activate expression of the *delta* gene (at 24 h in *S. purpuratus*; Sweet *et al.*, 2002; P.O. and E.H.D., unpublished data). As it becomes active in the veg$_2$ mesoderm, the *delta* gene is silenced in the skeletogenic mesenchyme, once these cells ingress. The significance of the late expression of *delta* in the definitive mesoderm is unknown, nor are the inputs to the *delta* gene that cause this expression known (Note 1 of Fig. 6). A positive input from the N signal transduction system was required earlier for the activation of *gcm* and *hmx*, two genes encoding transcription factors which are expressed in the definitive mesoderm. This input could perdure for these genes into the pregastrular period of Fig. 6.

Other inputs into mesodermal specification genes are presented by transcription factors that were generated earlier in the endomesoderm (Note 2; *viz* GataE, Elk, Otx; see Table 3; Yuh *et al.*, 2002). An internal network is set up, in which the *gatac* gene is particularly important because it cross-regulates the *hmx* and *not* genes (Note 3). Finally, we see several inputs into mesodermal differentiation genes. One of those is certainly indirect: a number of such genes respond to perturbations of *bra* expression (three are shown here), and yet Bra protein never appears in any cells which give rise to mesoderm (Note 4; this was shown by Gross and McClay, 2001). The implication is that a heretofore undetected signal emanates from the blastoporal area where the *bra* gene is expressed, and that this signal affects expression of these genes in the veg$_2$ mesoderm. In this respect, the model is clearly yet incomplete, since the transcription factor that transduces this signal is not known. Two of the known mesodermal differentiation genes are also affected by N signaling (Note 5; see Table 1), probably indirectly. The *pks* gene, which encodes a pigment cell enzyme, is likely to be a direct target of the *gcm* gene, since *gcm* appears to control pigment cell differentiation (A.R. and E.H.D., unpublished data). This is also true for the *fvmo* gene. The expression of the *pks* gene in the same cells that express the *gcm* gene is affected by a morpholino antisense oligonucleotide that blocks Gcm translation (Note 5; Appendix). Even though the regulatory linkages to differentiation genes remain largely obscure, it is clear that the veg$_2$ mesoderm, like the veg$_2$ endoderm, has by now entered on a terminal stage of cell type differentiation.

DISCUSSION

In the sea urchin embryo, the territorial specification process is followed by morphogenesis: the archenteron invaginates and is divided into its tripartite domains; the skeleton is laid down; the coelomic pouches, muscle bands, and pigment cells are formed in their appropriate positions; and so forth. Other gene regulatory networks will be necessary to understand these later processes in the terms of which we have now begun to understand the pregastrular process of endomesodermal specification. Additional signaling inputs will clearly be involved (McClay and Logan, 1996), as will additional regulatory genes, and also some of the same genes, perhaps under the control of different *cis*-regulatory modules. But, although development of the embryo moves on beyond specification, in terms of its *cis*-regulatory logic, the network considered here has a beginning and an end. It begins with a rather crude set of spatial differences, due to maternal cytoplasmic functions that are distributed to the appropriate blastomeres in the course of the invariant early cleavages. Their purpose is just to provide the first spatially differential regulatory inputs to the zygotic transcriptional apparatus (Davidson, 1990, 2001). Though without these initial inputs endomesodermal specification cannot occur, the actual development of the endomesoderm depends entirely on the operation of the zygotic regulatory apparatus.

Once its early interactions are instituted, the endomesoderm specification network is so organized that it locks in its successive regulatory states and proceeds inexorably forward. The end, or the termini, or periphery of this particular network is the activation of genes encoding differentiation proteins. We can see that the network is not very deep. From activation of the *brachyury* gene, for instance, to activation of its endodermal differentiation gene targets, is probably only one step, and the same probably goes for the *gcm* gene and its pigment cell targets, such as the *pks* gene. The *bra* and *gcm* regulatory genes are themselves only one or two linkages downstream from the regulatory genes that interpret the initial specification cues. So, much of the organization of the network is "sidewise": it is equipped with cross-regulations, autoregulations, feed-

back loops, and other devices that endow it with forward progress and stability.

Though it is provisional and incomplete in almost every respect, though it yet contains only 42 genes, and though it undergoes continuous revision, the endomesodermal network that we show in Figs. 2–6 already has very considerable explanatory power. Among the aspects the network already reveals are:

- What the initial cues are used for.
- Why in zygotic regulatory terms the veg_2 lineage becomes an endomesodermal domain.
- How the zygotic activities of the micromere lineage are confined thereto.
- How the signal transduction systems that affect endomesoderm specification are linked into the zygotic regulatory apparatus.
- How endodermal regulatory genes are set in action.
- How the states of specification become stabilized.

The network illuminates these things, and will provide real explanations for them when it is no longer provisional, that is when its linkages have been verified at the level of *cis*-regulatory function and genomic sequence. Most important, the explanations that emerge are in the appropriate terms. From the time of Boveri, it has been inescapable that deep explanations of embryonic development must ultimately be couched in terms of genomic properties. The relevant properties are of course the heritable, genomic *cis*-regulatory sequence code for development. The network in Fig. 2 will lead directly to a structure/function level understanding of how DNA sequence is causally responsible for a robust and evolutionarily ancient developmental specification process.

APPENDIX

QPCR Data Relevant to Endomesoderm Network[a]

Perturbation Ct:[Ub-C]-[Ub-Exp][a]

Gene	Perturbation	12-16 h	18-21 h	24-27 h	30-36 h	Data of:
apobec	Bra MASO			-3.8		J. Rast
bra	GataE MASO		-2.8, -2.4/ -2.8, -3.2	-1.6,-0.8/ -3.3,-3.2/ -2.1,-2.2		P.-Y. Lee
	Krox-En	-2.0, -3.3/	-9.0, -9.0/	-5.2,-5.5/ -5.0		C. Livi
	Eve MASO			-3.3		A. Ransick
	FoxA MASO			NS	+1.6	P. Oliveri
	Cad MOE	NS/NS	-4.7/-2.1	-4.4/-3.0		A. Ransick & T. Minokawa
	Elk-En			-3,2, –2.2, -4.1/		M. Arnone
	GataC MASO			-1.9		P. Oliveri J. Rast
	Otx-En		-2.5/-3.4/ -4.3/-8.2	-2.4/-2.0/ -5.2/-4.6		A. Ransick & T. Minokawa
capk	Bra MASO[1]			-1.7		J. Rast
cyclophilin	Dri MASO		-3.6			G. Amore
	Pmar1 MOE	+3.3		+5.5		P. Oliveri
	Pmar1-En	NS		+3.4		P. Oliveri
delta	Cad MOE[15]	-3.3/-3.9	-1.6/NS	NS/-1.9		A. Ransick & T. Minokawa
	Pmar1 MOE	+1.8/+3.2		+2.6/+2.6		P. Oliveri
	Pmar1-En	+1.8/+2.0		+3.1/+2.0		P. Oliveri
dpt	Bra MASO[1]			-6.8		J. Rast
	DnN MOE[8]		-1.7			C. Calestani
	KrlMASO	-3.8,-3.8		-3.1,-2.8		C. Livi
	Elk-En[5]			-2.1,-2.2,-4.6		M. Arnone
	GCM Maso		-2.6	-5.7/-3.6		A. Ransick
dri	Pmar1 MOE	+3.6		+6.7		P. Oliveri
	Pmar1-En	+2.4		+7.2		P. Oliveri
endo16	Otx-En	-1.6	-4.2	-3.1/-4.3/ -6.6		A. Ransick & T. Minokawa
	Cad MOE[10]	-2.9/NS	-6.2/-2.3	-5.2/-2.2		A. Ransick & T. Minokawa
	Krl MASO	-2.7,-2.7/NS		-2.5,-2.5/ -4.0,-3.3		C. Livi
	GataE-MASO		NS/NS	-2.8, -3.5/ -2.7, -1.8		P.-Y. Lee
ephx	Bra-MASO[1]			-3.8		J. Rast
ets	Pmar1 MOE	NS		+5.4		P. Oliveri
	Pmar1-En	NS		+4.5		P. Oliveri

APPENDIX—*Continued*

Gene	Perturbation	12-15	18-21	24-27	30-36	Author
eve	Krox-En	NS	-4.4, -4.3	-2.8, -4.3/ -5.3		C. Livi
	Cad MOE	-6.3, -4.9/ -3.9	-5.2, -4.0/ -3.1	-5.5, -4.3/ -4.6		A. Ransick & T. Minokawa
	Otx-En[12]			-2.9/-1.3		A. Ransick & T. Minokawa
ficolin	Bra MASO[1]			-1.7		J. Rast
	Dri MASO	NS/-3.6/ -2.7/-4.6	NS/NS/ -1.8/-2.5			G. Amore
	Dri-En		-4.6/-3.0			G. Amore
	Pmar1 MOE	+4.0		+5.9		P. Oliveri
	Pmar1-En	NS		+3.1		P. Oliveri
foxa	Otx-En		NS/NS/-2.8	-3.3/-1.4/ -4.1/-2.8		A. Ransick & T. Minokawa
	Dri-En[19]		-4.6/-3.0			G. Amore
	Dri MASO[19]			NS/NS	-2.3/-2.0/ -2.0/-2.0/ -3.0/-2.0/ 2.0	G. Amore
	Gsc MASO[2]			NS/NS	-3.2/-1.5/ -3.0/-3.0/ 3.8/-3.8	G. Amore
	Eve MASO			-2.6		A.Ransick
	GataE MASO		-2.1, -1.6/ NS/-1.3, -1.8	-2.9, -1.4, -2.0/-3.4, -3.1		P.-Y. Lee
	FoxA MASO			+2.3/NS, +1.6	+1.4, +1.8	P. Oliveri
	Cad MOE	-0.7, -2.0, -1.5/-2.2	-1.4, -4.1, -4.6/-2.1	-2.6, -5.0, -4.6/-4.9		A. Ransick & T. Minokawa
	GataC MASO			-2.7		P. Oliveri & J. Rast
	DnN MOE[3]		-2.5/NS			C. Calestani
	Elk-En[4]			-3.0, -3.7, -4.7		M. Arnone
foxb	Krox-En		-5.0	-5.7/-4.2		C. Livi
	GataE MASO			-6.4, -4.4/ -4.8, -5.2		P.-Y. Lee
	Eve MASO			-5.3		A. Ransick
	FoxA MASO			+1.41, +1.94		P. Oliveri
	Cad MOE[17]		NS	-5.1/-3.3		A. Ransick & T. Minokawa
	GataC MASO			-1.9		P. Oliveri & J. Rast
fvmo	DnN MOE[7]		-4.2/-3.0/ -2.5			C. Calestani
	Gcm MASO			-4.2		A. Ransick
gatac	Elk-En			6.0, -5.2,-5.7		M. Arnone
	Cad MOE	-6.1	-2.5	-3.6/-6.4		A. Ransick & T. Minokawa

APPENDIX—*Continued*

Gene	Perturbation	12-15 h	18-21 h	24-27	30-36	Author
	GataE MASO		NS/NS/NS/ NS, NS	-2.4/-2.3/NS		P.-Y. Lee
	Otx-En[18]		NS	-2.1,/-3.3		A. Ransick & T.Minokawa
gatae	Cad MOE	NS/NS	-5.1/-3.1	-4.7/-3.6		A. Ransick & T. Minokawa
	DnN MOE		-2.4/-1.4			C. Calestani
	Otx-En		-3.3, -2.2/ -2.8, -3.0/ -3.2	NS/-1.5, -2.4/-2.4/ -4.1/-2.4		A. Ransick & T. Minokawa
gcm	Cad MOE	-6.3, -5.1, -5.2/-3.3	-5.2, -6.1, -4.6/-4.3	-5.0, -5.7/ -5.5		A. Ransick & T. Minokawa
	DnN MOE		-3.5/-3.1			C. Calestani
	Gcm MASO			-2.0/-1.7		A. Ransick
	Elk-En			-4.3, -3.0, -6.4		M. Arnone
gell	Bra MASO			-4.4		J. Rast
gsc	Dri MASO	-5.0/-7.0	-7.0/8.6			G. Amore
	Dri-En		-3.6/-2.7			G. Amore
hmx	Krox-En	NS	-6.8, -7.5	-7.2, -7.7/ -5.9		C. Livi
	Cad MOE	NS	-2.8	-3.3		A. Ransick & T. Minokawa
	DnN MOE		-3.4/-6.0			C. Livi
	GataC MASO			-1.7		P. Oliveri J. Rast
	Otx-En		-3.0/-3.8/ -6.7/-7.7	-3.6/-3.2/ -9.7/6.3/-7.4		A. Ransick & T. Minokawa
	GataE MASO		-2.8, -2.8/ -3.6/-3.6	-2.6, -4.0/ -3.9, -3.1		P.-Y. Lee
hnf6	Pmar1 MOE	-4.4/NS		-7.6/-3.5		P. Oliveri
	Pmar1-En	-4.5/NS		-8.0, NS, -2.8		P. Oliveri
hox11/13b	Krox-En	-7.6	-6.2	-3.2		C. Livi
	Cad MOE	-5.4	-5.6/-2.7	-5.2/-3.1		A. Ransick & T. Minokawa
	Otx-En		-4.8/-4.3/ -6.5	-3.0/-4.3/ -6.6/-8.8		A. Ransick & T. Minokawa
kakapo	Bra MASO			-2.0		J. Rast
krox1	Krox-En	-4.0, -4.8	-7.8, -8.0	-7.4, -5.0/ -5.3		C. Livi
	Cad MOE	-5.9/-4.8	-4.6/-7.8	-5.1/-14.3		A. Ransick & T. Minokawa
	Eve MASO			-1.9		A. Ransick
	Otx-En	-2.3/-1.6	NS/NS/NS	-2.5/-3.3/NS		A. Ransick & T. Minokawa

APPENDIX—*Continued*

Gene	Perturbation	12-15 h	18-21 h	24-27	30-36	Author
lim1	GataE MASO		-1.6/-2.3, -1.9/NS, NS/ NS, NS	-2.6/-2.2, -2.3/-2.3		P.-Y. Lee
	Cad MOE	NS	-2.5/NS	-5.0/NS		A. Ransick T. Minokawa
	Elk-En[6]			-6.3, -3.7, -6.5		M. Arnone
	Otx-En		-2.5/NS	-3.3/-5.4		A. Ransick & T. Minokawa
	Krox-En		-3.1			C. Livi
msp130-like	Bra MASO[1]			-4.3		J. Rast
	Pmar1 MOE	+3.2		+5.8		P. Oliveri
	Pmar1-En	+2.4		+4.9		P. Oliveri
msp130	Dri MASO	-3.6/-5.1	-2.0/-2.5/ -3.0			G. Amore
	Dri-En		-3.6/-1.9			G. Amore
	Pmar1 MOE	+3.1		+6.4		P. Oliveri
	Pmar1-En	+1.6		+4.7		P. Oliveri
not	GataE MASO		NS/-1.9, -1.2/NS/NS/ NS, +2.8	-2.2, -2.1/ -2.4,-2.3		P.-Y. Lee
	Otx-En	-3.6	-5.0	-2.1/-3.3		A. Ransick & T. Minokawa
	GataC MASO			-1.7/+1.2		P. Oliveri & J. Rast
nrl1	GataE MASO		+2.3, +0.8, +1.8	+2.1, +1.6, +0.3/+2.3, +1.8		P.-Y. Lee
	Bra MASO[1]			+2.4, +4.3		J. Rast
	DnN MOE		+3.5/NS			C. Calestani
	Otx-En[3]		NS	+3.5		A. Ransick & T. Minokawa
orct	Bra MASO			-6.7		J. Rast
otxβ1/2	Krox-En	-5.6	-4.6	-3.6/-4.3		C. Livi
	GataE MASO		-2.5, -3.1/ -3.6/-3.6	-1.8, –0.4, -1.2/-3.0, -3.7/-2.6		P.-Y. Lee
	Cad MOE	NS	+10.3/+1.7/ +2.1	NS/NS		A. Ransick & T. Minokawa
	Otx-En	-3.8/-3.7	-3.2/-4.1/ -3.3	-2.1/-2.3/ -3.2/-3.8		A. Ransick & T. Minokawa

