

Appendix B

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

**Eric H. Davidson, 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, Peter J. C. Clarke[†], Alistair G. Rust, Zheng jun Pan, Maria I.
Arnone, Lee Rowen, R. Andrew Cameron, David R. McClay, Leroy Hood and
Hamid Bolouri**

(Published 2002)

Reprinted from *Developmental Biology*, Vol number 246, Davidson, E.H., Rast, J.P., Oliveri, P., Ransick, A., Calestani, C., Yuh, C.H., Minokawa, T., Amore, G., Hinman, V., Arenas-Mena, C., Otim, O., Brown, C.T., Livi, C.B., Lee, P.Y., Revilla, R., Schilstra, M.J., Clarke, P.J., Rust, A.G., Pan, Z., Arnone, M.I., Rowen, L., Cameron, R.A., McClay, D.R., Hood, L., Bolouri, H., Title of article A provisional regulatory gene network for specification of endomesoderm in the sea urchin embryo, Pages No. 162-190, Copyright (2002), with permission from Elsevier.

ABSTRACT

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: <http://www.its.caltech.edu/~mirsky/endomeso.htm> (End-mes Gene Network Update) for 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: *i*, use of morpholino antisense oligonucleotides targeted to many of the key regulatory genes in the network; *ii*, transformation of other regulatory factors into dominant repressors by construction of Engrailed repressor domain fusions; *iii*, ectopic expression of given regulatory factors, from genetic expression constructs and from injected mRNAs; *iv*, blockade of the β -catenin/Tcf pathway by introduction of mRNA encoding the intracellular domain of cadherin; and *v*, 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

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 6th 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 eight cells of the veg_1 ring bounded on top by the equatorial third cleavage plane and below by their eight sister cells of the veg_2 ring; plus four large and four 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 Etensohn (1993, 1996) by use of *diI* 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:

i. Initial specification of the veg₂ domain. The experimental evidence indicates that under normal conditions two inputs are required for the specification of veg₂ 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 6th cleavage veg₂ ring. This very early signaling function implies that, at least in some measure, the micromeres are already specified when they are born at 4th cleavage. New experimental insights into the mechanism by which micromere functions are confined to their lineage are reported by Oliveri *et al.* (2002), elsewhere in this Issue. The second input required for specification of the veg₂ 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 7th 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.

ii. 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 (Wikramanayake *et al.*, 2002). 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.

iii. The micromere Delta signal. During the 7th-9th 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 eight, 6th cleavage *veg*₂ cells). This event is specifically required for *veg*₂ mesodermal specification (see legend to Fig. 1).

*iv. Late specification of *veg*₁ 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, this Issue); and the *krox1* gene (C. Livi and E. Davidson, 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 were 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: STRATEGY FOR SOLVING THE NETWORK

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 amongst 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 in this Issue (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 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 three-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 in this Issue, *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/endomeso.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 amongst 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 Fig. 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 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 5th 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 (*op cit*).

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 expression (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 not useful in 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 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 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.* (2000). 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 veg₂ Endomesodermal Lineages

The initial zygotic phase of endomesoderm specification is completed during cleavage. As summarized above, the endomesoderm derives from the micromeres, veg₂, and part of the veg₁ lineages: the micromere lineage is born at 4th cleavage and their skeletogenic daughters at 5th cleavage; and the veg₂ lineage, all progeny of which contribute to endomesoderm, is born at 6th cleavage, together with their veg₁ sister cells. The early micromere signal to the grandparents and parents of the veg₂ lineage is passed between 4th and 6th 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 6th cleavage (Logan *et al.*, 1999); the Delta signal from the skeletogenic micromeres to the adjacent veg_2 blastomeres is passed between 7th and about 9th 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 4th and 9th 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 Sox1 (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. Oliveri and E. Davidson, unpublished data). As described in detail elsewhere in this

issue (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, 5; Kurokawa *et al.*, 1999; Oliveri *et al.*, 2002; G. Amore and E. Davidson, unpublished data). Ectopic expression of *pmar1* mRNA causes global derepression of these genes and a general, irreversible conversion of the embryonic blastomeres to cells 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. Amore and E. Davidson, 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 *veg2* 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 *veg2* endomesoderm lineage derives, and the nuclearization of β -catenin in the

veg₂ 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 of 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 Soxb 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 *$\beta1/2otx$* systems (see web site QPCR data, *op cit*). Zygotic activation of the *$\beta1/2otx$* *cis*-regulatory system occurs only toward the end of the 4th-9th cleavage period here considered (see Yuh *et al.*, 2002, this Issue, 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 amongst 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 *veg2* founder cells, in that cells expressing *pmar1* ectopically cause adjacent cells to express *endo16*, just as do transplanted 4th cleavage micromeres (Ransick and Davidson, 1993; Oliveri *et al.*, 2002). The ultimate transcriptional targets of the early signal in the *veg2* lineages are unknown, but as mentioned above it 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 (Oliveri *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 8th Cleavage to Mesenchyme Blastula

As the blastula stage begins after about 8th 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 veg₁ endoderm domain exactly as *wnt8* expression does after 24 h, and this plus the strength 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, we here see why introduction of a dominant negative form of Wnt8 (Wikramanayake *et al.*, 2002), or interference with β -catenin nuclearization by introduction 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 *$\beta1/2otx$* cis-regulatory system (Table 1; Yuh *et al.*, 2002, this Issue). 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 *veg₂* 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, this Issue). 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. Ransick and E. Davidson, 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 midlate 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 $\beta 1/2otx$ transcription unit becomes active at the beginning of the period covered by Fig. 4 (unpublished data of C. Livi and E. Davidson). 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 earlier used 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 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. Cameron, 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, α Otx 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 veg₂ Mesoderm Cells just before Gastrulation

Figure 6 provides a similar picture for the veg₂ mesoderm domain just before gastrulation and delamination of its mesenchymal cell types. At this stage the prospective mesodermal cells display processed intracellular N (in *Lytechinus 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. Oliveri and E. Davidson, unpublished data). As it becomes active in the veg₂ 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 and also for *delta* 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 and 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 here shown), 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 *veg₂* 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. Ransick and E. Davidson, 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 *veg₂* mesoderm, like the *veg₂* 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, feedback 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 Fig. 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.

ACKNOWLEDGEMENTS

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, *op cit*, and in papers elsewhere in this issue, *viz* Ransick *et al.*, 2002; Rast *et al.*, 2002; Oliveri *et al.*, 2002; Brown *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 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. 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 NCRN (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 NICHD. 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.*, in press.
- 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," p. 85. Appleton, New York.]
- Boveri, T. (1918). Zwei Fehlerquellen bei Merogonieversuchen und die Entwicklungsfähigkeit merogonischer partiellmerogonischer Seeigelbastarde. *Arch. Entwicklungsmech. Org.* **44**, 417-471.
- Brown, C. T., Rust, A. G., Clarke, P. J. C., Pan, Z., Schilstra, M. J., De Buyscher, T., Griffin, G., Cameron, R. A., Davidson, E. H. and Bolouri, H. (2002). New computational approaches for computational-experimental analysis of *cis*-regulatory networks. *Dev. Biol.*, in press.
- Cameron, R. A. and Davidson, E. H. (1997). LiCl perturbs ectodermal *veg1* 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. Dynam.* **207**, 450-460.

- Emily-Fenouil, F., Ghiglione, C., Lhomond, G., Lepage, T. and Gache, C. (1998). GSK3 β /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 *Strongylocentrotus 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., 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, 1997.
- 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., 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.*, in press.
- 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 veg₁ 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.*, in press.
- 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.*, submitted for publication.
- 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 Etensohn, 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 Etensohn, C. A. (2002). LvDelta is a mesoderm-inducing signal in the sea urchin embryo and can endow blastomeres with organizer-like properties. *Development*, in press.
- 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 *Strongylocentrotus 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.

- Wikramanayake, A. H., Peterson, R., Chen, J., Huang, L., McClay, D. R. and Klein, W. H. (2002). Selective expression of Wnt8 in vegetal cells of the early sea urchin embryo mediates mesoendoderm specification in a nuclear β -catenin-dependent manner. Submitted for publication.
- 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., Comisso, 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 *endol6* 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.*, in press.
- Zhu, X., Mahairas, G., Illies, M., Cameron, R. A., Davidson, E. H. and Etensohn, C. A. (2001). A large-scale analysis of mRNAs expressed by primary mesenchyme cells of the sea urchin embryo. *Development.* **128**, 2615-27.

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: XSox β 17 α/β and XSox3 physically interact with β -catenin. *Mol. Cell* **4**, 487-498.

Table 1. Experimental Approaches to the Endomesoderm Network

| Requirement | Method | Ref |
|---|--|--|
| Gene discovery | Differential macroarray screen ¹ <ul style="list-style-type: none"> •cadherin vs. LiCl •DnN vs. LiCl •Ectopic <i>bra</i> vs. Bra MASO² Homology screens bearing probes from other species | Rast <i>et al.</i> , 2000 Ransick <i>et al.</i> , 2002 Calestani, unpublished data Rast <i>et al.</i> , 2002 Table 3 |
| Finding <i>cis</i> -regulatory elements | Interspecific sequence comparison <ul style="list-style-type: none"> •BAC libraries from <i>S. purpuratus</i> and <i>L. variegatus</i> •Isolate BACs containing relevant genes³ •Obtain and annotate genomic sequences⁴ •Determine and test conserved elements⁵ | Brown <i>et al.</i> , 2002 Cameron <i>et al.</i> , 2000 Table 2 Yuh <i>et al.</i> , 2002 |
| Gene expression characterization | WMISH QPCR | Table 3 Table 3 |
| Perturbation analysis | Cadherin MOE DnN MOE MASO (many genes) Engrailed fusions (five genes) Transcription factor MOE (one gene) | Ransick <i>et al.</i> , 2002; Table 3; Appendix 1 " " " " Oliveri <i>et al.</i> , 2002 |

Notes to Table 1

Note 1. Subtractive hybridization, in which single-stranded driver RNA lacking a given class of endomesodermal sequences is reacted to high driver C₀t 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 Q-Bot 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 <5 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, this Issue).

Note 2. 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.

Note 3. 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.

Note 4. 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 $\geq 8x$ 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).

Note 5. 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.

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.

Table 2. BAC Clones Sequenced for the Endomesoderm Network Project

| Gene | Clone name | Original source of <i>S. purpuratus</i> cDNA probe | Size (kb) | Sequencing center |
|--------------------------------------|----------------|--|-----------|-------------------|
| <i>Sp apo bec</i> | 112F15 | Brachyury target screen | 50 | ISB |
| <i>Lv apo bec</i> | 031L12 | HS | 55 | ISB |
| <i>Lv β-catenin</i> | 48J5 | Miller and McClay, 1997; HS | 48 | ISB |
| <i>Sp brachyury</i> | 117A3 | Peterson <i>et al.</i> , 1999 | 146 | JGI, ISB |
| <i>Lv brachyury</i> | 187F03 | HS | 70 | ISB |
| <i>Sp capk</i> | 83N11 | Brachyury target screen | 56 | ISB |
| <i>Lv capk</i> | 177F13 | HS | 49 | ISB |
| <i>Lv decorin</i> | 114C22 | Dominant negative Notch screen; HS | 70 | ISB |
| <i>Sp delta</i> | 046A16 | Zhu <i>et al.</i> , 2001 | 157 | JGI |
| <i>Lv delta</i> | 20B17, 71J9 | Cadherin over-expression screen; HS | 70 | ISB |
| <i>Sp dpt</i> | 188B16 | Brachyury target screen | 36 | JGI |
| <i>Lv dpt</i> | 037C14 | HS | 60 | ISB |
| <i>Sp eve</i> | 079A02 | Cadherin over-expression screen | 173 | JGI |
| <i>Lv eve</i> | 112E15 | HS | 63 | ISB |
| <i>Sp foxa</i> | 041I19 | Harada <i>et al.</i> , 1996; HS | 150 | JGI |
| <i>Lv foxa</i> | 004G18 | Harada <i>et al.</i> , 1996; HS | 55 | ISB |
| <i>Lv foxb</i> | 92H22 | Luke <i>et al.</i> , 1997; HS | 87 | ISB |
| <i>Sp gatac</i> | 081C18 | Pancer <i>et al.</i> , 1999 | 140 | JGI |
| <i>Lv gatac</i> | 044D13 | HS | 71 | ISB |
| <i>Sp gatae</i> | 091A10 | Pancer <i>et al.</i> , 1999 | 184 | JGI |
| <i>Lv gatae</i> | 032P20 | HS | 71 | ISB |
| <i>Sp gcm</i> | 033O18 | Cadherin over-expression screen | 57 | ISB |
| <i>Lv gcm</i> | 018J3 | HS | 70 | ISB |
| <i>Sp gelsolin-like</i> | 118E15 | Brachyury target screen | 50 | ISB |
| <i>Lv hmx</i> | 076P20 | Martinez and Davidson, 1997 | 70 | ISB |
| <i>Sp hox11/13b</i> | 135O12 | Dobias <i>et al.</i> , 1996 | 125 | JGI |
| <i>Lv hox11/13b</i> | 235L15 | HS | 50 | ISB |
| <i>Lv kakapo</i> | 229D15 | Brachyury target screen; HS | 40 | ISB |
| <i>Sp kakapo</i> | 12G10 | Brachyury target screen | 50 | ISB |
| <i>Sp krox</i> | 163O19 | Wang <i>et al.</i> , 1996 | 114 | JGI |
| <i>Lv krox</i> | 060B16 | HS | 56 | ISB |
| <i>Sp lim</i> | 108P4 | Kawasaki <i>et al.</i> , 1999; HS | 158 | JGI |
| <i>Lv lim</i> | 097A18 | HS | 50 | ISB |
| <i>Lv not</i> | 219N8 | Peterson <i>et al.</i> , 1999 | 55 | ISB |
| <i>Lv notch</i> | 97A7 | Sherwood and McClay, 1997; HS | 71 | ISB |
| <i>Sp notch</i> | 191I13 | Sherwood and McClay, 1997; HS | 178 | JGI |
| <i>Lv nr1l</i> | 024M4 | Brachyury target screen; HS | 70 | ISB |
| <i>Sp orct</i> | 095C14 | Brachyury target screen | 57 | ISB |
| <i>Sp otx</i> | 006F13 | Li <i>et al.</i> , 1997 | 160 | JGI |
| <i>Lv otx</i> | 229L5 | HS | 62 | ISB |
| <i>Sp pks</i> | 080H21 | Dominant negative Notch screen | 147 | JGI |
| <i>Lv pks</i> | 53J24 | HS | 38 | ISB |
| <i>Lv pmar1</i> | 170H13 | Oliveri <i>et al.</i> , 2001; HS | 62 | ISB |
| <i>Lv soxb</i> | 208L3 | Kenny <i>et al.</i> , 1999; HS | 70 | ISB |
| <i>Sp soxb</i> | 58L24 | Kenny <i>et al.</i> , 1999; HS | 150 | ISB |

| | | | | |
|-------------------|--------|------------------------------------|-----|-----|
| <i>Lv t-brain</i> | 192I24 | Croce <i>et al.</i> , 2001; HS | 48 | ISB |
| <i>Sp t-brain</i> | 31J8 | Croce <i>et al.</i> , 2001; HS | 130 | OSB |
| <i>Sp wnt8</i> | 099F11 | Ferkowicz <i>et al.</i> , 1998; HS | 135 | JGI |
| <i>Lv wnt8</i> | 183H12 | Ferkowicz <i>et al.</i> , 1998 | 64 | ISB |