APPENDIX—*Continued*

Gene	Perturbation	12-15	18-21	24-27	30-36	48	Author
pks	GataE MASO		-2.2/-3.2	-3.4/-1.6, -1.1/-3.2, -3.6			P.-Y. Lee
	Gcm MASO		-6.1	-5.2/-3.3/-12.3			A. Ransick
	Bra MASO[1]			-2.6			J. Rast
	DnN MOE[7]		-3.3/-2.5/ -2.7, -3.1				C. Calestani
	Elk-En[7]			-7.1, -5.2 -6.5			I. Arnone
pmar1	Cad MOE	-4.0 (6 h)/ -3.0 (6 h)					A. Ransick & T. Minokawa
	Otx-En	-3.1, -2.9/ -2.8					A. Ransick & T. Minokawa
sm27	Pmar1 MOE	NS		+4.0			P. Oliveri
	Pmar1 En	NS		+2.7			P. Oliveri
pm27	Dri MASO		-3.2/-50	-4.0/-4.3			G Amore
	Dri-En			-3.9/-3.9			G. Amore
sm30	Dri MASO		-6.2/-5.4/ -5.6/-5.6		-7.5/-5.0/-6.0	-5.0/-3.2	G. Amore
	Dri-En				-6.2/-5.4/-5.6/-5.6		G. Amore
sm50	Dri MASO	-3.0/-2.1/-4.8/NS	-2.5/-4.0/-4.8/NS		-4.0/-3.3/NS -1.3/4.0/3.4		G. Amore
	Eve MASO			-3.8			A. Ransick
	Bra MASO[1]			-2.2			J. Rast
	Dri -En		-3.2/-2.3				G. Amore
	Gsc MASO[2]				NS/-3.0-3.0		G. Amore
	Pmarl MOE[11]			+4.4/+2.3			P. Oliveri
	Pmar1-En[11]			+2.5/NS			P. Oliveri
soxB1	SoxB1 MASO			+3.6,+3.3/ +3.0,+3.4	+3.8,+3.9/ +4.3,+4.3		C. Livi
sutx	DnN MOE			-2.7/-2.6			C. Calestani
	Gcm MASO		-4.3	-6.7			A. Ransick
tbr	Bra MASO[11]			-1.8			J. Rast
	Pmar1 MOE	+2.0/+1.1		+5.9/+3.4			P. Oliveri
	Pmar1-En	+1.7/+1.1		+4.7/+2.6			P. Oliveri
ui (via endo16)	Cad MOE[9]	-2.9/NS	-6.2/-2.3	-5.2/-2.2			A. Ransick & T. Minokawa
	GataE MASO		NS/NS	-2.8, -3.5/ -2.7, -1.8			P.-Y. Lee
wnt8	Krox-En	-4.2 (6 h), -2.0, -3.7	-3.2	NS			C. Livi

APPENDIX—*Continued*

Gene	Perturbation	12-15	18-21	24-27	30-36	48	Author
	Cad MOE	-4.5 (6 h)/ -5.4/-5.3	-3.9/-2.8	-5.3/-5.0			A. Ransick
	Otx-En	NS/NS	+2.6/+7.7/+ 2.1	NS/NS/NS			A. Ransick & T. Minokawa
	GataE MASO[14]		+1.5/+2.3/ +1.4	NS,NS/NS, NS/NS, NS			P.-Y. Lee

Abbreviations:

MASO: Morpholino-substituted Antisense OligoNucleotide
MOE: Messenger RNA Overexpression
EN: Engrailed repressor domain

Perturbation:	Genes not affected or probably affected only indirectly, as indicated in notes below
Krox-En	*endo16, pks, dpt, gcm, not, pmar1, tbr, nrl-1, αotx, foxa, soxb1, dec, gatae, elk*
GataE MASO	*endo16, hmx, dri, dec, eve, gatae, gcm, hox11/13b, krox1, wnt8, αotx, elk*
FoxA MASO	*endo16, gatae, krox1, eve, nrl, hmx, αotx, β1/2otx, wnt8, pks, hox11/13b*
Cad MOE	*dec, endo16, nrl, tbr, aotx, foxb*
Bra MASO	*nrl, ephx, ficolin, sm50, msp130-like, bra, tbr, capk, dpt, pks, dec, delta*
Dri MASO	*lim1, bra, eve, ets, dri (in pmc's), gatae, pmar1, tbr,ficolin,hnf6 ,delta,dec, gcm,fvmo,pks*
DnN MOE	*foxa, dec, hmx, gatac, elk, tbr, foxb, wnt8, β1/2otx, krox1, hox11/13b, eve, not, bra, fvmo*
Dri-En	*sm50*
Gsc MASO	*bra, eve, gatae*
Elk-En	*delta, elk, hnf6*
GataC MASO	*delta, dpt, gatac, gatae, gcm, capk, elk, pks, not*
Otx-En	*dec, delta, elk, foxb, gcm, αotx, tbr, eve, pks, foxb, gatac*
Pmar1 mRNA	*sm30, foxa, krox1, elk, wnt8*
Pmar1-En	*ephx, sm30, foxa, krox1, elk, wnt8*
Gcm MASO	*hmx, dec, gatac, nrl, delta, not, elk, capk, notch, krox, eve, bra, foxa, foxb*
Eve MASO	*wnt, krl, Tbr, gatae, gcm*

(a) Ub, ubiquitin mRNA control used for internal standardization; C_T, cycle no. at threshold; C, control embryo sample; Exp, sample from perturbed embryos. The quantity of transcript at any point in the reaction is $I(1.94^{C_T})$, where I is the initial amount in the reaction mixture. A positive number means the number of transcripts is increased by the perturbation; a negative number means the number of transcripts is decreased. Data are listed that are considered significant where significant means more than three-fold increase or decrease from control transcript levels (normalized C_T difference from control is <-1.6 or >+1.6). Smaller effects are shown as "NS", except where individual NS data are included together with other significant measurements to display scatter or inconsistencies amongst different batches of cDNA. Commas separate replicate measurements in the same cDNA batch; slashes indicate different batches of cDNA from independent experiments.

[1]Must be affected indirectly via intercellular signaling because Bra is not present in cells where these genes are active.
[2]Affected indirectly via cell signaling since at 30 h Gsc exprerssed only in OE.
[3]Thought to be indirect input, via GataE.
[4]Probably via GataC.
[5]Possibly indirect, via Bra.
[6]Probably in OE where both Elk and Lim1 are expressed then; not relevant to endomesoderm network.
[7]Indirect via Gcm; gene continues to be expressed as does *gcm* after N signaling phase is over.

APPENDIX—*Continued*

[8]Likely to be indirect: could be via E(S)→GataE→Bra as it is a small effect; or via *gcm*.
[9]Inputs are inferred from effects on *endo16*; see note 10 under *endo16*.
[10]There are no necessary GataE or Tcf sites in *endo16* (Yuh *et al.*, 2000). There is no Tcf input into α-Otx transcription unit, which is unaffected by Cad MOE. α -Otx is known to provide a direct input into *endo16* (Yuh *et al.*, 2002). Since the only significant driver inputs in *endo16* are Ui and α -Otx, the GataE input is indirect via Ui; the Cad MOE effect is probably via GataE rather than direct on *ui*, because the effect is too late (>18 h) for a direct Tcf input.
[11]Probably via *ets*; based on ECRA (Kurokawa *et al.*, 1999; Oliveri *et al.*, 2002).
[12]Indirect, because rescued by *krox* MOE.
[14]Weak effect 18 h only, probably not significant, since *wnt8* expression is disappearing in *gatae* domain at 18 h. Effect seen is probably via Otx which represses *wnt8* in this period.
[15]Indirect affect since introduction of *pmar1* mRNA rescues delta expression in Cad MOE embryos.
[16]Probably indirect via *gcm*, since *gcm* is an early target of N signaling system, and *fvmo* expression depends on *gcm* expression.
[17]Probably indirect via GataE since effect is too late for a Tcf input since in most of *foxb* expression domain nuclear β-catenin has disappeared by 24-27 h.
[18]Probably indirect because if there were an Otx site in *gatac* effect would have been seen earlier than 24 h.
[19]Indirect effect from OE, since *dri* is expressed only in OE at this time.

ACKNOWLEDGMENTS

The work summarized in this paper is the product of many minds, many kinds of technology, many different experiments, and several different laboratory groups and institutions, as indicated in the list of authors. Much of the computational work and model assembly was done with Hamid Bolouri and his group at the University of Hertfordshire, the remainder with C. Titus Brown of Caltech (see Bolouri and Davidson, 2002; Brown *et al.*, 2002). Most of the experimental molecular biology directly underlying the model was carried out by the first author's group at Caltech (individual contributions that are otherwise yet unpublished are indicated in the QPCR data Web site, and in papers elsewhere in this issue, *viz* Ransick *et al.*, 2002; Rast *et al.*, 2002; Oliveri *et al.*, 2002; Yuh *et al.*, 2002). Much of the experimental embryology on which the model rests comes from the laboratory of David R. McClay of Duke University. Most of the BAC sequences were obtained at Leroy Hood's sequencing center at the Institute for Systems Biology, Seattle, Washington. We thank Scott Bloom at ISB for his technical assistance. We are also extremely grateful to Elbert Branscombe of JGI who jump-started this whole project by offering to arrange for JGI to sequence *S. purpuratus* BACs containing endomesodermal genes, and in the event provided about a third of the BAC sequences listed in Table 2. The work described in this paper required an intense expenditure of technical effort, and would never have been possible without the contributions of Jane Wyllie, Ping Dong, Niñon Le, Miki Jun, Jina Jun, and Patrick Leahy, all superbly able, experienced technical assistants. We are grateful to Maria Rosa Ponce and José Luis Micol, who participated in the early phases of the genomics underlying the network analysis when we were first feeling our way forward. The project also depended heavily on the Arraying Facility at the Beckman Institute, directed by R. Andrew Cameron and staffed by Julie Hahn, Arnufo Lorico, and Ted Biondi; and supported by a grant from NCRR (RR-15044). The project has intensively utilized the computational and technological facilities in the Transcription Factor Research Center of the Beckman Institute, headed by Chiou-Hwa Yuh. The embryological and molecular biology aspects of the network analysis were supported by grants from NIH (HD-37105 and RR-06591); the computational aspects by a grant from NIGMS (GM-61005); the comparative aspects by a grant from NASA's Fundamental Space Biology program (NAG2-1368); and other support was provided by the Stowers Institute for Medical Research, the Beckman Institute, and the Lucille P. Markey Trust. The first author would particularly like to acknowledge the continuing support and encouragement of Dr. Richard Tasca of NICHHD. Many other scientists have contributed in very important ways to this project. In addition to those whose published contributions are referred to in text, we are very grateful to Koji Akasaka of Hiroshima University, William Klein of MD Anderson Hospital; and Athula Wikramanayake of the University of Hawaii for making available for reference their unpublished data. Prof. Ellen Rothenberg of Caltech provided a perspicacious and valuable critical review of the manuscript, for which it benefited greatly. Finally, one should perhaps acknowledge the unusual properties of sea urchin embryos, which seem preordained for developmental gene network analysis, and many millions, if not billions, of which contributed directly to each of the diagrams in this paper.

REFERENCES

Arnone, M., and Davidson, E. H. (1997). The hardwiring of development: Organization and function of genomic regulatory systems. *Development* **124,** 1851–1864.
Bolouri, H., and Davidson, E. H. (2002). Modeling DNA sequence-based *cis*-regulatory gene networks. *Dev. Biol.* **246,** 2–13.
Boveri, T. (1902). Über mehrpolige Mitosen als Mittel zur Analyse des Zellkerns. *Verh. Phys. Med. Ges. Würzburg* **35**, 67–90. [Parts of this paper have been translated by B. R Voeller (1968). "The Chromosome Theory of Inheritance," pp. 85. Appleton, New York].
Boveri, T. (1918). Zwei Fehlerquellen bei Merogoniegversuchen und die Entwicklungsfähigkeit merogonischer partiellmerogo-

nischer Seeigelbastarde. *Arch. Entwicklungsmech. Org.* **44**, 417–471.

Brown, C. T., Rust, A. G., Clarke, P. J. C., 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.

Cameron, R. A., and Davidson, E. H. (1997). LiCl perturbs ectodermal veg$_1$ lineage allocations in *Strongylocentrotus purpuratus* embryos. *Dev. Biol.* **187**, 236–239.

Cameron, R. A., Hough-Evans, B. R., Britten, R. J., and Davidson, E. H. (1987). Lineage and fate of each blastomere of the eight-cell sea urchin embryo. *Genes Dev.* **1**, 75–85.

Cameron, R. A., Fraser, S. E., Britten, R. J., and Davidson, E. H. (1991). Macromere cell fates during sea urchin development. *Development* **113**, 1085–1092.

Cameron, R. A., Mahairas, G., Rast, J. P., Martinez, P., Biondi, T. R., Swartzell, S., Wallace, J. C., Poustka, A. J., Livingston, B. T., Wray, G. A., Ettensohn, C. A., Lehrach, H., Britten, R. J., Davidson, E. H., and Hood, L. (2000). A sea urchin genome project: Sequence scan, virtual map, and additional resources. *Proc. Natl. Acad. Sci. USA* **97**, 9514–9518.

Chuang, C.-K., Wikramanayake, A. H., Mao, C.-A., Li, X., and Klein, W. H. (1996). Transient appearance of *Strongylocentrotus purpuratus* Otx in micromere nuclei: Cytoplasmic retention of SpOtx possibly mediated through an α-actinin interaction. *Dev. Genet.* **19**, 231–237.

Croce, J., Lhomond, G., Lozano, J., and Gache, C. (2001). *ske-T*, a T-box gene expressed in the skeletogenic mesenchyme lineage of the sea urchin embryo. *Mech. Dev.* **107**, 159–162.

Davidson, E. H. (2001). "Genomic Regulatory Systems. Development and Evolution." Academic Press, San Diego.

Davidson, E. H., Cameron, R. A., and Ransick, A. (1998). Specification of cell fate in the sea urchin embryo: Summary and some proposed mechanisms. *Development* **125**, 3269–3290.

Dobias, S. L., Zhao, A. Z., Tan, H., Bell, J. R., and Maxson, R. (1996). SpHbox7, a new Abd-B class homeobox gene from the sea urchin *Strongylocentrotus purpuratus*: Insights into the evolution of *Hox* gene expression and function. *Dev. Dyn.* **207**, 450–460.

Emily-Fenouil, F., Ghiglione, C., Lhomond, G., Lepage, T., and Gache, C. (1998). GSK3β/shaggy mediates patterning along the animal–vegetal axis of the sea urchin embryo. *Development* **125**, 2489–2498.

George, N. C., Killian, C. E., and Wilt, F. H. (1991). Characterization and expression of a gene encoding a 30.6-kDa *Stronglylocentrotus pupuratus* spicule matrix protein. *Dev. Biol.* **147**, 334–342.

Gross, J. M., and McClay, D. R. (2001). The role of Brachyury (T) during gastrulation movements in the sea urchin *Lytechinus variegatus*. *Dev. Biol.* **239**, 132–147.

Gurdon, J. B. (1988). A community effect in animal development. *Nature* **336**, 772–774.

Gurdon, J. B., Lemaire, P., and Kato, K. (1993). Community effects and related phenomena in development. *Cell* **75**, 831–834.

Harada, Y., Akasaka, K., Shimada, H., Peterson, K. J., Davidson, E. H., and Satoh, N. (1996). Spatial expression of a forkhead homologue in the sea urchin embryo. *Mech. Dev.* **60**, 163–173.

Harkey, M. A., Klueg, K., Sheppard, P., and Raff, R. A. (1995). Structure, expression, and extracellular targeting of PM27, a skeletal protein associated specifically with growth of the sea urchin larval spicule. *Dev. Biol.* **168**, 549–566.

Hörstadius, S. (1939). The mechanics of sea urchin development, studied by operative methods. *Biol. Rev. Cambridge Philos. Soc.* **14**, 132–179.

Howard, E. W., Newman, L. A., Oleksyn, D. W., Angerer, R. C., and Angerer, L. M. (2001). Spkrl: A direct target of β-catenin regulation required for endoderm differentiation in sea urchin embryos. *Development* **128**, 365–375.

Huang, L., Li, X., El-Hodiri, H. M., Dayal, S., Wikramanayake, A. H., and Klein, W. H. (2000). Involvement of Tcf/Lef in establishing cell types along the animal-vegetal axis of sea urchins. *Dev. Genes Evol.* **210**, 73–81.

Jacobsen, T. L., Brennan, K., Arias, A. M., and Muskavitch, M. A. T. (1998). *Cis*-interactions between Delta and Notch modulate neurogenic signaling in *Drosophila*. *Development* **125**, 4531–4540.

Katoh-Fukui, Y., Noce, T., Ueda, T., Fujiwara, Y., Hashimoto, N., Higashinakagawa, T., Killian, C. E., Livingston, B. T., Wilt, F. H., Benson, S. C., Sucov, H. M., and Davidson, E. H. (1991). The corrected structure of the SM50 spicule matrix protein of *Strongylocentrotus purpuratus*. *Dev. Biol.* **145**, 201–202.

Kawasaki, T., Mitsunaga-Nakatsubo, K., Takeda, K., Akasaka, K., and Shimada, H. (1999). Lim1 related homeobox gene (HpLim1) expressed in sea urchin embryos. *Dev. Growth Differ.* **41**, 273–282.

Kenny, A. P., Kozlowski, D., Oleksyn, D. W., Angerer, L. M., and Angerer, R. C. (1999). SpSoxB1, a maternally encoded transcription factor asymmetrically distributed among early sea urchin blastomeres. *Development* **126**, 5473–5483.

Kurokawa, D., Kitajima, T., Mitsunaga-Nakatsubo, K., Amemiya, S., Shimada, H., and Akasaka, K. (1999). HpEts, an *ets*-related transcription factor implicated in primary mesenchyme cell differentiation in the sea urchin embryo. *Mech. Dev.* **80**, 41–52.

Lee, Y.-H., Britten, R. J., and Davidson, E. H. (1999). *SM37*, a new skeletogenic gene of the sea urchin embryo isolated by regulatory target site screening. *Dev. Growth Differ.* **41**, 303–312.

Li, X., Chuang, C.-K., Mao, C.-A., Angerer, L. M., and Klein, W. H. (1997). Two Otx proteins generated from multiple transcripts of a single gene in *Strongylocentrotus purpuratus*. *Dev. Biol.* **187**, 253–266.

Logan, C. Y., and McClay, D. R. (1997). The allocation of early blastomeres to the ectoderm and endoderm is variable in the sea urchin embryo. *Development* **124**, 2213–2223.

Logan, C. Y., Miller, J. R., Ferkowicz, M. J., and McClay, D. R. (1999). Nuclear β-catenin is required to specify vegetal cell fates in the sea urchin embryo. *Development* **126**, 345–357.

Luke, N. H., Killian, C. E., and Livingston, B. T. (1997). Spfkh1 encodes a transcription factor implicated in gut formation during sea urchin development. *Dev. Growth Differ.* **39**, 285–294.

Makabe, K. W., Kirchhamer, C. V., Britten, R. J., and Davidson, E. H. (1995). *Cis*-regulatory control of the *SM50* gene, an early marker of skeletogenic lineage specification in the sea urchin embryo. *Development* **121**, 1957–1970.

Martinez, P., and Davidson, E. H. (1997). SpHmx, a sea urchin homeobox gene expressed in embryonic pigment cells. *Dev. Biol.* **181**, 213–222.

Martinez, P., Rast, J. P., Arenas-Mena, C., and Davidson, E. H. (1999). Organization of an echinoderm *Hox* gene cluster. *Proc. Natl. Acad. Sci. USA* **96**, 1469–1474.

McClay, D. R., and Logan, C. Y. (1996). Regulative capacity of the archenteron during gastrulation in the sea urchin. *Development* **122**, 607–616.

McClay, D. R., Peterson, R. E., Range, R. C., Winter-Vann, A. M., and Ferkowitz, M. J. (2000). A micromere induction signal is activated by β-catenin and acts through Notch to initiate specification of secondary mesenchyme cells in the sea urchin embryo. *Development* **127**, 5113–5122.

Nocente-McGrath, C., Brenner, C. A., and Ernst, S. G. (1989). Endo-16, a lineage-specific protein of the sea urchin embryo, is first expressed just prior to gastrulation. *Dev. Biol.* **136**, 264–272.

Oliveri, P., Carrick, D. M., and Davidson, E. H. (2002). A regulatory gene network that directs micromere specification in the sea urchin embryo. *Dev. Biol.* **246**, 209–228.

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**, 773–786.

Parr, B. A., Parks, A. L., and Raff, R. A. (1990). Promoter structure and protein sequence of msp130, a lipid-anchored sea urchin glycoprotein. *J. Biol. Chem.* **265**, 1408–1413.

Peterson, K. J., Harada, Y., Cameron, R. A., and Davidson, E. H. (1999). Expression pattern of *Brachyury* and *Not* in the sea urchin: Comparative implications for the origins of mesoderm in the basal deuterostomes. *Dev. Biol.* **207**, 419–431.

Ransick, A., and Davidson, E. H. (1993). A complete second gut induced by transplanted micromeres in the sea urchin embryo. *Science* **259**, 1134–1138.

Ransick, A., and Davidson, E. H. (1995). Micromeres are required for normal vegetal plate specification in sea urchin embryos. *Development* **121**, 3215–3222.

Ransick, A., and Davidson, E. H. (1998). Late specification of veg_1 lineages to endodermal fate in the sea urchin embryo. *Dev. Biol.* **195**, 38–48.

Ransick, A., Ernst, S., Britten, R. J., and Davidson, E. H. (1993). Whole mount *in situ* hybridization shows *Endo16* to be a marker for the vegetal plate territory in sea urchin embryos. *Mech. Dev.* **42**, 117–124.

Ransick, A., Rast, J. P., Minokawa, T., Calestani, C., and Davidson, E. H. (2002). New early zygotic regulators of endomesoderm specification in sea urchin embryos discovered by differential array hybridization. *Dev. Biol.* **246**, 132–147.

Rast, J. P., Amore, G., Calestani, C., Livi, C. B., Ransick, A., and Davidson, E. H. (2000). Recovery of developmentally defined gene sets from high-density cDNA macroarrays. *Dev. Biol.* **228**, 270–286.

Rast, J. P., Cameron, R. A., Poustka, A. J., and Davidson, E. H. (2002). Downstream target of Brachyury in the sea urchin embryo. *Dev. Biol.* **246**, 191–208.

Ruffins, S. W., and Ettensohn, C. A. (1993). A clonal analysis of secondary mesenchyme cell fates in the sea urchin embryo. *Dev. Biol.* **160**, 285–288.

Ruffins, S. W., and Ettensohn, C. A. (1996). A fate map of the vegetal plate of the sea urchin (*Lytechinus variegatus*) mesenchyme blastula. *Development* **122**, 253–263.

Sherwood, D. R., and McClay, D. R. (1997). Identification and localization of a sea urchin Notch homologue: Insights into vegetal plate regionalization and Notch receptor regulation. *Development* **124**, 3363–3374.

Sherwood, D. R., and McClay, D. R. (1999). LvNotch signaling mediates secondary mesenchyme specification in the sea urchin embryo. *Development* **126**, 1703–1713.

Sherwood, D. R., and McClay, D. R. (2001). LvNotch signaling plays a dual role in regulating the position of the ectoderm-endoderm boundary in the sea urchin embryo. *Development* **128**, 2221–2232.

Sucov, H. M., Benson, S., Robinson, J. J., Britten, R. J., Wilt, F., and Davidson, E. H. (1987). A lineage-specific gene encoding a major matrix protein of the sea urchin embryo spicule. II. Structure of the gene and derived sequence of the protein. *Dev. Biol.* **120**, 507–519.

Sweet, H. C., Hodor, P. G., and Ettensohn, C. A. (1999). The role of micromere signaling in Notch activation and mesoderm specification during sea urchin embryogenesis. *Development* **126**, 5255–5265.

Sweet, H. C., Gehring, M., and Ettensohn, C. A. (2002). LvDelta is a mesoderm-inducing signal in the sea urchin embryo and can endow blastomeres with organizer-like properties. *Development* **129**, 1945–1955.

Vonica, A., Weng, W., Gumbiner, B. M., and Venuti, J. M. (2000). TCF is the nuclear effector of the β-catenin signal that patterns the sea urchin animal–vegetal axis. *Dev. Biol.* **217**, 230–243.

Wang, W., Wikramanayake, A. H., Gonzalez-Rimbau, M., Vlahou, A., Flytzanis, C. N., and Klein, W. H. (1996). Very early and transient vegetal-plate expression of *SpKrox1*, a Krüppel/Krox gene from *Stronglyocentrotus purpuratus*. *Mech. Dev.* **60**, 185–195.

Wikramanayake, A. H., Huang, L., and Klein, W. H. (1998). β-Catenin is essential for patterning the maternally specified animal–vegetal axis in the sea urchin embryo. *Proc. Natl. Acad. Sci. USA* **95**, 9343–9348.

Yeh, E., Zhou, L., Rudzik, N., and Boulianne, G. L. (2000). Neuralized functions cell autonomously to regulate Drosophila sense organ development. *EMBO J.* **19**, 4827–4837.

Yeh, E., Dermer, M., Commisso, C., Zhou, L., McGlade, C. J., and Boulianne, G. L. (2001). Neuralized functions as an E3 ubiquitin ligase during *Drosophila* development. *Curr. Biol.* **11**, 1675–1679.

Yuh, C.-H., Bolouri, H., and Davidson, E. H. (1998). Genomic *cis*-regulatory logic: Functional analysis and computational model of a sea urchin gene control system. *Science* **279**, 1896–1902.

Yuh, C.-H., Bolouri, H., and Davidson, E. H. (2001). *cis*-Regulatory logic in the *endo16* gene: Switching from a specification to a differentiation mode of control. *Development* **128**, 617–628.

Yuh, C.-H., Brown, C. T., Livi, C., Clarke, P. J. C., and Davidson, E. H. (2002). Patchy interspecific sequence similarities efficiently identify active *cis*-regulatory elements in the sea urchin. *Dev. Biol.* **246**, 148–161.

Zhu, X., Mahairas, G., Illies, M, Cameron, R. A, Davidson, E. H., and Ettensohn, C. A. (2001). A large-scale analysis of mRNAs expressed by primary mesenchyme cells of the sea urchin embryo. *Development.* **128**, 2615–2627.

Zorn, A. M., Barish, G. D., Williams, B. O., Lavender, P., Klymkowsky, M. W., and Varmus, H. E. (1999). Regulation of Wnt signaling by Sox proteins: XSoxβ17α/β and XSox3 physically interact with β-catenin. *Mol. Cell* **4**, 487–498.

Received for publication January 3, 2002
Revised February 20, 2002
Accepted February 20, 2002

APPENDIX 2

fragility). A key research problem is distinguishing among laws, protocols, and historical accidents.
36. The central dogma can be thought of as a protocol, with DNA, RNA, protein, RNA polymerase, ribosome, etc., as modules.
37. A compelling case, but using very different terminology, is made in (*18*).
38. Steady state here means simply that all variables in Fig. 2 (r, d, y, A, C, etc.) approach constants, which can be solved for algebraically.
39. "iff" means "if and only if."
40. An important use of positive feedback is to deliberately destabilize equilibria and amplify small differences to create switches and to break symmetries and homogeneities. This can create patterns that are then maintained using negative feedback. Positive feedback is also critical to autocatalysis in growth and metabolism.
41. $da/dt = a' = gu$ means that a (the output of A) is a time integral of gu, where u is the input to A.
42. Stability is easily shown using standard methods of linear systems. Steady-state values can be found (in a stable system) by setting all time derivatives to 0, yielding $gk_1y = gk_2r$ or $y = (k_2/k_1)r$.
43. Mechanisms often exist that allow controller parameters (e.g., k_1 and k_2) to be much less uncertain than g and d. It is often even easier to make ratios such as k_2/k_1 largely invariant to variations in underlying physical quantities affecting the individual k_1 and k_2.
44. If precise gain is required, then the ratio k_2/k_1 must also be precise.
45. The national power grid has integral control at the >2000 power plants to regulate frequency and voltages of delivered power, oil refineries have >10,000 such control loops, and Internet congestion control involves a form implemented as part of TCP (transport control protocol). See (*12*) for more details, proofs, and examples.
46. T. M. Yi, Y. Huang, M. I. Simon, J. Doyle, *Proc. Natl. Acad. Sci. U.S.A.* **97**, 4649 (2000).
47. A. Levchenko, P. A. Iglesias, *Biophys. J.* **82**, 50 (2002).
48. H. El-Samad, J. P. Goff, M. Khammash, *J. Theor. Biol.* **214**, 17 (2002).
49. This argument can be made rigorous and is a standard elementary result in control theory. It is a special case of the internal model principle.
50. For sufficiently large g, the frequency domain peak and time domain transients become unacceptably large, although still stable.
51. One interpretation is that negative feedback is always balanced by an equal and opposite positive feedback. Strictly speaking, with dynamics this is not well defined, and $\log|S(\omega)|$ gives the correct generalization.
52. Relatively rare circumstances can involve an inequality ($\geq$). This is worse, but it means that Eq. 4 is an inequality constraint rather than a pure "conservation" law. See (*12*, *63*).
53. The robust yet fragile nature of highly optimized complex regulatory networks can be mistakenly attributed to various kinds of bifurcations and "order-disorder" transitions (e.g., phase transitions, critical phenomena, "edge-of-chaos," pattern formation, etc.). See (*12*, *24*).
54. The development of the Boeing 777 alone required a global software and computing infrastructure with roughly 10,000 workstations, terabytes of data, and a billion-dollar price tag.
55. Systems Biology Workbench (www.cds.caltech.edu/sbw).
56. D. T. Gillespie, *J. Chem. Phys.* **113**, 297 (2000).
57. K. Zhou, J. Doyle, K. Glover, *Robust and Optimal Control* (Prentice-Hall, Upper Saddle River, NJ, 1996).
58. E. Sontag, *Mathematical Control Theory: Deterministic Finite Dimensional Systems* (Springer, New York, 1998).
59. The idea of "closed-loop" systems biology is a purely metaphorical application of the message of this article to biology research itself. Biologists already iterate (loop) between experiment and at least informal modeling and "simulation." This process might be enhanced by "closing the loop" with theory and infrastructure "modules," which would also benefit.
60. One such example is Internet technology, which is rich in protocols and feedback and is beginning to have a rich theory. Even though it is poorly understood by nonexperts and has become a focus of many specious theories, details and enormous data sets are available, and it makes an attractive example to compare with biological networks. See (*12*).
61. N. Wiener, *Cybernetics* (Wiley, New York, 1948).
62. A. R. Mushegian, E. V. Koonin, *Proc. Natl. Acad. Sci. U.S.A.* **93**, 10268 (1996).
63. H. W. Bode, *Network Analysis and Feedback Amplifier Design* (Krieger, Melbourne, FL, 1945).