Notes to Table 2:

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.

| Abbrev | Protein encoded | Spatial Expression | | | Class | Source of Gene, Information | Inputs | |
|--------------------|---|--------------------|--------------------|---------|--------------------------|--|--|---|
| | | 15-18 h | 20-24 h | 30-40 h | | | Source | Data type |
| <i>apobec</i> | Cytidine deaminase | unk | EM | E | RNA editing enzyme | Brachyury target screens | Bra | Bra MASO |
| <i>bra</i> | Brachyury | EM | E | E, OE | Transcription factor | Peterson <i>et al.</i> , 1999; Gross and McClay, 2001 | GataE FoxA Krox | GataE MASO, CCRA, ECRA FoxA MASO, CCRA, ECRA Krox-En |
| <i>capk</i> | c-AMP protein kinase | unk | M | SMC | Protein-modifying enzyme | Brachyury target screens | GataC Tcf Elk Otx ⁺ Bra ¹ | GataC MASO, CCRA Cad MOE, CCRA Elk-En Otx-En Bra MASO |
| <i>cyclophilin</i> | Cyclophilin, peptidyl-prolyl <i>cis-trans</i> isomerase | PMC | PMC | PMC | Protein-modifying enzyme | Zhu <i>et al.</i> , 2001; G. Amore <i>et al.</i> , unpublished | unk act | Pmar1-En, Pmar1 MOE |
| <i>dri</i> | DeadRinger | PMC | PMC, OE | OE | Transcription factor | 40h-7h subtractive screen | unk rep | Pmar1-En, Pmar1 MOE |
| <i>decorin</i> | Decorin | unk | M | unk | Extracellular matrix | Dominant negative Notch screen | unk | |
| <i>delta</i> | Delta | m (8-18 h) | PMC (20), SMC (24) | | Ligand | Zhu <i>et al.</i> , 2001 | Tcf unk rep Ub act ⁴ E(S) ¹⁴ | Cad MOE, CCRA Pmar1-En, Pmar1 MOE Unk DnN MOE |
| <i>dpt</i> | d-Dopachrome tautomerase | unk | M | SMC | Enzyme | Brachyury target screens | Bra ¹ E(S) ⁸ Gem Elk ⁵ | Bra MASO DnN MOE Gem MASO Elk-En |
| <i>elk</i> | Ets family factor | unk | E | E, OE | Transcription factor | M. Arnone <i>et al.</i> , unpublished | GataE | GataE MASO |

| | | | | | | | | |
|----------------|-----------------------------------|-------|-----|---------|----------------------|--------------------------------------|--|---|
| <i>endo16</i> | Endo16 | EM | EM | E | Cell surface protein | Nocente-McGrath <i>et al.</i> , 1989 | Otx ⁺ | Otx-En (Li <i>et al.</i> , 1991; Yuh <i>et al.</i> , 2001), ECRA |
| <i>eptx</i> | p33/HEH epoxide hydrolase | | PMC | PMC | Enzyme | Brachyury target screens | Ui GataE ¹⁰ Bra ¹ | ECRA GataE/MASO Bra MASO |
| <i>ets</i> | Ets | Ma-Ub | PMC | PMC | Transcription factor | Kurokawa <i>et al.</i> , 1999 | unk rep | Pmar1 MOE, Pmar1-En |
| <i>eve</i> | Even-skipped orthologue | E | E | Veg.(E) | Transcription factor | Cadherin over-expression screen | Tcf Krox Otx ¹² | Cad MOE Krox-En Otx-En |
| <i>ficolin</i> | Ficolin-like | PMC | PMC | PMC | TGFβ membrane BP | Incidental screen | unk act Bra ¹ | Pmar1-En, Pmar1 MOE Bra MASO |
| <i>foxa</i> | Hepatocyte NF 3 orthologue (FoxA) | EM | E | E | Transcription factor | Harada <i>et al.</i> , 1996; HS | FoxA Otx GataE GataC Tcf E(S) ³ Dri ² Elk | FoxA MASO, CCRA Otx-En GataE/MASO GataC/MASO Cad MOE DnN MOE Dri MASO Elk-En |
| <i>foxb</i> | Winged helix factor | ND | E | E | Transcription factor | Luke <i>et al.</i> , 1997 | Krox Tcf GataE GataC FoxA Otx ^{+,12} | Krox-En Cad MOE GataE/MASO GataC/MASO FoxA MASO, CCRA Otx-En |
| <i>fmo</i> | Flavine mono-oxygenase | unk | M | SMC | Enzyme | Dominant negative Notch screen | E(S) Gcm | DnN MOE |
| <i>gatac</i> | GataC | ND | M | ND | Transcription factor | Pancer <i>et al.</i> , 1999 | Otx GataE ¹⁴ Elk Tcf | Otx-En GataE/MASO Elk-En Cad MOE |
| <i>gatae</i> | GataE | EM | EM | E | Transcription | Pancer <i>et al.</i> , 1999 | Otx | Otx-En |

| | | | | | | factor | | | | E(S) | |
|------------------|-----------------------------|-----|-----|-----|--|--------------------------|---|--|--|--|---|
| <i>gell</i> | Gelsolin-like | unk | EM | E | | Actin-modulating protein | Brachyury target screens | | | Tcf | DnN MOE Cad MOE Bra MASO |
| <i>gcm</i> | Gial cells missing | M | M | SMC | | Transcription factor | Cadherin over-expression screen | | | Gcm Elk Tcf | Gcm MASO Elk-En Cad MOE DnN MOE |
| <i>hmx</i> | Homeodomain DBP | unk | unk | EM | | Transcription factor | Martinez and Davidson, 1997 | | | E(s) ⁺ , ⁰ Otx Krox Tcf | Otx-En Krox-En Cad MOE GataC MASO DnN MOE GataE MASO |
| <i>hnf6</i> | Hepatocyte NF 6 orthologue | unk | unk | unk | | Transcription factor | K. Makabe, unpublished; HS | | | Pmar1 | Pmar1-En, Pmar1 MOE |
| <i>hox11/13b</i> | Homeobox protein, Hox11/13b | Ub | Ub | E | | Transcription factor | Dobias <i>et al.</i> , 1996; Martinez <i>et al.</i> , 1999 | | | Krox Tcf Otx ¹² | Krox-En Cad MOE Otx-En |

| Abbrev | Protein encoded | Spatial Expression | | | Class | Source of Gene, Information | Inputs | |
|---------------|-------------------------|--------------------|---------|---------|----------------------|-------------------------------|--|--|
| | | 15-18 h | 20-24 h | 30-40 h | | | Source | Data type |
| <i>kakapo</i> | Calponin domain protein | unk | EM | E | Cytoskeletal protein | Brachyury target screens | Bra | Bra MASO |
| <i>krox1</i> | Krox | E | E | E | Transcription factor | Wang <i>et al.</i> , 1996 | Otx Krox Tcf | Otx-En Krox-En Cad MOE |
| <i>krl</i> | Krüppel-like | EM | unk | unk | Transcription factor | Howard <i>et al.</i> , 2001 | Tcf | Cad-MOE (Howard <i>et al.</i> , 2001) |
| <i>lim1</i> | LIM-1 | Veg1, OE | OE | OE | Transcription factor | Kawasaki <i>et al.</i> , 1999 | Tcf Otx GataE [®] Elk ⁶ Krox | Cad MOE Otx-En GataE MASO Elk-En Krox-En |

| <i>msh130</i> | MSP130 | PMC | PMC | PMC | PMC | Cell surface protein | Parr <i>et al.</i> , 1990 | Dri unk act | Dri MASO Pmar1-En, Pmar1 MOE |
|--------------------|--|-----------------------|---------|---------|--------------------|-----------------------------|--------------------------------|--|--|
| <i>msh130-like</i> | MSP130-like | PMC | PMC | PMC | PMC | Cell surface protein | Parr <i>et al.</i> , 1990 | unk act | Pmar1-En, Pmar1 MOE Bra MASO |
| <i>not</i> | Not | unk | M | M | SMC | Transcription factor | Peterson <i>et al.</i> , 1999 | Otx GataE [@] GataC | Otx-En GataE MASO GataC MASO |
| <i>notch</i> | Notch | Ma-Ub (protein) ND | Ub | Ub | unk | receptor | Sherwood and McClay, 1997 | E(S) ¹³ | |
| <i>nrl-1</i> | Neutralized-like-1 | ND | M | M | PMC, M | Signaling intermediate | Brachyury target screens | GataE Bra Otx ³ E(S) | GataE MASO Bra MASO Otx-En DnN MOE Bra MASO |
| <i>orct</i> | Organic Cation Transporter | unk | EM | EM | EM | Transporter | Brachyury target screens | Bra | Bra MASO |
| <i>otxβ1/2</i> | Orthodenticle orthologue (β1/2 transcription unit) | unk | OE | OE | E, OE | Transcription factor | Li <i>et al.</i> , 1997 | Tcf GataE Otx ⁺ Krox | Cad MOE GataE MASO, CCRA Otx-En, CCRA Krox-En, CCRA |
| <i>pks</i> | Polyketide synthase | unk | M | M | SMC | Enzyme | Dominant negative Notch screen | E(S) ⁷ Bra ¹ GataE Gem Elk | DnN MOE Bra MASO GataE MASO Gem MASO Elk-En |
| <i>pmar1</i> | Paired-class homeodomain protein | m (6-12h), ND | ND | ND | ND | Transcription factor | Oliveri <i>et al.</i> , 2002 | Otx Tcf | Otx-En Cad MOE |
| Abbrev | Protein encoded | Spatial Expression | | | Class | Source of Gene, Information | Inputs | | |
| | | 15-18 h | 20-24 h | 30-40 h | | | Source | Data type | |
| <i>sm27</i> | Spicule matrix protein-27 | ND | PMC | PMC | Structural protein | Harkey <i>et al.</i> , 1995 | unk act | | Pmar1-En, Pmar1 MOE |
| <i>Sm30</i> | Spicule matrix protein-30 | ND | PMC | PMC | Structural protein | George <i>et al.</i> , 1991 | Dri | | Dri MASO |

| | | | | | | | | | |
|--------------|---------------------------|-----------------------|-----|-----|-----|----------------------------|--|---|--|
| <i>sm37</i> | Spicule matrix protein-37 | unk | PMC | PMC | PMC | Structural protein | Lee <i>et al.</i> , 1999 | | No inputs |
| <i>sm50</i> | Spicule matrix protein-50 | PMC | PMC | PMC | PMC | Structural protein | Sucov <i>et al.</i> , 1987; Katoh-Fukui <i>et al.</i> , 1991 | Ets ¹¹ Ets Dri Bra | neg-Ets (Kurokawa <i>et al.</i> , 1999) Pmar1-En, Pmar1 MOE Dri MASO Bra MASO |
| <i>soxbl</i> | Sox protein | Ma-Ub | Ec | Ec | unk | HMG transcription factor | Kenny <i>et al.</i> , 1999 | Krl | Krl MASO (Howard <i>et al.</i> , 2001) |
| <i>sutx</i> | Sulfotransferase | unk | M | M | SMC | Enzyme | Dominant negative Notch screen | E(S) | DnN MOE |
| <i>tbr</i> | T-brain | Ma-Ub, m (>8), Zy-PMC | PMC | PMC | PMC | T-box transcription factor | Croce <i>et al.</i> , 2001; HS | Ets unk rep Bra ¹ | Neg-Ets (Kurokawa <i>et al.</i> , 2000) Pmar1-En, Pmar1 MOE Bra MASO |
| <i>ui</i> | Unknown | unk | unk | unk | unk | Transcription factor | C.-H. Yuh <i>et al.</i> , unpublished | Tcf ⁹ GataE ⁹ | Cad MOE GataE MASO |
| <i>wnt8</i> | Wnt8 | m (5), EM (10-15), | E | E | unk | Wingless family ligand | A. Wikramanayake & W. Klein, unpublished | Tcf GataE ¹⁵ Krox Otx | Cad MOE GataE MASO Krox-En Otx-En |

Notes to Table 3

¹Ubiquitous activating input inferred from ubiquitous expression in presence of Pmar1 MOE and Pmar1-En.

²Role 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.

³Otx-En effect cannot be rescued by GataE MOE.

⁴Otx-En effect cannot be rescued by Krox MOE.

⁵GataE MASO effect cannot be rescued by Otx MOE.

†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.

*Effect concluded to be indirect or irrelevant: see note regarding specific data in Appendix

ABREVIATIONS.

Spatial Expression

The expression patterns are indicated by abbreviations for each embryonic territory: Ma, maternal; Ub, ubiquitous; m, micromere; E, definitive endoderm; EM, veg₂ endomesoderm; M, veg₂ mesoderm; OE, oral ectoderm; PMC, primary skeltogenic mesenchyme; SMC, secondary mesenchyme cells of veg₂ 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 hours after fertilization.

Data Type

Experimental manipulations used for quantitative PCR are : antisense morpholino oligonucleotide injection (MASO); gene overexpression (GOE); 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 below:

Cad MOE: Cadherin RNA overexpressioin (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)

FIGURE LEGENDS

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 4th-6th 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. McClay, 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 *veg2* mesodermal domain; the production of excess mesodermal cell types on introduction of constitutively active N receptor; and the failure of *veg2* 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 *veg1* endoderm, see Logan and McClay (1997); Ransick and Davidson (1998). The late endodermal specification of *veg1* cells depends on a signal from *veg2*, which is likely to be Wnt8 (McClay and Logan, 1996; Davidson *et al.*, 1998; unpublished data of D. McClay).

(B) Some representative developmental stages: 5 h, 4th 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 6th cleavage. 6 h, 5th cleavage, 28-cell stage, vegetal view. The micromeres in the center directly above the six veg_2 cells. 12 h, 8th 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. 15 h, early 9th cleavage, approximately 300 cells, viewed from the side. There are a total of nine cleavages in *S. purpuratus*, but these are increasingly asynchronous after 6th cleavage. 24 h, mesenchyme blastula stage, approximately 650 cells, viewed from the side. 36 h, 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.