REVIEW

A Genomic Regulatory Network for Development

Eric H. Davidson,[1]* Jonathan P. Rast,[1] Paola Oliveri,[1] Andrew Ransick,[1] Cristina Calestani,[1] Chiou-Hwa Yuh,[1] Takuya Minokawa,[1] Gabriele Amore,[1] Veronica Hinman,[1] César Arenas-Mena,[1] Ochan Otim,[1] C. Titus Brown,[1] Carolina B. Livi,[1] Pei Yun Lee,[1] Roger Revilla,[1] Alistair G. Rust,[2]† Zheng jun Pan,[2]‡ Maria J. Schilstra,[2] Peter J. C. Clarke,[2] Maria I. Arnone,[3] Lee Rowen,[4] R. Andrew Cameron,[1] David R. McClay,[5] Leroy Hood,[4] Hamid Bolouri[2]

Development of the body plan is controlled by large networks of regulatory genes. A gene regulatory network that controls the specification of endoderm and mesoderm in the sea urchin embryo is summarized here. The network was derived from large-scale perturbation analyses, in combination with computational methodologies, genomic data, cis-regulatory analysis, and molecular embryology. The network contains over 40 genes at present, and each node can be directly verified at the DNA sequence level by cis-regulatory analysis. Its architecture reveals specific and general aspects of development, such as how given cells generate their ordained fates in the embryo and why the process moves inexorably forward in developmental time.

The mechanism causing cats to beget cats and fish to beget fish is hardwired in the genomic DNA, because the species specificity of the body plan is the cardinal heritable property. But despite all the examples of how individual genes affect the developmental process, there is yet no case where the lines of causality can be mapped from the genomic sequence to a major process of bilaterian development. One reason for this is that most of the developmental systems that have been intensively studied produce adult body parts, such as the third instar *Drosophila* wing disc, or the vertebrate hindbrain during rhombomere specification, or the heart anlagen of flies and mice (*1*). These systems present tough challenges because they go through successive stages of pattern formation in order to generate complex morphologies, and their development is initiated from states that are already complex. Furthermore, traditional molecular, genetic, and developmental biological approaches have focused on determining the functions of one or a few genes at a time, an approach that is not adequate for analysis of large regulatory control systems organized as networks. The heart of such networks consists of genes encoding transcription factors and the cis-regulatory elements that control the expression of those genes. Each of these cis-regulatory elements receives multiple inputs from other genes in the network; these inputs are the transcription factors for which the element contains the specific target site sequences. The functional linkages of which the network is composed are those between the outputs of regulatory genes and the sets of genomic target sites to which their products bind. Therefore, these linkages can be tested and verified by cis-regulatory analysis. This means identifying the control elements and their key target sites, and experimentally determining their functional significance. The view taken here is that "understanding" why a given developmental process

[1]Division of Biology, California Institute of Technology, Pasadena, CA 91125, USA. [2]Science and Technology Research Centre, University of Hertfordshire, AL10 9AB, UK. [3]Stazione Zoologica Anton Dohrn, 80121 Naples, Italy. [4]Institute for Systems Biology, Seattle, WA 98105, USA. [5]Department of Zoology, Duke University, Durham, NC 27708–0325, USA.

*To whom correspondence should be addressed. E-mail: davidson@caltech.edu

†Present address: European Bioinformatics Institute, Wellcome Trust Genome Campus, Hinxton, Cambridgeshire CB101 1SD, UK.

‡Present address: Altera European Technology Centre, Holmers Farm Way, High Wycombe, Buckinghamshire HP12 4XF, UK.

occurs as it does requires learning the key inputs and outputs throughout the genomic regulatory system that controls the process as it unfolds.

In mechanistic terms, development proceeds as a progression of states of spatially defined regulatory gene expression. Through this progresssion, specification occurs: This is the process by which cells in each region of the developing animal come to express a given set of genes. The spatial cues that trigger specification in development are generally signaling ligands produced by other cells, in consequence of their own prior states of specification. In addition to intercellular signals, maternal molecules of regulatory significance are distributed to particular cells with the egg cytoplasm and partitioned spatially during cleavage. Ultimately, either inter- or intracellular spatial cues affect the course of events in development by causing the activation (or repression) of particular genes encoding transcription factors. But although it is these genes that do the transcriptional regulatory work of spatial specification, the locus of programmatic control for each developmental event is the sequence of the particular cisregulatory elements that respond to the inputs presented. Genes encoding transcription factors are typically used at many times and places in the life cycle, and so the uniqueness of any given developmental regulatory network lies in its operative cis-regulatory modules. Such cisregulatory systems produce new and often more refined spatial patterns than those described by their inputs: They add regulatory or informational value. For example, cis-regulatory elements active in spatial specification often use "and" logic, in that two different transcription factors, each present in a given spatial domain, must be bound to the cis-regulatory DNA at once in order for transcription to be activated (*1*). The gene is expressed only where the input patterns overlap, and this defines a new spatial regulatory state. By determining the succession of DNA sequence–based cis-regulatory transactions that govern spatial gene expression, closure can be brought to the question of why any particular piece of development actually happens.

The most closely examined example of a cis-regulatory information processing system is that which controls developmental expression of the *endo16* gene of the sea urchin embryo. *Endo16* encodes a large polyfunctional protein that is secreted into the lumen of the embryonic and larval midgut. *Endo16* is expressed in the early embryo in the progenitors of the endomesoderm, then throughout the gut, and finally only in the midgut (*2*–*4*), a not very elaborate spatial sequence. But its control system turns out to be an elegantly organized and complex information processing device that responds to both positive and negative inputs to set the boundaries of expression. Early and late expression phases are controlled by two different subregions of the regulatory sequence, or modules, each several hundred base pairs long. Together these are serviced by nine different DNA sequence–specific transcription factors. The functional role(s) of each interaction were determined (*5*, *6*), and a computational model was derived to describe how this system responds to its time-varying regulatory inputs and to mutations and combinations of its target sites. The functions that the *endo16* regulatory system performs are conditional on the inputs, and they include linear amplification of these inputs, but also many nonlinear operations such as an intermodule switch that transfers control from the early to the late module, detection of input thresholds, and various logic operations (*5*, *6*). The model affords precise predictions of the responses of this cis-regulatory system under all conditions.

Uses of a First-Stage Regulatory Network Model

A complete cis-regulatory network model would portray both the overall intergenic architecture of the network and the information processing functions of each node, at the level achieved for the *endo16* cis-regulatory system. The complete model could then handle the kinetic flow of regulatory inputs around the whole system. Because of the nonlinear processing functions at each node, inputs into the network are unlikely to be propagated through it in a linear fashion. But the primary necessity is to discover the logic map of the intergenic regulatory interactions, and to represent this map as a first-stage regulatory network model. Its function is just to define precisely those inputs and outputs to each cis-regulatory element that derive from other genes in the network. We have derived such a model for endomesoderm specification in the sea urchin embryo. Although in absolute terms there is an uncomfortably large number of genes in the endomesoderm network (almost 50 at present), they are only a tiny fraction of the total being expressed in the embryo, which is estimated at about 8500 (*1*).

There are two ways to consider such network models, which are roughly equivalent to the functional genomics point of view and the developmental biology point of view (*7*, *8*). In what we term the "view from the genome," all relevant inputs into each cis-regulatory element that occur in all cells at all times in the developmental process are shown at once. This gives the genetically determined architecture of the network and predicts the target site sequences that should be functional in the genomic cis-regulatory DNA. The second, the "view from the nucleus," highlights only those interactions occurring in given nuclei in the particular time frame of that view. It explains why given genes are or are not being expressed at given times and in given cells.

Endomesoderm Specification in the Sea Urchin Embryo

The biology of the sea urchin embryo offers natural advantages for a regulatory network analysis of development. Not many regulatory steps separate the initial zygotic gene expressions that first distinguish a given patch of embryonic cells from the activation of terminal differentiation genes in the progeny of these cells (*1*, *9*, *10*). Furthermore, the sea urchin embryo gives rise only to a very simply constructed larva that consists of single-cell-thick structures and only 10 to 12 cell types (*10*), rather than to a morphologically complex juvenile version of the adult body plan, as in the development of insects and vertebrates.

Not only is the molecular and developmental biology of the sea urchin embryo well known (*1*, *10*–*12*), but dozens of developmentally regulated genes have been cloned, the overall embryonic expression patterns are well described, and the genome has been at least somewhat characterized (*13*–*15*). A large collection of arrayed cDNA and bacterial artificial chromosome (BAC) libraries is available (*13*). Most important for present purposes, the sea urchin embryo provides a high-throughput test bed for cis-regulatory analysis by gene transfer (*6*, *16*–*18*).

The endomesoderm of the sea urchin embryo forms from cell lineages at the south pole (the "vegetal" pole) of the early embryo (Fig. 1). The endomesodermal constituents of the embryo ultimately consist of the skeletogenic mesenchyme, which arises from the micromere lineage; several other mesodermal cell types; and the gut endoderm. Most of the gut endoderm and all but the skeletogenic mesodermal cell types derive from the progeny of a ring of eight sixth cleavage cells, called "veg_2"; the remainder of the gut endoderm derives from their eight sister cells, "veg_1", which also give rise to some ectoderm. What happens in the specification of the lineages is now reasonably well understood as a result of a long series of experimental studies to which many different labs have contributed [see the compressed summary of major steps in Table 1, and see (*10*) and (*19*) for reviews]. The specification of the micromere lineages occurs as soon as these cells are formed at fourth cleavage, because if isolated then and cultured, their progeny will express skeletogenic functions just as they do in their natural situation (*10*). Their specification depends initially on localized maternal cues.

Specification of the veg_2 lineage in endomesodermal progenitor cells begins immediately as well. There are two inputs required: one a signal passed from the micromeres to the immediate ancestors of the veg_2 ring, at fourth to sixth cleavage (*20*, *21*), and the other the nuclearization of β-catenin (that is, its accumulation in the nuclei of all prospective endomesodermal cells) (*22*). β-catenin is a cofactor of the Tcf transcription factor, and its initial nuclearization is autonomous rather

than signal dependent. However, the endomesodermal cells soon activate a gene encoding the signaling ligand Wnt8 (*23*), which, when bound by the adjacent cells, stimulates a signal transduction pathway that results in further nuclearization of β-catenin/Tcf. Endomesodermal functions downstream of the Tcf transcription input are thereby reinforced by an intra-endomesodermal signaling loop (*19*).

At seventh through ninth cleavage, the descendants of the micromeres, now located in the center of the disc of veg_2 cells (Fig. 1, 10-hour embryo), emit the ligand Delta (*24*, *25*), which activates the Notch (N) signal transduction system in the adjacent veg_2 cells and is required to specify them as mesoderm [Fig. 1, 15-hour embryo (*26*–*28*)]. If we now imagine the specification map from the bottom rather than from the side as in Fig. 1, the pattern of cell fates (and by now of gene expression) would display a concentric arrangement (*10*): In the center are the "small micromeres," the fifth-cleavage sister lineage of the skeletogenic micromeres; surrounding them are the skeletogenic precursors; the veg_2 mesoderm precursors; and finally the veg_2 endoderm precursors. The embryo is still an indifferent-looking hollow ball of cells, but the specification map is well on its way to completion. At 20 to 24 hours, the skeletogenic cells move inside the blastocoel (Fig. 1, 24-hour embryo), leaving behind a now fully specified central disc of prospective mesodermal cell types, and peripheral to them, the endoderm precursors. After this, a late Wnt8 signal from the veg_2 endoderm causes the adjacent veg_1 progeny to become specified as endoderm as well, and gastrular invagination ensues. The problem that we set ourselves was to discover the network of regulatory interactions underlying the events of endomesoderm specification during the first 24 hours, by which point some mesodermal and endodermal differentiation genes are already being expressed in a cell type–specific manner.

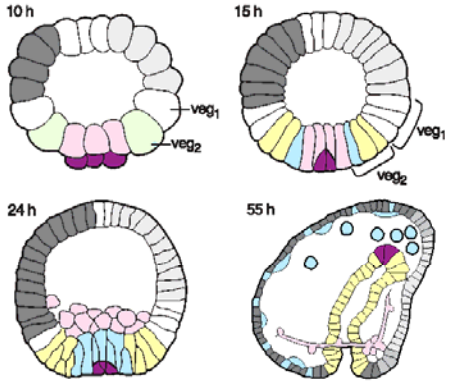


Fig. 1. Schematic diagrams of *S. purpuratus* embryos displaying specified territories (*10*). Drawings were traced off differential interference contrast images of embryos. The color coding shows the disposition of endomesoderm components and also refers to the network diagrams that follow: lavender, skeletogenic lineage; darker purple, the small micromere precursors of adult mesoderm; light green, endomesodermal veg_2 lineage that later gives rise to endoderm, yellow, and to mesoderm, light blue. Light gray indicates oral ectoderm; darker gray indicates aboral ectoderm; white indicates regions yet to be specified at the stages shown. Ten-hour (10 h) embryo: a median optical section of an early blastula, at about seventh cleavage. 15 h blastula: a similar view, at about ninth cleavage. There is now a single cell-deep ring of mesodermal precursors directly abutting the skeletogenic micromere lineage. 24 h mesenchyme blastula-stage embryo: specification of veg_2 endoderm and of mesodermal cell types completed. 55 h late gastrula stage embryo (about 800 cells): The drawing shows the later disposition of all the endomesodermal cell types about midway through embryonic morphogenesis.

Analyzing the Network

The cis-regulatory network for endomesoderm specification that we show in the following was derived in part from a large-scale perturbation analysis in which the expression of many different regulatory genes and the operation of several signaling processes were altered experimentally. The effects on many other genes were then measured with quantitative real-time fluorescence polymerase chain reaction [QPCR (*29*)] (see Fig. 2 for the kinds of perturbations applied and illustration of their effects). For an input to be considered significant, the effect of the perturbation had to be greater than threefold with respect to the control; that is, the level of the target gene transcript must be <30% or >300% of normal as a result of the perturbation. Numerical QPCR data (updated as additional measurements are made) are available online (*30*).

Most of the network linkages discovered in this study were based on perturbations that remove functions (*19*), such as morpholino-substituted antisense oligonucleotides (Fig. 2A), or blockade of all endomesoderm specification (Fig. 2C), or blockade of mesoderm specification (Fig. 2D). One mRNA encoding a transcription factor and mRNAs encoding four different Engrailed domain fusions to transcription factors were used as well (*31*, *32*). These mRNAs were all introduced into the egg in amounts that would produce levels within an

Table 1. Phenomenological aspects of endomesoderm specification in sea urchin embryos: developmental process (*55*).

1. Autonomous cues of maternal origin
   - Nuclearization of β-catenin (*22*) in micromeres (by fourth cleavage) and veg_2 cells (from sixth cleavage on)
   - Exclusion of ectodermal transcription factors from vegetal-most cell nuclei (*11*)
   - Nuclearization of Otx factor in micromeres at fourth cleavage (*56*)
2. Early micromere signal
   - Micromere signal to veg_2 (fourth through sixth cleavage) required for normal endomesodermal specification (*20*, *21*)
3. Wnt8/Tcf loop
   - Wnt8 ligand expressed throughout endomesodermal domain maintains and strengthens β-catenin/Tcf input in these nuclei (*19*, *23*)
   - β-catenin/Tcf input required for endomesoderm specification [(*22*); reviewed in (*1*, *10*, *19*)]
4. Late micromere signal
   - Expression of Delta ligand in micromeres (*24*, *25*)
   - Activation of Notch signal transduction in veg_2 descendants adjacent to micromeres that receive Delta signal (*26*–*28*, *57*)
5. Skeletogenesis
   - Skeletogenic functions expressed after ingression of skeletogenic cells in late blastula
6. Specification of veg_2 mesoderm and endoderm
   - Segregation of cell type precursors within vegetal plate complete by late blastula (*58*, *59*)
   - Mesoderm cells turn off endoderm genes, leaving endoderm genes expressed in peripheral veg_2 cells (*19*, *59*)
7. Specification of veg_1 endoderm
   - Wnt8 signal from veg_2 to veg_1 and activation of β-catenin nuclearization in abutting veg_1 cells (*19*, *22*)
8. Invagination of archenteron
   - veg_2 mesoderm carried inward at tip of archenteron on gastrulation
   - Followed by roll-in of veg_1 endoderm, contributing mainly hindgut (*60*, *61*)

order of magnitude of the natural mRNA concentrations per cell, sometimes within a few fold of these concentrations (in reality less because of continuing decay of the exogenous mRNA).

In itself, perturbation analysis cannot distinguish between direct and indirect effects: Blockade of the expression of a gene that encodes a transcriptional activator may decrease expression of both immediately and secondarily downstream target genes; and if it encodes a repressor, blockade of its expression may increase expression of both. Direct effects are those in which a perturbation in the expression or function of a transcription factor causes changes in the expression of another gene, because target sites for that factor are included in a cis-regulatory element of the gene. cis-Regulatory analysis can therefore be used to resolve whether effects on a given control element are indeed direct. Another approach that we have used at several key nodes of the network is the attempted rescue of a perturbation effect by introduction of appropriate amounts of mRNA encoding a different factor, which might be mediating an indirect effect of the perturbation (*33*). Where a rescue experiment indicates an indirect effect, or where the effect must be indirect because the affected and the perturbed genes are expressed in different cells or at different times, the implied relationships are omitted from the network models. This is because only direct effects imply specific genomic target site sequences in the cis-regulatory systems of the affected genes, and an object of the network model is to make explicit a testable map of cis-regulatory interrelations.

In an iterative process, the inferences from the experimental perturbation results were checked against the network model, further experiments were designed, the model was altered according to their results if necessary, and so forth. The model was constructed with the program Netbuilder (*34*), a new tool for the construction of computational models that allows simulations to be performed, so as to test whether its relationships generate the appropriate outputs. But from the start, the model had to conform to the facts from experimental embryology (Table 1).

A major gene discovery effort was undertaken in order to clothe with real genes the armature of interactions implied by the embryology, and to add to the collection of genes already known to be involved in endomesoderm specification. Several screens were carried out (Table 2) in which endomesoderm specification was perturbed so as to generate material for use with a very sensitive subtractive hybridization technology designed for use with large-scale arrays of ~10^5 clone cDNA libraries (macroarrays) (*35*). The purpose was to create probes in which sequences differentially expressed in the endo-

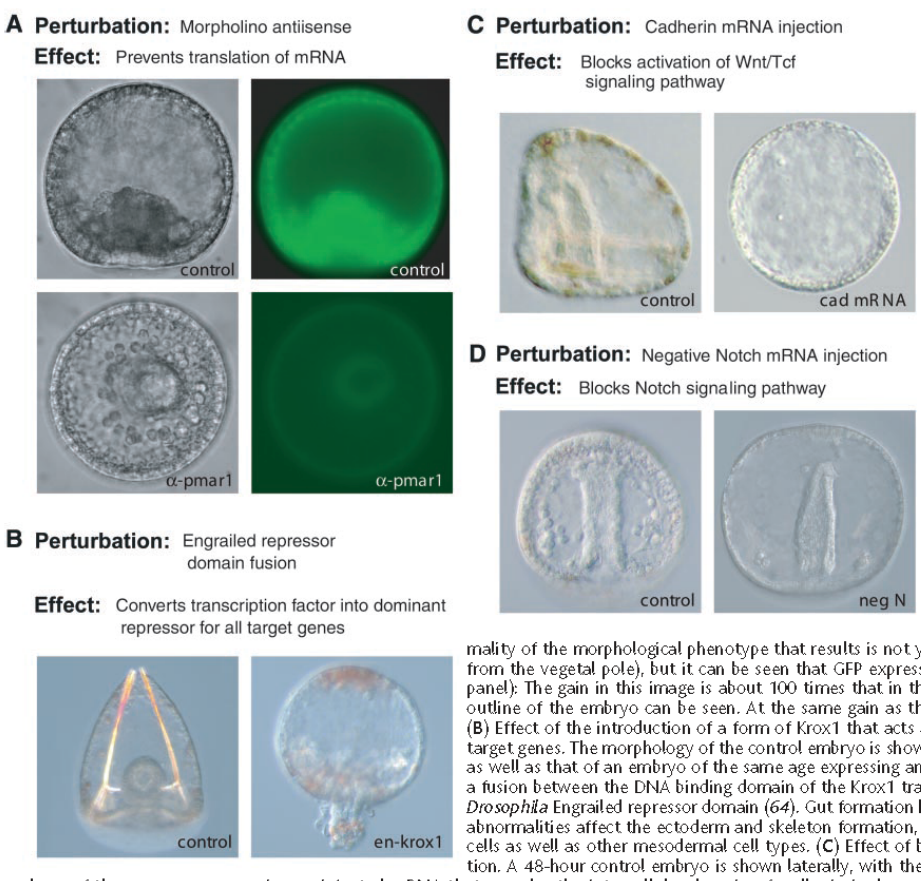


Fig. 2. Perturbations and functional knockouts used in the network analysis. (**A**) Effect of a MASO, from (*25*). Eggs giving rise to control embryos were injected with an mRNA encoding a fusion between the 5′ leader plus the initial part of the coding sequence of a gene encoding the Pmar1 transcription factor (*25*), fused to the GFP coding sequence. The control eggs also contained an irrelevant morpholino oligonucleotide. Lateral views of control embryos are shown. The top left panel displays normal embryonic morphology at 24 hours (compare Fig. 1), and the fluorescence display, top right, shows that all cells in the embryo express GFP. Eggs giving rise to the embryos in the two bottom panels were injected with the same GFP fusion plus a MASO targeted to the leader sequence of the *pmar1* mRNA. The abnormality of the morphological phenotype that results is not yet evident (left panel, viewed from the vegetal pole), but it can be seen that GFP expression is totally abolished (right panel): The gain in this image is about 100 times that in the top right panel, so that the outline of the embryo can be seen. At the same gain as the control, the image is black. (**B**) Effect of the introduction of a form of Krox1 that acts as an obligate repressor of its target genes. The morphology of the control embryo is shown at 72 hours, oral side down, as well as that of an embryo of the same age expressing an injected mRNA that encodes a fusion between the DNA binding domain of the Krox1 transcription factor (*63*) and the *Drosophila* Engrailed repressor domain (*64*). Gut formation has not occurred, other severe abnormalities affect the ectoderm and skeleton formation, and there are excess pigment cells as well as other mesodermal cell types. (**C**) Effect of blocking β-catenin nuclearization. A 48-hour control embryo is shown laterally, with the oral side on the left; and an embryo of the same age expressing an injected mRNA that encodes the intracellular domain of cadherin is shown on the right (image from A. Ransick). The cadherin embryo consists of a hollow ball of ectoderm; endomesodermal specification has been completely wiped out. (**D**) Effect of the introduction of a negatively acting derivative of the N receptor. A control 37-hour late gastrula is shown on the left, and on the right is an embryo of the same age expressing an injected mRNA encoding the extracellular domain of the N receptor (negN) (image from C. Calestani). This embryo has a normal complement of skeletogenic mesenchyme cells and a well-formed gut but only a very few mesodermal cells of veg_2 origin as compared with the control.

mesoderm are greatly enriched (by 20- to 30-fold, which affords the possibility of isolating very rare transcripts). The probes were used for high C_0t (concentration × time) hybridization to the macroarrays, and the results were digitized and analyzed with a new image analysis program, BioArray, which was designed for analysis of differential macroarray screens (*34*). New regulatory genes were recovered, as well as genes encoding differentiation proteins of the endoderm and mesoderm (*19*, *36*–*39*). Most of the transcriptional regulatory genes that are specifically involved in endomesoderm specification up to 24 hours are probably now known (*36*). On the other hand, only a small sample of endomesodermal differentiation genes have so far been recovered, because most of the screens were directed at the earlier stages of the specification process (Table 2).

Direct cis-regulatory analysis is essential to test the predicted network linkages, but the task of finding these elements on the scale of the network required an approach different from the traditional methods, which boil down to searching experimentally over all the genomic DNA surrounding a gene of interest [the average intergenic distance in *Strongylocentrotus purpuratus* is about 30 kb (*13*)]. To solve this problem, we turned to computational interspecific sequence analysis. BAC recombinants containing the genes of interest in a more or less central position were recovered from two sea urchin species. These were *S. purpuratus*, on which all the experiments were carried out, and *Lytechinus variegatus*, which develops in a very similar manner. The last common ancestor of these species lived about 50 million years ago (*40*, *41*). The sequences of BACs representing most of the genes in the network at present were obtained and annotated (*19*). A new program, FamilyRelations, was built for the purpose of recognizing short patches of conserved sequence in long stretches of genomic DNA (*34*). Applied to the *Strongylocentrotus-Lytechinus* species pair, this approach efficiently served to identify cis-regulatory elements that score positively in gene transfer tests (*42*).

In summary, three software packages were developed and used for this project: Netbuilder, FamilyRelations, and BioArray (*34*). These programs are all available online; for access, go to http://sea-urchin.caltech.edu/software.

Provisional Endomesoderm cis-Regulatory Network: The View from the Genome

The overall network (Fig. 3) combines all significant perturbation data (*19*, *30*); information on time and place of gene expression, as determined by whole mount in situ hybridization (WMISH) and QPCR measurements (*19*); computational and experimental cis-regulatory data where available; the results of rescue experiments; and all the underlying information from experimental embryology. The outputs from each gene in the diagram are color-coded: for instance, that from the *gatae* gene (GenBank accession number, AF077675), shown in dark green, provides inputs to the *lim*, *otx*β, *foxa*, *foxb*, *not*, *bra*, *elk*, *pks*, and *nrl* genes. These particular relations were derived from studies (*19*, *43*) of the effects of an α-*gatae* morpholino antisense oligonucleotide (MASO). Of course many other genes were entirely unaffected by this MASO treatment (*30*).

The early cleavage stage events in endomesoderm specification take place in the veg_2 endomesoderm lineage, indicated in light green above the triple line at the top, and in the micromere lineage shown in lavender at the left. The central light green endomesodermal domain of the diagram in Fig. 3 portrays genes that ultimately (that is, by 24 hours) function in either endoderm or mesoderm; however, many of these genes are initially expressed throughout the veg_2 domain. At the bottom, in three boxes, are shown several differentiation genes: skeletogenic genes on the left, mesodermal genes (mainly pigment cell genes) in the center, and endodermal genes on the right. So the first take-home lesson of the diagram in Fig. 3 is that, except for these differentiation genes, almost every gene in the network encodes a DNA sequence–specific transcription factor, and that most of the linkages in the network consist of cis-regulatory interactions amongst these genes. There are also three genes encoding signaling ligands: the *wnt8* gene, the *delta* gene, and the unknown gene responsible for the micromere-to-veg_2 signal (M→V2L). But on the network scale, it is plain to see that most of the regulatory work of specification is done by the cis-regulatory elements of genes encoding transcription factors. This is a general fact of life that should be true for all major developmental programs (*1*).

The model provides explanations of specific developmental processes. One example is spatial control by negative transcriptional interactions, illustrated here by the functions of the *foxa* gene. The *foxa* gene is expressed in the endoderm, as gastrulation proceeds, primarily in the foregut and midgut. Perturbation experiments with α-*foxa* MASO resulted in a sharp increase in target gene transcript levels (*30*), implying that *foxa* encodes a repressor (black barred lines emanating from this gene in Fig. 3). Two target genes are *foxb* and *bra*: *foxb* is expressed in the hindgut and blastopore (*19*, *44*) and *bra* in the blastopore (*37*, *45*). We see from the network diagram that the repression is likely to be spatial restriction due to *foxa*. Hence, an experiment was carried out in which a reporter gene controlled by a cis-regulatory element of *bra* introduced into embryos bearing an α-*foxa* MASO. The result was that expression now spread forward into the anterior gut (*46*). Comparative observations have also been made on the embryo of a starfish, a distantly related echinoderm. Here too, *foxa* is used in endomesoderm specification as a repressor, servicing the same target genes as in the *S. purpuratus* network (*47*). So the network provides an explanation of why those target genes are expressed where they are: partly as a result of spatial transcriptional repression. In addition, the network implies a temporal aspect of *foxa* expression. The *foxa* gene is seen to repress itself as well; combined with the continuing positive inputs (from GataE and other factors), the result should in principle be an oscillation. And indeed, QPCR measurements of *foxa* mRNA show that its level rises, falls, and then rises again late in gastrulation (*48*).

Table 2. Differential gene discovery screens. Macroarray filter screens were carried out with probes prepared by high-C_0T subtractive hybridization, using single-stranded driver and selectate, as described (*35*). "Selectate" denotes the cDNA preparation that contains the sequences of interest, in contrast to the nucleic acid present in excess in the hybridization reaction: The "Driver," which lacks these sequences. In the subtractive hybridizations, the reactions were carried out to near termination with respect to driver, and nonhybridized selectate sequences were recovered by hydroxyapatite chromatography (*35*).

Driver from	Selectate from	Ref.
1. Embryos expressing intracellular Cad*	LiCl-treated embryos†	*36*
2. Embryos expressing extracellular N‡	LiCl-treated embryos†	*19*, *39*
3. Control embryos too young to express *bra*§	Embryonic cells ectopically expressing *bra*‖	*37*
4. Embryos bearing α-*bra* MASO¶	Embryonic cells ectopically expressing *bra*#	*37*

*Cad, intracellular domain of cadherin (Fig. 2C). This domain sequesters β-catenin, which is thereby localized at the inner surface of the cell membrane. An excess of the cadherin intracellular domain severely decreases the availability of β-catenin for transit into the nucleus. †LiCl-treated embryos produce excess endomesoderm (*12*, *62*). ‡The extracellular domain of N acts as a repressor of N function in mesoderm specification (*27*) (Fig. 2D). §The *brachyury* (*bra*) gene is active by about 18 hours. Driver mRNA was extracted from normal 15-hour embryos. ‖Ectopic *bra*-expressing cells were obtained by disaggregating 18-hour embryos expressing genetic constructs that produce *bra* mRNA under the control of a ubiquitously active cis-regulatory element. The transgenic cells were tagged with green fluorescent protein (GFP) and isolated by fluorescence-activated cell sorting (FACS) (*37*). ¶MASO embryos were collected at 24 to 27 hours (late blastula stage). #Cells expressing *bra* were obtained by FACS as above, but at 24 to 27 hours (*37*).