Figure 1

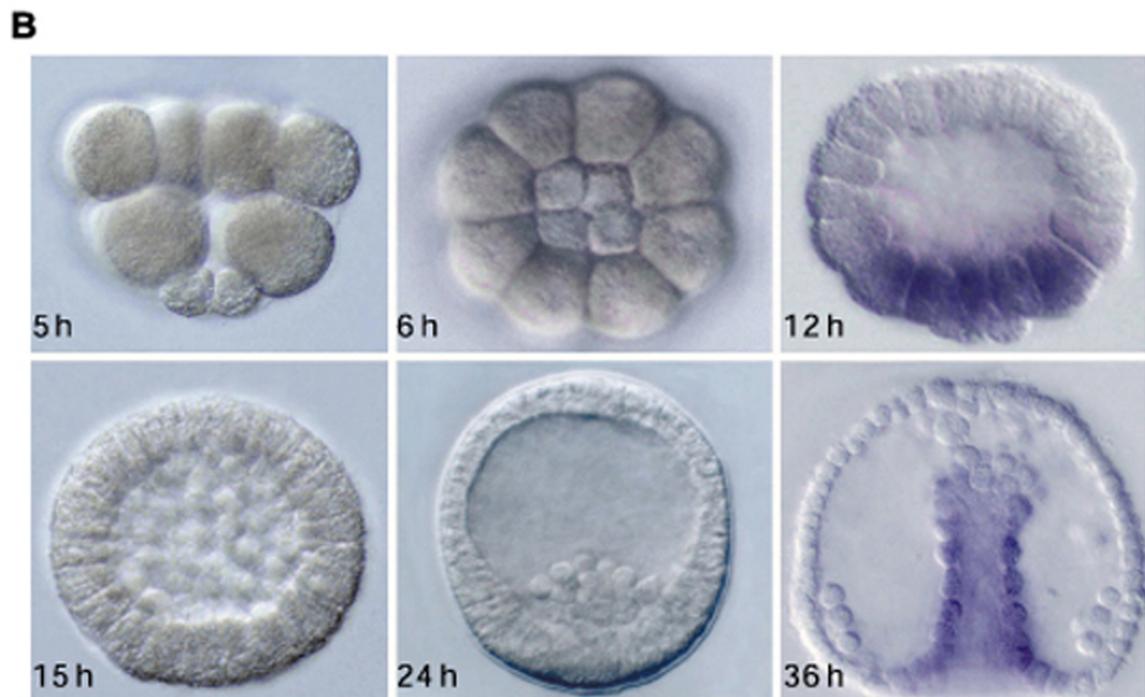
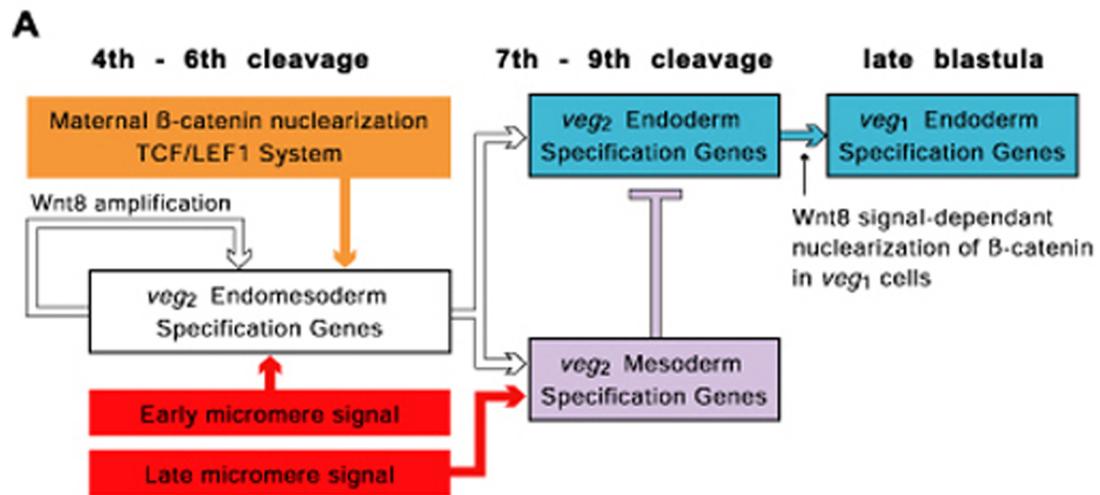


FIG. 2. Regulatory gene network for endomesoderm specification: the view from the genome. The architecture of the network is based on perturbation and expression data listed in Table 3 and compiled in Appendix, on data from *cis*-regulatory analyses for several genes, and on other experiments discussed in text. See <http://www.its.caltech.edu/~mirsky/endomeso.htm> (End-mes Gene Network Update) for current version of the model in this Figure, and <http://www.its.caltech.edu/~mirsky/qpcr.htm> (End-mes Network QPCR Data) for current list of quantitative results of perturbation experiments and temporal details. Each short horizontal line from which bent arrows extend represent 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 ($c\beta$) nuclearization system (χ) which autonomously causes accumulation of β -catenin in the nuclei of all future endomesodermal cells. This transcriptional system ($n\beta$ /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 4th 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 left represents the skeletogenic micromere (pmc) domain prior to ingression; the light green area indicates the veg₂ 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 veg₁ 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.