The network explains some of the phenotypes observed when given processes are perturbed, in terms of its consequential regulatory logic. For example, as shown in Fig. 2C, if β-catenin nuclearization is prevented by introduction of mRNA encoding the intracellular domain of cadherin, neither endodermal nor mesodermal cell types and structures appear. In default of β-catenin/Tcf inputs, the embryo becomes a hollow ball of ectoderm. Note, however, that all the perturbation data underlying the network in Fig. 3 were obtained between 6 and 24 hours, long before any gastrulation phenotypes can be seen (*30*). Initiation of β-catenin nuclearization produces such a catastrophic re-

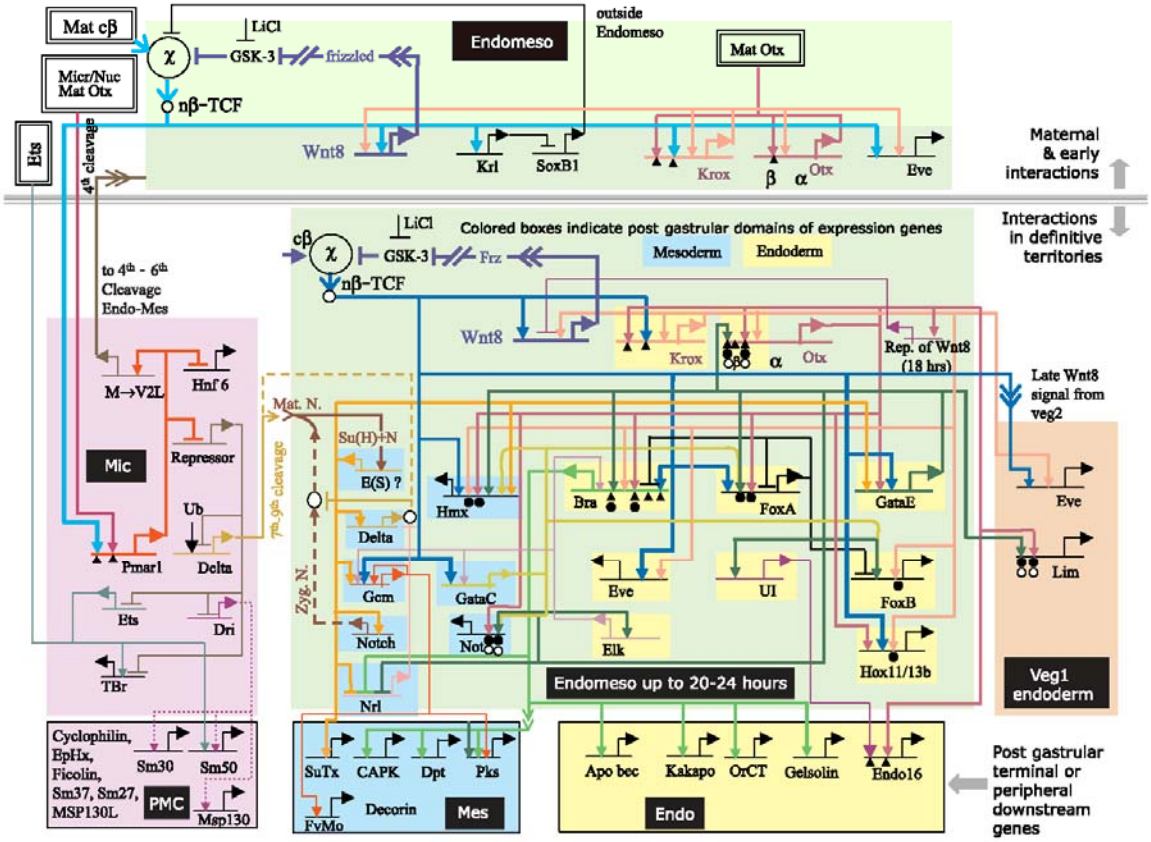


Fig. 3. Regulatory gene network for endomesoderm specification: the view from the genome. The current version of the model in this figure and the perturbation data on which it is based are available on a Web site (*30*); for additional details and discussion, see (*19*). At the top, above the triple line, are the earliest interactions; in the middle tier, the spatial domains are color-coded (Fig. 1), and genes are placed therein according to their final loci of expression. As indicated (black background labels), the lavender area at the left represents the skeletogenic micromere (mic) domain before ingression; the light green area indicates the veg_2 endomesoderm domain, with genes eventually expressed in endoderm on yellow backgrounds and genes eventually expressed in mesoderm on blue backgrounds; the tan box at right represents the veg_1 endoderm domain. Many genes are initially expressed over broader ranges, and their expression later resolves to the definitive domains. The rectangles in the lower tier of the diagram show downstream differentiation genes (PMC, "primary" or skeletogenic mesenchyme). Short horizontal lines from which bent arrows extend represent cis-regulatory elements responsible for expression of the genes named beneath the line. Embryonic gene expression was perturbed in specific ways as in Fig. 2. The arrows and barred lines indicate the inferred normal function of the input (activation or repression), as deduced from changes in transcript levels due to the perturbations. Each input arrow constitutes a prediction of specific transcription factor target site sequence(s) in the cis-regulatory control element. In some cases, the predicted target sites have been identified in experimentally defined cis-regulatory elements that generate the correct spatial pattern of expression (solid triangles). At the upper left, the light blue arrow represents the maternal β-catenin (cβ) nuclearization system (χ). This transcriptional system (nβ–TCF) is soon accelerated and then taken over by zygotic Wnt8 (dark blue lines); its initial activation, of mixed zygotic and maternal origins, is shown in light blue. Data for the roles of SoxB1 and Krüppel-like (Krl) are from (*50*, *51*). Data for the role of Ets are from (*52*, *65*). "Micr/Nuc Mat Otx" refers to the early localization of maternal Otx in micromere nuclei at fourth cleavage (*56*). Genes labeled "Repressor" are inferred; all other genes shown are being studied at the DNA sequence level and by multiplexed QPCR. "Ub" indicates a ubiquitously active positive input inferred on the basis of ubiquitous expression seen by whole-mount in situ hybridization, under conditions in which a spatial repression system that normally confines expression has been disarmed. Dotted lines in the diagram indicate inferred but indirect relationships. Arrows inserted in arrow tails indicate intercellular signaling interactions. Small open or closed circles indicate perturbation effects that resist rescue by the introduction of mRNA where there is a possibility that the effect seen is actually an indirect result of an upstream interaction; that is, this possibility of such an indirect effect has been experimentally excluded, and both sites are shown as probable direct inputs (*19*). Large open ovals represent cytoplasmic biochemical interactions at the protein level, such as those responsible for nuclearization of β-catenin, for the effect of Delta on N (*66*); or for the effect of Neuralized, an E3 ubiquitin ligase with specificity for Delta (*67*, *68*).

sult because multiple endodermal and mesoderm regulatory genes depend on a β-catenin/Tcf input. For these genes, only a few percent of control transcript levels survive cadherin mRNA injection (*19*, *30*). Another interesting phenotype is obtained when embryos are treated with α-*gcm* MASO. The result is albino larvae (*49*). The gene *gcm* is ultimately expressed in pigment cells (*36*), and a downstream target of *gcm* is the *pks* (*polyketide synthase*) gene, which is also expressed in pigment cells (*38*, *39*). This product (and other pigment cell genes under *gcm* control, not shown) is likely to be required for synthesis of the red quinone pigment these cells produce. Upstream, the network shows *gcm* to be a target of the N signaling system, because its expression is severely depressed by the introduction of a negatively acting N derivative (*19*) (Fig. 2D). In fact, *gcm* expression begins in the single ring of mesoderm progenitor cells that directly receives the Delta micromere signal (*36*). So we now have a sequence of DNA-based interactions that leads from the initial specification to the terminal differentiation of pigment cells and that explains the albino phenotype. Similarly, the network explains the α-*gatae* MASO phenotype. This treatment produces a severe interference with endoderm specification and gut development (*43*), which is no less than would be expected from the branching regulatory effects of *gatae* expression indicated in the network.

The network explains the role of the signaling interactions required in endomesodermal specification in terms of their inputs into cis-regulatory systems (except for the early micromere-to-veg$_2$ signal, the targets of which remain unknown). The gene encoding Wnt8 is itself a target of a β-catenin/Tcf input and it is, in addition, under the control of the early endomesoderm regulator *krox*. These inputs show how the autonomous nuclearization of β-catenin soon causes the Wnt8 loop to start up in all endomesoderm cells, strengthening the set of regulatory relationships indicated by the blue lines in Fig. 3.

The view from the genome provides a qualitative DNA-level explanation for the spatial domains of expression of many endomesodermal regulatory genes. No two of these genes have identical inputs: Each cis-regulatory information processing system has its own job to do. The network shows that the downstream targets of a few of these regulatory genes, such as *bra* (*37*), include differentiation proteins that were discovered in our differential screens, but for many of the regulatory genes the downstream targets are still unknown.

System-Level Insights into the Developmental Process

Physiological transcriptional responses flicker on after the advent of stimuli, then return to their ground state; for example, after changes in the level of nutrients or the advent of toxins in the bloodstream, or after the appearance of pathogens. In contrast, the fundamental feature of developmental transcriptional systems in higher (bilaterian) animals is that it always moves inexorably forward, never reversing direction. This property is clearly evident in the developmental process considered here, and the network provides a concrete mechanistic explanation. To see this, we consider views from the nuclei at successive stages (Figs. 4 and 5).

The initial events in endomesoderm specification occur in the micromeres and in the veg$_2$

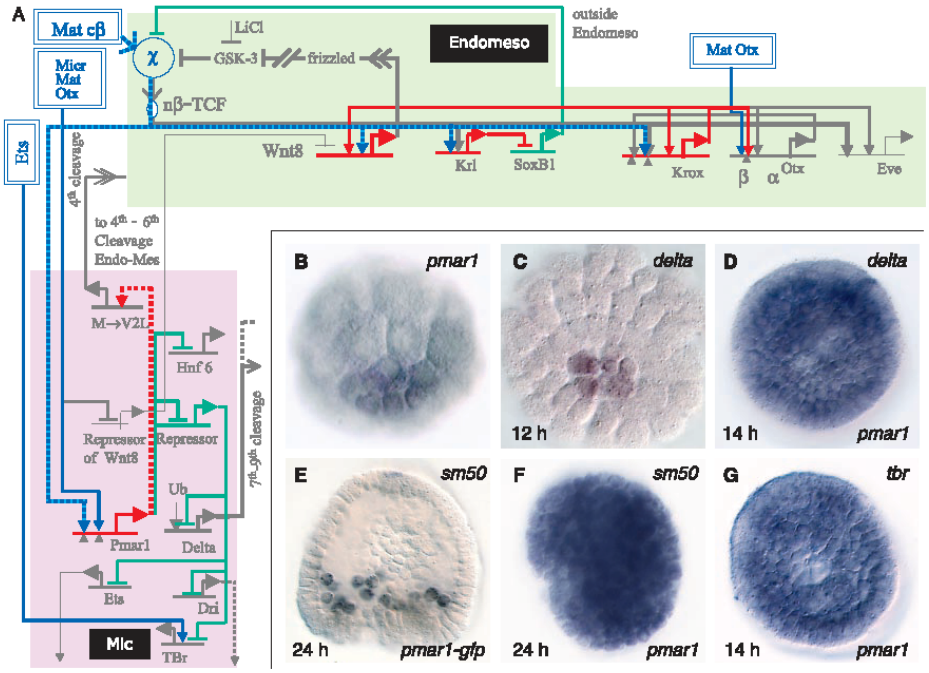


Fig. 4. Initial events in endomesoderm specification. (**A**) View from veg$_2$ endomesoderm and micromere nuclei, about fourth to seventh cleavage. Maternal inputs are shown in blue boxes (see Fig. 3 for abbreviations) and blue lines, except for the autonomous nuclearization of β-catenin, shown in a hatched blue line. Four early zygotic transcriptional activations are indicated in red: *krox*, *krl*, *wnt8* in the endomesodermal domain (all of which require the β-catenin/Tcf input), and *pmar1* in the micromere (mic) domain, which requires this and a maternal Otx input [suggested by cis-regulatory as well as perturbation evidence (*19*)]. Directly or indirectly, *pmar1* is also required for expression of the ligand conveying the early micromere to veg$_2$ signal (M→V2L). The negatively acting subnetworks discussed in text are shown in green. All other gene expressions and interactions in the network are indicated in gray. (**B** through **G**) Whole-mount in situ hybridization displays, from (*25*). The gene, expression of which is being displayed, is shown at the upper right, and the mRNA injected into the egg at the lower right; the age of the embryo is at lower left. (B) Expression of *pmar1* specifically in micromeres. (C) Expression of *delta* specifically in micromeres. (D) Expression of *delta* in all embryonic cells when *pmar1* mRNA is translated everywhere, after injection into the egg. Exactly the same result is obtained if an Engrailed domain fusion is instead expressed (*25*); because the Engrailed fusion acts as an obligaterepressor of *pmar1* targetgenes, *pmar1* must normally act as a repressor. (E) Expression of *sm50*, a skeletogenic differentiation gene exclusively in skeletogenic mesenchyme cells (*69*). (F) Global expression of *sm50* in embryos expressing *pmar1* globally. (G) Expression of the skeletogenic regulator *tbr* in embryos expressing *pmar1* mRNA globally. (F) and (G) show that the whole embryo has been converted to a state of skeletogenic mesenchyme differentiation. Note the rounded form of the cells at 24 hours in (F), as compared to the control in (E), due to their tendency to behave mesenchymally.

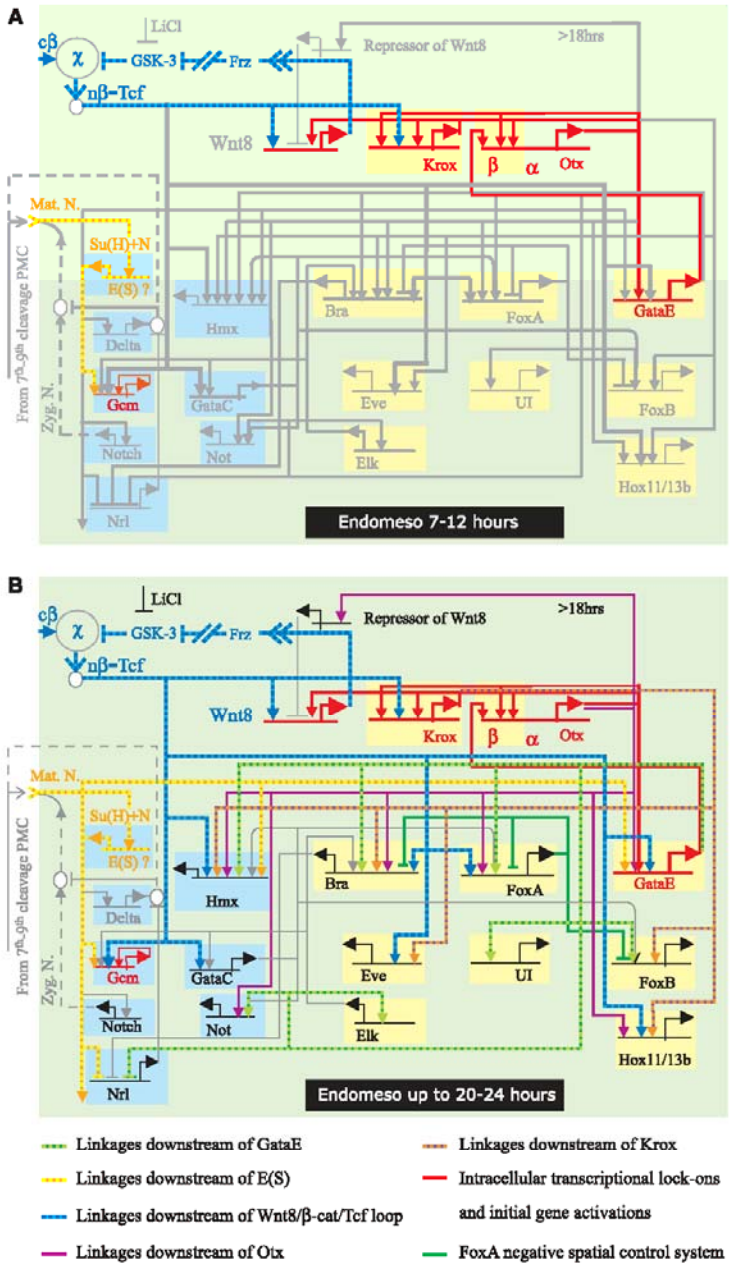


Fig. 5. Lock-down functions and expression of the complete regulatory state. (**A**) Institution of regulatory lock-down devices, shown in color. This view from the endomesoderm nuclei extends from about sixth cleavage to midblastula stage (Fig. 1). The features illuminated are the zygotic Wnt8/Tcf loop (hatched blue), and zygotic auto- and cross-regulations (red), as discussed in text. The N signal transduction input into the *gcm* gene is shown in hatched orange. (**B**) Complete activation of the endomesodermal regulatory system: the view from the nuclei from midblastula to after mesenchyme blastula (Fig. 1). By this point, both endoderm and mesoderm specifications have become final, and all genes shown are being expressed. All can be accounted for in terms of the set of inputs included in the color key at the bottom. Except for the Delta and Wnt8 signal-mediated inputs, which are transient, these regulatory inputs have by now achieved stabilization by the interactions shown in (A).

lineage, as summarized above. The maternal inputs provide the initial state, with respect to regulatory transactions. There are two consequences of the initial zygotic transcriptional responses (Fig. 4A, shown in red). The first is to begin the activation of the endomesodermal zygotic control apparatus; here, by turning on the *krox* (*35*) and *krl* [krüppel-like (*50*)] genes in the veg_2 endomesoderm and the *pmar1* gene in the micromeres. The second is a surprise: An immediate sequel, in both domains, is to engage repressive subnetworks (shown in green) of interactions that have the effect of stabilizing the initial definition of the endomesodermal and mesomere territories by cutting off the possibility of similar transcriptional activations elsewhere. The *krl* gene encodes a repressor that prevents expression of *soxb1* in the endomesoderm, though it is expressed everywhere else (*50*, *51*). The SoxB1 protein antagonizes nuclearization of β-catenin. The *krl*/*soxb1* loop is an early lock-down device to keep the endomesodermal cells endomesodermal (because they have elevated nuclear β-catenin from the start) and to prevent other cells from going the same way. The *pmar1* gene active in the micromeres also encodes a repressor. Its target is an unknown gene that produces another repressor of key regulators of micromere-specific function. Like *soxb1*, it too is potentially active everywhere, except where it itself is repressed, which is the role accomplished by *pmar1* in the micromeres. Micromere regulators that are micromere-specific only because of the *pmar1* repression system include the gene that produces the Delta signal to the surrounding veg_2 cells and the regulatory genes that are responsible for installing the skeletogenic state of differentiation in the micromere progeny [the *t-brain* (*tbr*) gene, the *ets* gene, and the *deadringer* (*dri*) gene (*19*, *25*, *52*)]. Some evidence for the *pmar1* repression system is reproduced in Fig. 4, B through G. Expression of the *delta* gene, the *tbr* skeletogenic control gene, and *sm50*, a skeletogenic differentiation gene, all occur globally if *pmar1* mRNA is expressed globally (*25*) (Fig. 4). Almost the first thing accomplished by zygotic genes activated in both the veg_2 endomesoderm and the micromeres is to activate local negative control of otherwise global repressors of the respective states of specification. The network reveals active repression of these endomesodermal regulatory states in all the cells of the embryo, except those where *krl* and *pmar1* are respectively activated.

The system next proceeds to stabilize positively, and to expand, the endomesodermal regulatory state (Fig. 5A, red interactions). The result is essentially to lock the process into forward drive: "commitment," here seen to be hardwired into the regulatory circuitry. The Wnt8/Tcf loop discussed above is a piece of this process, which consists mainly of positive cis-regulatory feedbacks; that is, auto- and cross-regula-

tions. In the future mesodermal domain, the *gcm* gene autoregulates after its initial activation though the N pathway (*49*). Similarly, the *krox* gene positively autoregulates, in addition to stimulating expression of the *wnt8* gene, which locks *wnt8* and *krox* in a positive regulatory embrace. The *krox* gene product also activates one of the transcription units of the *otx* gene (*19*, *30*, *42*). In turn, Otx stimulates the *krox* gene. The *otx* gene now provides an input into the *gatae* gene, the importance of which was discussed above; but note that the β-*otx* cis-regulatory system in turn responds positively to GataE input (*30*, *43*). This is a further positive feedback that links the *gatae* gene, a dedicated endomesodermal activator, into the stabilization circuitry. As illustrated by the color coding in Fig. 5B, the regulatory state illustrated in Fig. 5A suffices to provide inputs to every one of the known transcriptional regulatory genes in the endomesodermal domain. The drivers are Krox, Otx, GataE, Tcf, and whatever Enhancer of Split-like factor operates in this embryo downstream of N signal transduction. After this, the expression of the *wnt8* gene falls off [probably the gene is repressed by one of the Otx isoforms (*19*, *30*, *42*, *53*)]; and during the late blastula stage, β-catenin disappears from the veg_2 endomesoderm nuclei (*22*). By now, the regulatory system is locked in and has no further need of this input, which was so important in the initial phases of the specification process.

Here we can see how an active cis-regulatory network produces the developmental phenomenon of progressivity. Later, epigenetic processes such as changes in chromatin structure, methylation, etc., may contribute to further stabilization of the differentiated state. But the processes highlighted in Figs. 4 and 5 are sufficient to explain the progression from the initial maternal inputs, to early zygotic responses and stabilization of the state of specification, and thence to the full-fledged program of regulatory gene expression.

Conclusions

Developmental regulatory network analysis can be done in any organism where the necessary genomics, a high-throughput method of gene transfer, and the ancillary molecular methods are available. But it requires a new mix of technologies and a new level of close interactions between system-minded biologists and computational scientists. It seems no more possible to understand development from an informational point of view without unraveling the underlying regulatory networks than to understand where protein sequence comes from without knowing about the triplet code. To understand the operation of whole systems of regulatory interactions, computational models are essential: for organizing experimental extensions and tests at each stage of construction of the model, to check on consistency, and to integrate experimental results with the current network architecture by means of simulation. The cis-regulatory systems at the nodes of the network in reality each process kinetic input information: the rise and fall of the activities of the transcription factors to which they respond. But even from the first-stage model, which just states the interactions that occur at each node, there emerge system properties that can only be perceived at the network level. Examples are the features of the system treated in Figs. 4 and 5. These features explain the means by which maternal spatial cues are used to activate the zygotic transcriptional network, the progressivity of the developmental process, and its lock-down mechanisms. The network model relates these and other developmental features of the process of endomesoderm specification (*19*) directly to the genome, because it is couched in terms of cis-regulatory interactions at the DNA level. The model thus represents an outline of the heritable developmental program, but the program is not the machine. The DNA regulatory network coexists with many other multicomponent systems that constitute the machine. These systems execute biochemical functions, produce signal transduction pathways, and cause cell biological changes to occur. They sum to the majority of the working parts of the cell. Their mobilization is controlled by the transcriptional switches that hook them into the genomic regulatory control system.

The development of complex body plans is a definitive property of the Bilateria, and encoding the developmental process is a major regulatory function of the genome. It has been clear for a long time that the evolution of body plans has occurred by change in the genomic programs for the development of these body plans (*54*), and it is now clear that we need to consider this in terms of change in regulatory networks. The bilaterians all have more or less the same genetic toolkit, and in particular rely on essentially the same repertoire of regulatory genes to control the developmental organization of their body plans (*1*). Network analysis affords the means to focus on the exact consequences of differences in the use of these genes. To solve the questions of body plan evolution will require learning how architectural changes in developmental networks could be added on at each evolutionary stage, while yet preserving the workability of what was there before. It will be necessary to consider regulatory gene networks as evolutionary palimpsests—patterns of regulatory interactions that are successively overlain with new regulatory patterns. In the last analysis, understanding what a given animal is, including us, will mean understanding where each linkage of our developmental networks arose, what other forms share them, which are new, and which are ancient.

References and Notes

1. E. H. Davidson, *Genomic Regulatory Systems. Development and Evolution* (Academic Press, San Diego, CA, 2001).
2. C. Nocente McGrath, C. A. Brenner, S. G. Ernst, *Dev. Biol.* **136**, 264 (1989).
3. A. Ransick, S. Ernst, R. J. Britten, E. H. Davidson, *Mech. Dev.* **42**, 117 (1993).
4. M. Soltysik-Espanola, D. C. Klinzing, K. Pfarr, R. D. Burke, S. G. Ernst, *Dev. Biol.* **165**, 73 (1994).
5. C.-H. Yuh, H. Bolouri, E. H. Davidson, *Science* **279**, 1896 (1998).
6. ———, *Development* **128**, 617 (2001).
7. M. Arnone, E. H. Davidson, *Development* **124**, 1851 (1997).
8. H. Bolouri, E. H. Davidson, *Dev. Biol.*, in press.
9. E. H. Davidson, *Development* **113**, 1 (1991).
10. E. H. Davidson, R. A. Cameron, A. Ransick, *Development* **125**, 3269 (1998).
11. L. M. Angerer, R. C. Angerer, *Dev. Biol.* **218**, 1 (2000).
12. S. Hörstadius, *Biol. Rev. Camb. Philos. Soc.* **14**, 132 (1939).
13. R. A. Cameron *et al.*, *Proc. Natl. Acad. Sci. U.S.A.* **97**, 9514 (2000).
14. X. Zhu, G. Mahairas, R. A. Cameron, E. H. Davidson, C. A. Ettensohn, *Development* **128**, 2615 (2001).
15. A. Poustka *et al.*, *Genomics* **59**, 122 (1999).
16. C. V. Kirchhamer, E. H. Davidson, *Development* **122**, 333 (1996).
17. A. P. McMahon *et al.*, *Dev. Biol.* **108**, 420 (1985).
18. C. N. Flytzanis, R. J. Britten, E. H. Davidson, *Proc. Natl. Acad. Sci. U.S.A.* **84**, 151 (1987).
19. E. H. Davidson *et al.*, *Dev. Biol.*, in press.
20. A. Ransick, E. H. Davidson, *Science* **259**, 1134 (1993).
21. ———, *Development* **121**, 3215 (1995).
22. C. Y. Logan, J. R. Miller, M. J. Ferkowicz, D. R. McClay, *Development* **126**, 345 (1999).
23. A. Wikramanayake *et al.*, in preparation.
24. H. C. Sweet, M. Gehring, C. A. Ettensohn, *Development*, in press.
25. P. Oliveri, D. M. Carrick, E. H. Davidson, *Dev. Biol.*, in press.
26. D. R. Sherwood, D. R. McClay, *Development* **124**, 3363 (1997).
27. ———, *Development* **126**, 1703 (1999).
28. ———, *Development* **128**, 2221 (2001).
29. For perturbation measurements by QPCR, RNA was extracted from embryos grown from eggs injected with the respective perturbation reagents and converted to cDNA. This was used in multiplexed QPCR experiments to assess the quantitative effects of each perturbation on the other genes indicated in Fig. 3, usually at several different stages. The effects for data included in the model ranged from threefold increases or decreases in the amount of transcript as a result of the perturbation up to several hundred fold increases or decreases. Results were obtained on at least two different independent batches of cDNA for each perturbation, sometimes more, and often several repeats were performed on each batch. For numerical data, see the URL in (*26*).
30. For the Gene Network Update Web site, see www.its.caltech.edu/~mirsky/endomes.htm; for the QPCR data underlying the network, see www.its.caltech.edu/~mirsky/qpcr.htm.
31. X. Li, A. H. Wikramanayake, W. H. Klein, *Dev. Biol.* **212**, 425 (1999).
32. The introduction of mRNA encoding sequence-specific transcription factors at levels grossly beyond those naturally present per cell of course breeds artifacts, by engendering interactions with weak target sites. The only mRNA encoding a natural transcription factor used in this work was *pmar1* mRNA, which produces a repressor. The level injected into the egg was only two to three times the normal level per cell (discounting mRNA decay), although after injection it was globally distributed. The identical phenotypes (shown in Fig. 4) were obtained after injection of mRNA encoding a Pmar1-Engrailed domain fusion. This result justifies both experiments (*25*). Three other mRNAs encoding Engrailed fusions were used. The specificity of the first, Otx-Engrailed, was thoroughly studied earlier (*31*). The

specificity of the effects of the second, encoding Krox-Engrailed, is attested to by the minor fraction of tested genes on which it had any effect whatsoever (*19*) and the fact that the only genes that were affected are required just where *krox* is expressed, even though the exogenous mRNA is present globally. The last, encoding an Elk-Engrailed construct, affects only three other genes out of all those tested (*19*) (Fig. 3). Elk plays a peripheral role in the network up to 24 hours, and its main importance may be for later events in development.