Figure 2

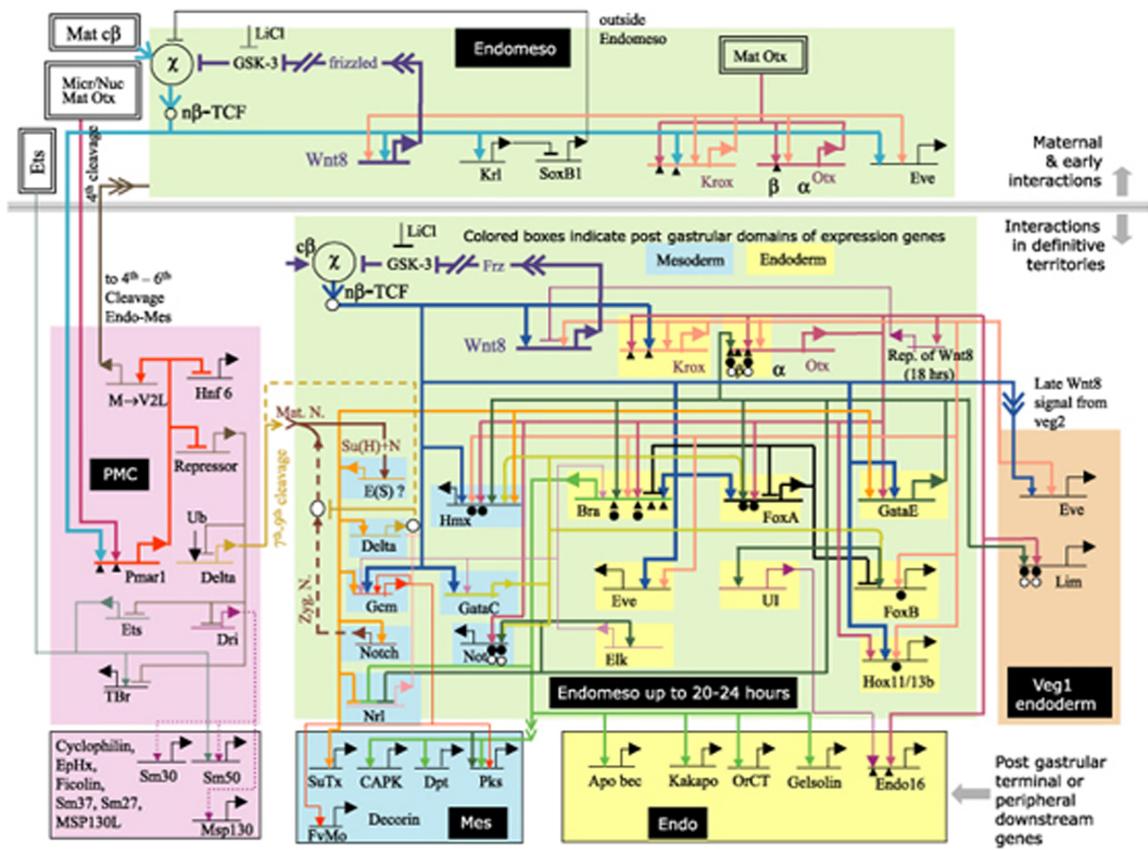
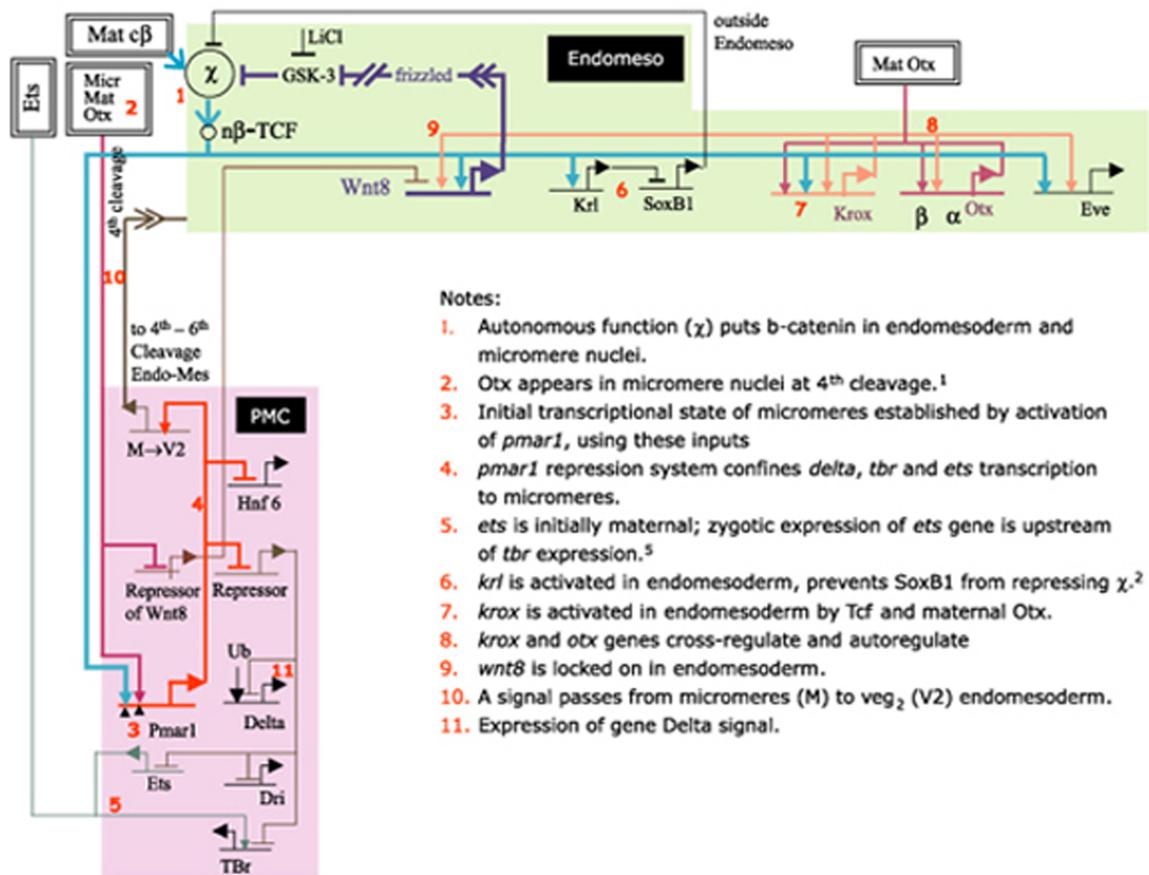


FIG. 3. Initial specification of micromere and veg_2 endomesodermal domains. The diagram includes the portion of the model in Fig. 2 that includes events occurring from about 4th to 9th cleavages within these domains. The "notes" (red numerals) indicate specific functional aspects. See legend to Fig. 2 for symbolism and architectural explanation of the form of the model. References: ¹Chuang et al. (1996); ²Kenny *et al.* (1999); ³Ransick and Davidson (1993, 1995); ⁴Oliveri *et al.* (2002); ⁵Kurokawa *et al.* (1999); unpublished data of K. Akasaka.

Figure 3

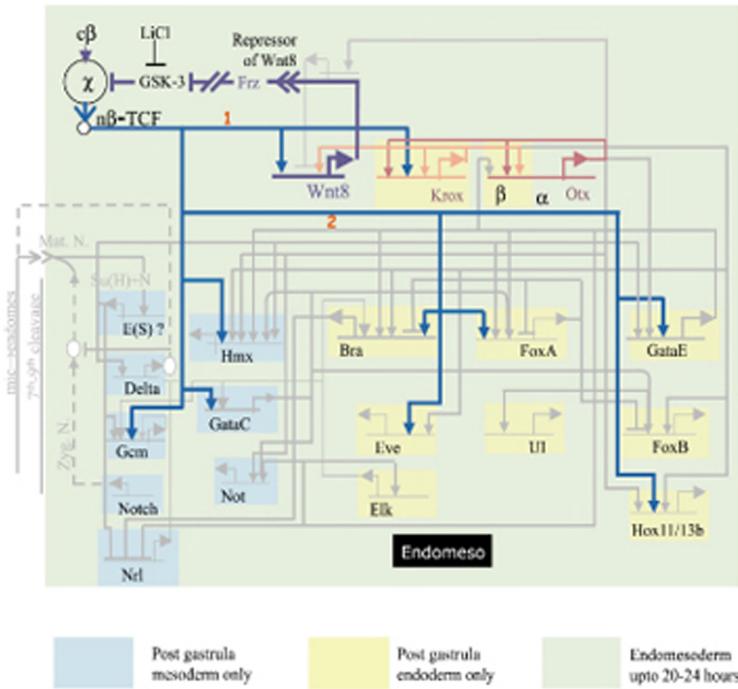


Notes:

1. Autonomous function (χ) puts b-catenin in endomesoderm and micromere nuclei.
2. Otx appears in micromere nuclei at 4th cleavage.¹
3. Initial transcriptional state of micromeres established by activation of *pmar1*, using these inputs
4. *pmar1* repression system confines *delta*, *tbr* and *ets* transcription to micromeres.
5. *ets* is initially maternal; zygotic expression of *ets* gene is upstream of *tbr* expression.⁵
6. *krl* is activated in endomesoderm, prevents SoxB1 from repressing χ .²
7. *krox* is activated in endomesoderm by Tcf and maternal Otx.
8. *krox* and *otx* genes cross-regulate and autoregulate
9. *wnt8* is locked on in endomesoderm.
10. A signal passes from micromeres (M) to veg₂ (V2) endomesoderm.
11. Expression of gene Delta signal.

FIG. 4. Views from the nuclei of the *veg*₂ endomesoderm from about 8th cleavage to the 20-24 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: ¹Logan *et al.* (1999); ²A. Wikramanayake *et al.* (2002); ³Yuh *et al.* (2002); C. Livi and E. Davidson, unpublished data.

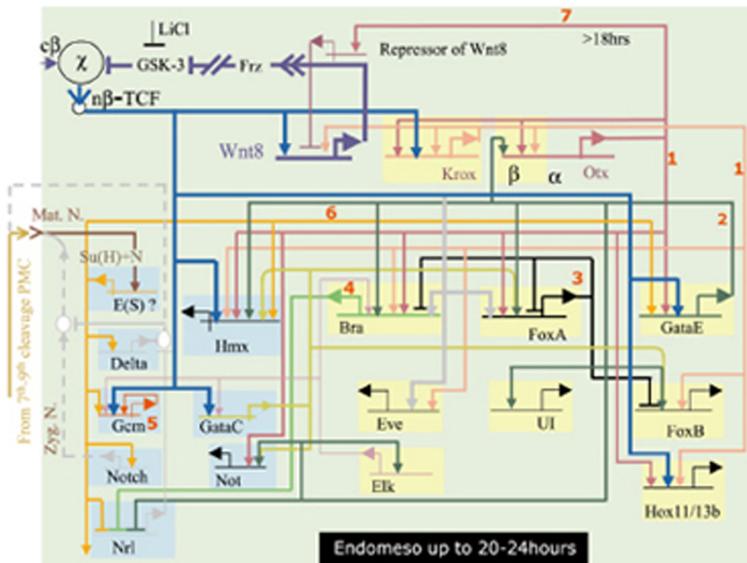
Figure 4A



Notes:

1. β -catenin/Tcf input is now produced by a zygotic signaling loop driven by Wnt8 expression in endomesoderm cells.
2. β -catenin/Tcf input is required for expression of many regulatory genes that become active in the veg₂ endomesodermal territory during early-mid blastula stage.

Figure 4B

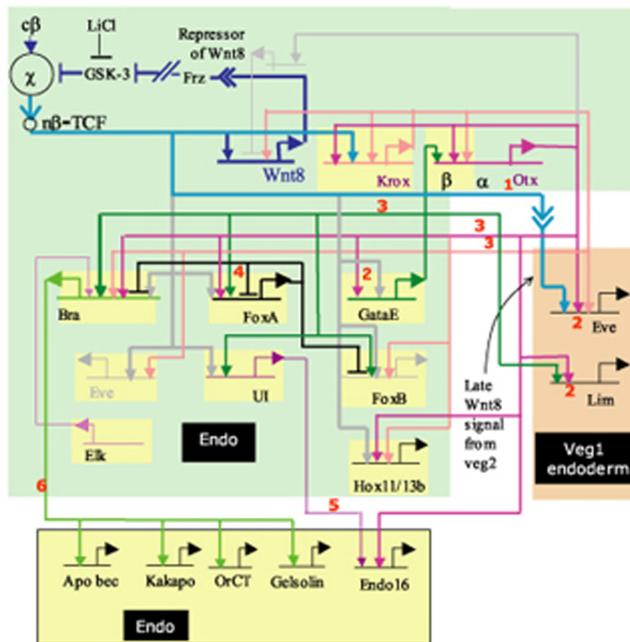


Notes:

1. *krox1* and *otx* provide inputs into other endomesodermal regulatory genes, linking them with the initial endomesodermal specification apparatus.
2. The *gatae* gene is a major regulator of other endomesodermal regulatory genes; it participates in a positive regulatory loop with the $\beta1/2otx$ regulatory element (β).
3. The *foxa* gene acts as a repressor of other endodermal genes, and of itself.
4. The *bra* gene is active in the blastoporal endoderm where it is required for gastrulation.
5. The *gcm* regulatory gene is a primary target of the Delta-Notch signal from the micromeres.
6. At this stage the Notch signal transduction system is linked to the *gatae* gene and its downstream regulatory targets.
7. At late blastula stage the wnt8- β -catenin/Tcf system ceases to function,^{1,2} possibly due (indirectly) to accumulation of Otx factor³ which may activate a repressor of wnt8.

FIG. 5. View from the nuclei of veg_2 and veg_1 endoderm after 24 h and to the beginning of gastrulation.

Figure 5

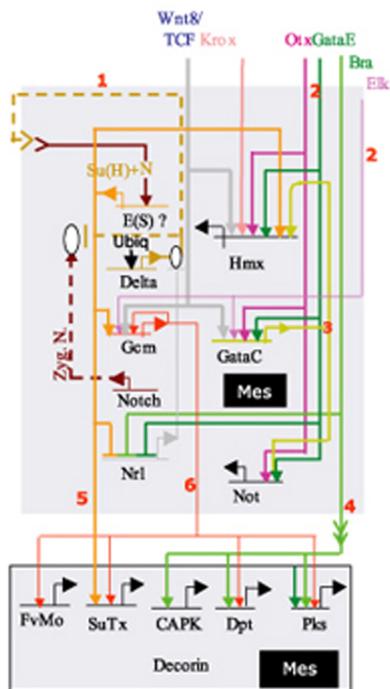


Notes:

1. β -Catenin/Tcf transcriptional inputs appear now in prospective veg₁ endoderm, probably induced by Wnt8 signals from edge of veg₂ endoderm.
2. Result is activation of regulatory genes in the veg₁ 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

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.

Figure 6



Notes:

1. Delta expression appears in definitive veg₃ 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.

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: |
|--------------------|-----------------------|-----------------|------------------------------|---------------------------------------|---------|-----------------------------|
| <i>apobec</i> | Bra MASO | | | -3.8 | | J. Rast |
| <i>bra</i> | 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 |
| | 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 |
| <i>capk</i> | Bra MASO ¹ | | | -1.7 | | J. Rast |
| <i>cyclophilin</i> | Pmar1 MOE | +3.3 | | +5.5 | | P. Oliveri |
| | Pmar1-En | NS | | +3.4 | | P. Oliveri |
| <i>delta</i> | Cad MOE ¹⁵ | -3.3/-3.9 | -1.6/NS | NS/-1.9 | | A. Ransick & T. Minokawa |
| | Pmar1 mRNA | +1.8/+3.2 | | +2.6/+2.6 | | P. Oliveri |
| | Pmar1-En | +1.8/+2.0 | | +3.1/+2.0 | | P. Oliveri |
| <i>dpt</i> | Bra MASO ¹ | | | -6.8 | | J. Rast |
| | DnN MOE ⁸ | | -1.7 | | | C. Calestani |
| | Elk-En ⁵ | | | -2.1,-2.2,- 4.6 | | M. Arnone |
| <i>dri</i> | Pmar1 MOE | +3.6 | | +6.7 | | P. Oliveri |
| | Pmar1-En | +2.4 | | +7.2 | | P. Oliveri |
| <i>elk</i> | GataE MASO | | +1.5/NS, NS/+1.2, +1.7 | -3.8,-3.0/ NS, NS | | P.-Y. Lee |
| <i>endo16</i> | 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 |
| | GataE-MASO | | NS/NS | -2.8, -3.5/ -2.7, -1.8 | | P.-Y. Lee |
| <i>ephx</i> | Bra-MASO ¹ | | | -3.8 | | J. Rast |
| <i>ets</i> | Pmar1 MOE | NS | | +5.4 | | P. Oliveri |
| | Pmar1-En | NS | | +4.5 | | P. Oliveri |