33. In such rescue experiments, if the effect is indirect via a second gene, then the introduction of mRNA generated from the second gene will suffice to correct the perturbation effect; but if it is direct, no rescue can be obtained by this route. For example, if gene A activates gene B, which in turn activates gene C, then the effect of a knockout of gene A expression is direct for gene B but indirect for gene C, and the effect of the gene A knockout on gene C would be rescued by the introduction of B mRNA. If, on the other hand, there are necessary target sites for the gene A product in the cis-regulatory elements of both genes B and C, then the effect on gene C of a gene A knockout cannot be rescued by the introduction of B mRNA.
34. C. T. Brown *et al.*, *Dev. Biol.*, in press.
35. J. P. Rast *et al.*, *Dev. Biol.* **228**, 270 (2000).
36. A. Ransick, J. P. Rast, T. Minokawa, C. Calestani, E. H. Davidson, *Dev. Biol.*, in press.
37. J. P. Rast, R. A. Cameron, A. J. Poustka, E. H. Davidson, in preparation.
38. See *Science* Online for supplementary Web material (www.sciencemag.org/cgi/content/full/295/5560/1669/DC1).
39. C. Calestani, E. Davidson, unpublished data. All QPCR measurements cited are posted on the QPCR Web site (*30*). The phenotype of the negatively acting N embryos is shown in Fig. 2D. The expression of the *pks* gene is posted on the *Science* Web site (*38*).
40. P. Gonzalez, H. A. Lessios, *Mol. Biol. Evol.* **16**, 938 (1999).
41. C. R. C. Paul, A. B. Smith, *Biol. Rev.* **59**, 443 (1984).
42. C.-H. Yuh *et al.*, *Dev. Biol.*, in press.
43. P.Y. Lee, E. Davidson, unpublished data. All QPCR data cited are posted on the QPCR Web site (*30*). An image of the α-GataE MASO phenotype is posted on the *Science* Web site (*38*).
44. N. H. Luke, C. E. Killian, B. T. Livingston, *Dev. Growth Differ.* **39**, 285 (1997).
45. D. R. McClay, D. R. Gross, *Dev. Biol.* **239**, 132 (2001).
46. R. Cameron, E. Davidson, unpublished data. The expanded expression of *bra* in sea urchin embryos bearing α-*foxa* MASO is illustrated on the Science Web site (*38*).
47. V. Hinman, E. Davidson, unpublished data. The expanded expression of *bra* in a starfish embryo bearing α-*foxa* MASO is illustrated on the *Science* Web site (*38*).
48. P. Oliveri, E. Davidson, unpublished data. All *foxa* QPCR data cited are posted on the QPCR Web site (*30*). Other data are cited in (*19*).
49. A. Ransick, E. Davidson, unpublished data. An image of the albino phenotype produced by α-*gcm* MASO is posted on the *Science* Web site (*38*). Evidence for *gcm* autoregulation comes from QPCR data posted on the web site indicated in (*30*).
50. E. W. Howard, L. A. Newman, D. W. Oleksyn, R. C. Angerer, L. M. Angerer, *Development* **128**, 365 (2001).
51. A. P. Kenny, D. Kozlowski, D. W. Oleksyn, L. M. Angerer, R. C. Angerer, *Development* **126**, 5473 (1999).
52. K. Akasaki, personal communication.
53. X. Li, C. K. Chuang, C. A. Mao, L. M. Angerer, W. H. Klein, *Dev. Biol.* **187**, 253 (1997).
54. R. J. Britten, E. H. Davidson, *Quart. Rev. Biol.* **46**, 111 (1971).
55. This description relates specifically to indirectly developing euechinoid species, such as *S. purpuratus*, *L. variegatus*, *Hemicentrotus pulcherrimus*, and *Paracentrotus lividus*.
56. C. Chuang, A. H. Wikramanayake, C. Mao, X. Li, W. Klein, *Dev. Genet.* **19**, 231 (1996).
57. H. C. Sweet, P. G. Hodor, C. A. Ettensohn, *Development* **126**, 5255 (1999).
58. S. W. Ruffins, C. A. Ettensohn, *Dev. Biol.* **160**, 285 (1993).
59. ———, *Development* **122**, 253 (1996).
60. A. Ransick, E. H. Davidson, *Dev. Biol.* **195**, 38 (1998).
61. C. Y. Logan, D. R. McClay, *Development* **124**, 2213 (1997).
62. R. A. Cameron, E. H. Davidson, *Dev. Biol.* **187**, 236 (1997).
63. W. Wang *et al.*, *Mech. Dev.* **60**, 185 (1996).
64. C. Livi, E. Davidson, unpublished data. QPCR data cited are posted on the QPCR Web site (*30*); the Krox-Engrailed phenotype is illustrated in Fig. 2B.
65. D. Kurokawa *et al.*, *Mech. Dev.* **80**, 41 (1999).
66. T. L. Jacobsen, K. Brennan, A. Martinez-Arias, M. A. T. Muskavitch, *Development* **125**, 4531 (1998).
67. E. Yeh, L. Zhou, N. Rudzik, G. L. Boulianne, *EMBO J.* **19**, 4827 (2000).
68. E. Yeh *et al.*, *Curr. Biol.* **11**, 1675 (2001).
69. H. M. Sucov, B. R. Hough-Evans, R. J. Britten, E. H. Davidson, *Genes Dev.* **2**, 1238 (1988).
70. We are particularly grateful to E. Branscombe, who arranged for the U.S. Department of Energy's Joint Genome Institute to provide many of the BAC sequences included in this work (*19*); the remainder were obtained at the Institute for Systems Biology, Seattle, WA. We thank E. Rothenberg of Caltech for perspicacious and very helpful comments on the manuscript. The major embryological aspects of this work were supported by grants from NIH (HD-37105 and RR-06591); the computational aspects were supported by a grant from NIH (GM-61005); the comparative aspects were supported by a grant from the NASA/Ames Fundamental Space Biology program (NAG2-1368); the arrayed cDNA and BAC libraries on which the project depended were produced with the support of a grant from NIH (RR-15044); and the work on the *brachyury* gene was supported by a Human Frontiers grant (RG0290) to R.A.C. Other support was provided by the Lucille P. Markey Charitable Trust, the Stowers Institute for Medical Research, the Beckman Institute of Caltech, and a grant from NSF (IBN99882875) to R.A.C.

REVIEW

Modeling the Heart—from Genes to Cells to the Whole Organ

Denis Noble

Successful physiological analysis requires an understanding of the functional interactions between the key components of cells, organs, and systems, as well as how these interactions change in disease states. This information resides neither in the genome nor even in the individual proteins that genes code for. It lies at the level of protein interactions within the context of subcellular, cellular, tissue, organ, and system structures. There is therefore no alternative to copying nature and computing these interactions to determine the logic of healthy and diseased states. The rapid growth in biological databases; models of cells, tissues, and organs; and the development of powerful computing hardware and algorithms have made it possible to explore functionality in a quantitative manner all the way from the level of genes to the physiological function of whole organs and regulatory systems. This review illustrates this development in the case of the heart. Systems physiology of the 21st century is set to become highly quantitative and, therefore, one of the most computer-intensive disciplines.

The amount of biological data generated over the past decade by new technologies has completely overwhelmed our ability to understand it. Genomics has provided us with a massive "parts catalog" for the human body; proteomics seeks to define these individual "parts" and the structures they form in detail. But there is as yet no "user's guide" describing how these parts are put together to allow those interactions that sustain life or cause disease. In many cases, the cellular, organ, and system functions of genes and proteins are unknown, although clues often come from similarity in the gene sequences. Moreover, even when we understand function at the protein level, successful intervention, for example, in drug therapy, depends on knowing how a protein behaves in context, as it interacts with the rest of the relevant cellular machinery to generate function at a higher level. Without this integrative knowledge, we may not even know in which disease states a receptor, enzyme, or transporter is relevant, and we will certainly encounter side effects that are unpredictable from molecular information alone.

Inspecting genome databases alone will not get us very far in addressing these problems. The reason is simple. Genes code for protein sequences. They do not explicitly code for the interactions between proteins

University Laboratory of Physiology, Parks Road, Oxford OX1 3PT, UK. E-mail: denis.noble@physiol.ox.ac.uk

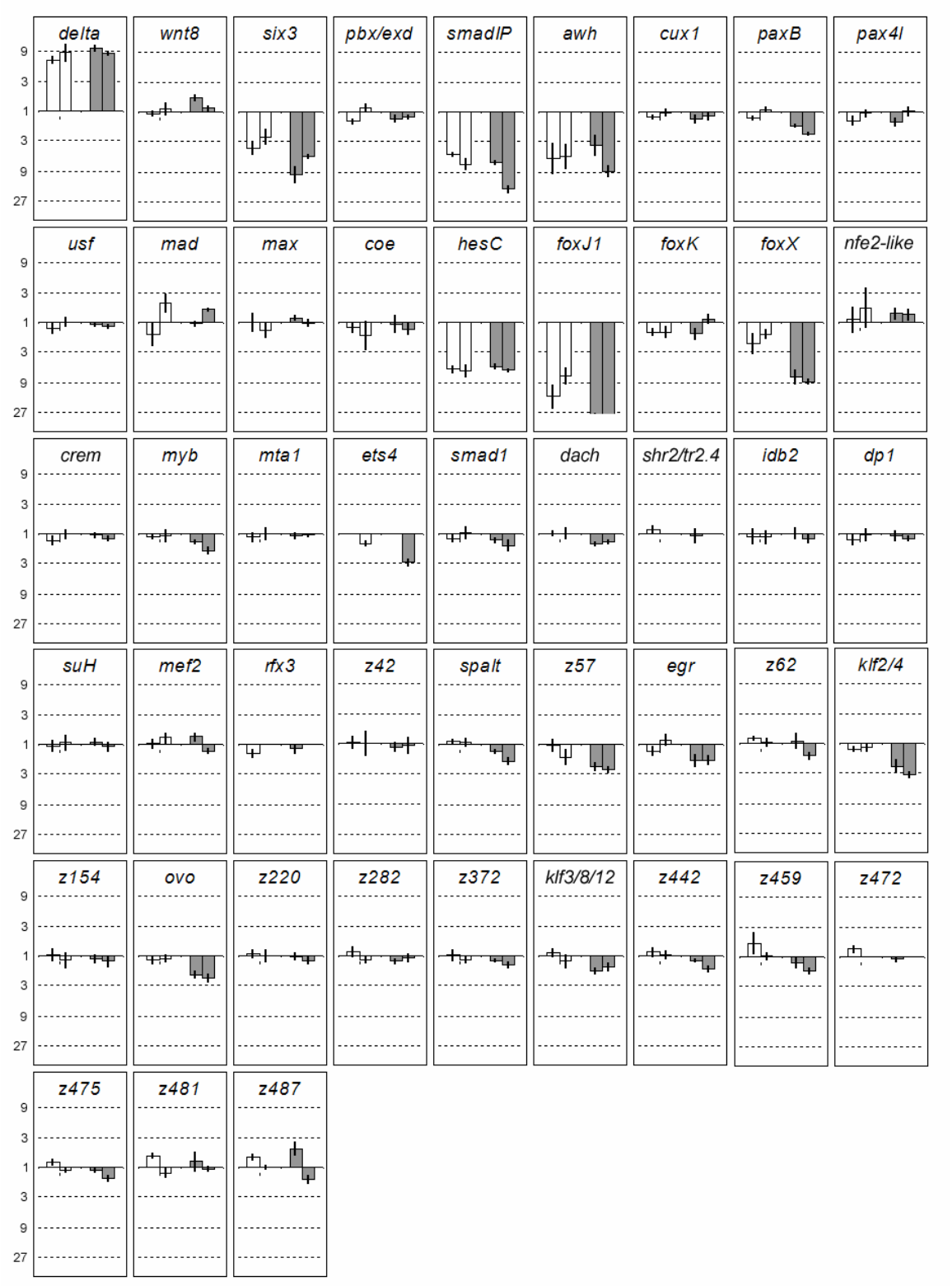
APPENDIX 3

Supplementary Material

Supplementary Table 1: List of R of mic candidate genes

Name	Family	Reported transcripts per embryo			Reported expression domain/s 12h af	References
		In the egg	6h af	12h af		
Sp-six3 (2)	Homeobox	0	0	1330	N.D.	(1)
Sp-pbx/exd (42)	Homeobox	940	1100	810	Ubq	(1)
Sp-smadIP (81)	Homeobox	40 (*)	60 (*)	1780 (*)	EM (*) / ABO and OE (**)	(*) see (1); (**) see (2)
Sp-awh (122)	Homeobox	0	0	540	Ubq	(1)
Sp-cux1 (262)	Homeobox	>500	>500	>500	N.D.	(1)
Sp-paxB (274)	Homeobox	2350	1910	2240	N.D.	(1)
Sp-pax4l (394)	Homeobox	550	530	670	N.D.	(1)
Sp-usf (182)	bHLH	980	850	900	Ubq	(3)
Sp-mad (364)	bHLH	40	70	600	N.D.	(3)
Sp-max (365)	bHLH	240	200	340	N.D.	(3)
Sp-coe (607)	bHLH	150	320	200	N.D.	(3)
Sp-hesC (617)	bHLH	300	110	2420	Ubq	(3)
SpFoxJ1	Forkhead	~600 (e)	~1000 (e)	~1500 (e)	Ubq (9h)	(4)
SpFoxK	Forkhead	~1000 (e)	~1000 (e)	~2000 (e)	Ubq (9h)	(4)
SpFoxX	Forkhead	~2000 (e)	~2500 (e)	~1500 (e)	Ubq (9h)	(4)
Sp-nfe2-like (7)	bzip	50	70	240	N.D.	(3)
Sp-crem (399)	bzip	1550	970	1030	N.D.	(3)
Sp-myb (284)	Myb	490	470	1000	N.D.	(3)
Sp-mta1 (285)	Myb	1630	1410	1980	N.D.	(3)
Sp-ets4	Ets	~4000 (*)	~8000 (*)(e)	~8000 (*)(e)	NVD (9h) (**)	(*) see (5); (**) see (6)
Sp-smad1 (23)	Smad	1620	1900	1890	N.D.	(3)
Sp-dach (27)	SKI/SNO/DAC	4270	780	680	Ubq	(3)
Sp-shr2/tr2.4 (155)	NHR	1200	1250	1330	Ubq	(3)
Sp-idb2 (295)	LIM domain	850	1060	540	N.D.	(3)
Sp-dp1 (318)	E2F	740	880	1060	Ubq	(3)
Sp-suH (326)	IPT/TIG domain	230	610	640	N.D.	(3)
Sp-mef2 (352)	MADS-box	840	980	300	N.D.	(3)
Sp-rfx3 (70)	Other TFs	1830	970	930	Ubq	(3)
z42	C2H2 ZnF	270	330	360	N.D.	(2)
Sp-spalt / z54	C2H2 ZnF	1960	2180	1610	Ubq	(2)
z57	C2H2 ZnF	80	80	470	N.D.	(2)
Sp-egr / z60	C2H2 ZnF	280	250	440	Ubq	(2)
z62	C2H2 ZnF	580	630	640	Ubq	(2)
Sp-klf2/4 / z85	C2H2 ZnF	320	240	1640	ABO or OE	(2)
z154	C2H2 ZnF	380	330	330	N.D.	(2)
Sp-ovo / z157	C2H2 ZnF	800	630	600	Ubq	(2)
z220	C2H2 ZnF	>120	>120	>250	N.D.	(2)
z282	C2H2 ZnF	610	480	470	N.D.	(2)
z372	C2H2 ZnF	300	270	280	N.D.	(2)
Sp-klf3/8/12 / z400	C2H2 ZnF	410	380	620	N.D.	(2)
z442	C2H2 ZnF	>500	>400	>600	Ubq	(2)
z459	C2H2 ZnF	>100	>100	>250	N.D.	(2)
z472	C2H2 ZnF	N.D.	N.D.	N.D.	N.D.	(2)
z475	C2H2 ZnF	780	600	700	N.D.	(2)
z481	C2H2 ZnF	580	450	490	N.D.	(2)
z487	C2H2 ZnF	330	210	520	Ubq	(2)

NOTES:
af: post-fertilization
Ubq: Ubiquitous
EM: Endomesoderm
ABO: Aboral Ectoderm
OE: Oreal Ectoderm
N.D.: No Data
NHR: Nuclear Hormone Receptor
TFs: Transcription Factors
ZnF: Zinc Finger
e: Trancript levels extrapolated from other time points in reference
NVD: Non-Vegetal Domain



Supplementary Figure 1: Pmar1 MOE screen. Graphs showing fold change in mRNA expression for *r of mic* candidate genes upon over-expression of Pmar1 mRNA. *delta* and *wnt8*

were used as positive and negative controls, respectively. A fold change of "1" (solid line) indicates "no change". The numbers situated above "1" indicate fold increase, and the numbers situated below "1", fold decrease (in logarithmic scale). A 3-fold or greater change was considered to be significant. White bars and grey bars represent data from samples at 9 h and 12 h of development, respectively. For each color, the two bars correspond to two independent batches of embryos. Error bars represent the standard deviation from three independent measurements on the same sample.

Supplementary Table 2: Folds of difference in expression levels of Ets1 and Tbr genes relative to uninjected embryos. Changes greater than 2.9 folds are considered significant.

	Tbr 18hr	Tbr 24hr	Ets1 18hr	Ets1 24hr
HesC MASO	2.23; 3.66	2.68; 6.60	3.65; 4.58	4.04; 6.03
Pmar1 MOE	4.12; 3.21	4.84; 1.80	5.01; 3.63	5.70; 2.64
Control	0.90; 0.60	1.25; 1.26	1.14; 1.01	0.59; 0.69

Experiment conducted by Mary Wahl.

REFERENCES

1. Howard-Ashby, M., Materna, S. C., Brown, C. T., Chen, L., Cameron, R. A. & Davidson, E. H. (2006) *Developmental Biology* **300,** 74-89.
2. Materna, S. C., Howard-Ashby, M., Gray, R. F. & Davidson, E. H. (2006) *Developmental Biology* **300,** 108-120.
3. Howard-Ashby, M., Materna, S. C., Brown, C. T., Chen, L., Cameron, R. A. & Davidson, E. H. (2006) *Developmental Biology* **300,** 90-107.
4. Tu, Q., Brown, C. T., Davidson, E. H. & Oliveri, P. (2006) *Developmental Biology* **300,** 49-62.
5. Rizzo, F., Fernandez-Serra, M., Squarzoni, P., Archimandritis, A. & Arnone, M. I. (2006) *Developmental Biology* **300,** 35-48.
6. Wei, Z., Angerer, L. & Angerer, R. (1999) *Development* **126,** 1729-1737.