| Gene | Perturbation | 12-15 | 18-21 | 24-27 | 30-36 | Author |
|----------------|-----------------------------|--------------------------|----------------------------------|---|---------------------------------------|-----------------------------|
| <i>eve</i> | 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 |
| <i>ficolin</i> | Bra MASO ¹ | | | -1.7 | | J. Rast |
| | Pmar1 MOE | +4.0 | | +5.9 | | P. Oliveri |
| | Pmar1-En | NS | | +3.1 | | P. Oliveri |
| <i>foxa</i> | Otx-En | | NS/NS/- 2.8 | -3.3/-1.4/ -4.1/-2.8 | | A. Ransick & T. Minokawa |
| | 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 |
| | Dri MASO ² | | | | -1.1/-2.0/ -1.8/-1.8/ -1.8/-3.5 | G. Amore |
| | 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 |
| <i>foxb</i> | Krox-En | | -5.0 | -5.7/-4.2 | | C. Livi |
| | GataE MASO | | | -6.4, -4.4/ -4.8, -5.2 | | P.-Y. Lee |
| | 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 |
| <i>fymo</i> | DnN MOE ⁷ | | -4.2/-3.0/ -2.5 | | | C. Calestani |
| | Gcm MASO | | | -4.2 | | A. Ransick |
| | Elk-En | | | -6.0, -5.2, -5.7 | | M. Arnone |
| | Otx-En ¹⁸ | | NS | -2.1,-/3.3 | | A. Ransick & T. Minokawa |
| | GataE MASO ¹³ | | NS/NS/N S/ NS, NS | -2.2, -2.5/ NS, NS/- 2.4, -2.5 | | P.-Y. Lee |

| Gene | Perturbation | 12-15 | 18-21 | 24-27 | 30-36 | Author |
|------------------|--------------|--------------------------|------------------------------------|-------------------------------------|-------|--------------------------|
| <i>gatae</i> | 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 |
| <i>gcm</i> | 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 |
| <i>gell</i> | Bra MASO | | | -4.4 | | J. Rast |
| <i>hmx</i> | 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 |
| <i>hnf6</i> | Pmar1 MOE | -4.4/NS | | -7.6/-3.5 | | P. Oliveri |
| | Pmar1-En | -4.5/NS | | -8.0, NS, -2.8 | | P. Oliveri |
| <i>hox11/13b</i> | 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 |
| <i>kakapo</i> | Bra MASO | | | -2.0 | | J. Rast |
| <i>krox1</i> | 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 |
| | Otx-En | -2.3/-1.6 | NS/NS/NS | -2.5/ 3.3/NS | | A. Ransick & T. Minokawa |

| Gene | Perturbation | 12-15 | 18-21 | 24-27 | 30-36 | Author |
|--------------------|-----------------------|-----------|---|--|-------|-----------------------------|
| <i>lim1</i> | 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 |
| <i>Krox-En</i> | | | -3.1 | | | C. Livi |
| <i>msp130-like</i> | Bra MASO ¹ | | | -4.3 | | J. Rast |
| | Pmar1 MOE | +3.2 | | +5.8 | | P. Oliveri |
| | Pmar1-En | +2.4 | | +4.9 | | P. Oliveri |
| <i>msp130</i> | Dri MASO | -3.6/-5.1 | | | | G. Amore |
| | Pmar1 MASO | +3.1 | | +6.4 | | P. Oliveri |
| | Pmar1-En | +1.6 | | +4.7 | | P. Oliveri |
| <i>not</i> | GataE MASO | | NS/-1.9, - 1.2/NS/N S/ 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 |
| <i>nrl1</i> | 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 |
| <i>orct</i> | Bra MASO | | | -6.7 | | J. Rast |
| <i>otxβ1/2</i> | 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 |
| | DnN MOE ³ | | -2.5/NS | | | C. Calestani |
| | Otx-En | -3.8/-3.7 | -3.2/-4.1/ -3.3 | -2.1/-2.3/ -3.2/-3.8 | | A. Ransick & T. Minokawa |

| Gene | Perturbation | 12-15 | 18-21 | 24-27 | 30-36 | Author |
|--------------------------------------|-----------------------------|---------------------------|--------------------------|----------------------------------|--------------------|-----------------------------|
| <i>pks</i> | GataE MASO | | -2.2/-3.2 | -3.4/-1.6, -1.1/-3.2, -3.6 | | P.-Y. Lee |
| | Gcm MASO | | | -5.2/-3.3 | | 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 |
| <i>pmar1</i> | 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 |
| <i>sm27</i> | Pmar1 MOE | NS | | +4.0 | | P. Oliveri |
| | Pmar1 En | NS | | +2.7 | | P. Oliveri |
| <i>sm30</i> | Dri MASO | | | | -4.5/-4.8/ -5.9 | G. Amore |
| <i>sm50</i> | Dri MASO | | | | -4.0/-3.3/ -1.3 | G. Amore |
| | Bra MASO ¹ | | | -2.2 | | J. Rast |
| | Pmar1 MOE ¹¹ | | | +4.4/+2.3 | | P. Oliveri |
| | Pmar1-En ¹¹ | | | +2.5/NS | | P. Oliveri |
| | <i>sutx</i> | DnN MOE | | | -2.7/-2.6 | |
| <i>tbr</i> | 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 |
| <i>ui</i> (via <i>endo16</i>) | 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 |
| <i>wnt8</i> | Krox-En | -4.2 (6 h), -2.0, -3.7 | -3.2 | NS | | C. Livi |
| | 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/N S | | A. Ransick & T. Minokawa |
| | GataE MASO ¹⁴ | | +1.5/+2.3/ +1.4 | NS,NS/N S, NS/NS, NS | | P.-Y. Lee |

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

(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.

¹Must be affected indirectly via intercellular signaling because Bra is not present in cells where these genes are active.

²Affected indirectly via cell signaling since at 30 h Dri expressed only in OE.

³Thought to be indirect input, via GataE.

⁴Probably via GataC.

⁵Possibly indirect, via Bra.

⁶Probably in OE where both Elk and Lim1 are expressed then; not relevant to endomesoderm network.

⁷Indirect via Gcm; gene continues to be expressed as does *gcm* after N signaling phase is over.

⁸Likely to be indirect: could be via E(S)→GataE→Bra as it is a small effect; or via *gcm*.

⁹Inputs are inferred from effects on *endo16*; see note 10 under *endo16*.

- ¹⁰There 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.
- ¹¹Probably via *ets*; based on ECRA (Kurokawa *et al.*, 1999; Oliveri *et al.*, 2002).
- ¹²Indirect, because rescued by *krox* MOE.
- ¹³Probably indirect via *elk*.
- ¹⁴Weak 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.
- ¹⁵Indirect affect since introduction of *pmar1* mRNA rescues delta expression in Cad MOE embryos.
- ¹⁶Probably indirect via *gcm*, since *gcm* is an early target of N signaling system, and *fymo* expression depends on *gcm* expression.
- ¹⁷Probably 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.
- ¹⁸Probably indirect via *elk* because if there were an Otx site in *gatac* effect would have been seen earlier than 24 h.