

- fragility). A key research problem is distinguishing among laws, protocols, and historical accidents.
36. The central dogma can be thought of as a protocol, with DNA, RNA, protein, RNA polymerase, ribosome, etc., as modules.
  37. A compelling case, but using very different terminology, is made in (18).
  38. Steady state here means simply that all variables in Fig. 2 ( $r, d, y, A, C$ , etc.) approach constants, which can be solved for algebraically.
  39. "iff" means "if and only if."
  40. An important use of positive feedback is to deliberately destabilize equilibria and amplify small differences to create switches and to break symmetries and homogeneities. This can create patterns that are then maintained using negative feedback. Positive feedback is also critical to autocatalysis in growth and metabolism.
  41.  $da/dt = a' = gu$  means that  $a$  (the output of  $A$ ) is a time integral of  $gu$ , where  $u$  is the input to  $A$ .
  42. Stability is easily shown using standard methods of linear systems. Steady-state values can be found (in a stable system) by setting all time derivatives to 0, yielding  $gk_2y = gk_2r$  or  $y = (k_2/k_1)r$ .
  43. Mechanisms often exist that allow controller parameters (e.g.,  $k_1$  and  $k_2$ ) to be much less uncertain than  $g$  and  $d$ . It is often even easier to make ratios such as  $k_2/k_1$  largely invariant to variations in underlying physical quantities affecting the individual  $k_1$  and  $k_2$ .
  44. If precise gain is required, then the ratio  $k_2/k_1$  must also be precise.
  45. The national power grid has integral control at the  $>3000$  power plants to regulate frequency and voltages of delivered power, oil refineries have  $>10,000$  such control loops, and Internet congestion control involves a form implemented as part of TCP (transport control protocol). See (12) for more details, proofs, and examples.
  46. T. M. Yi, Y. Huang, M. I. Simon, J. Doyle, *Proc. Natl. Acad. Sci. U.S.A.* **97**, 4649 (2000).
  47. A. Levchenko, P. A. Iglesias, *Biophys. J.* **82**, 50 (2002).
  48. H. El-Samad, J. P. Goff, M. Khammash, *J. Theor. Biol.* **214**, 17 (2002).
  49. This argument can be made rigorous and is a standard elementary result in control theory. It is a special case of the internal model principle.
  50. For sufficiently large  $g$ , the frequency domain peak and time domain transients become unacceptably large, although still stable.
  51. One interpretation is that negative feedback is always balanced by an equal and opposite positive feedback. Strictly speaking, with dynamics this is not well defined, and  $\log|S(\omega)|$  gives the correct generalization.
  52. Relatively rare circumstances can involve an inequality ( $\geq$ ). This is worse, but it means that Eq. 4 is an inequality constraint rather than a pure "conservation" law. See (12, 63).
  53. The robust yet fragile nature of highly optimized complex regulatory networks can be mistakenly attributed to various kinds of bifurcations and "order-disorder" transitions (e.g., phase transitions, critical phenomena, "edge-of-chaos," pattern formation, etc.). See (12, 24).
  54. The development of the Boeing 777 alone required a global software and computing infrastructure with roughly 10,000 workstations, terabytes of data, and a billion-dollar price tag.
  55. Systems Biology Workbench ([www.cds.caltech.edu/sbw](http://www.cds.caltech.edu/sbw)).
  56. D. T. Gillespie, *J. Chem. Phys.* **113**, 297 (2000).
  57. K. Zhou, J. Doyle, K. Glover, *Robust and Optimal Control* (Prentice-Hall, Upper Saddle River, NJ, 1996).
  58. E. Sontag, *Mathematical Control Theory: Deterministic Finite Dimensional Systems* (Springer, New York, 1998).
  59. The idea of "closed-loop" systems biology is a purely metaphorical application of the message of this article to biology research itself. Biologists already iterate (loop) between experiment and at least informal modeling and "simulation." This process might be enhanced by "closing the loop" with theory and infrastructure "modules," which would also benefit.
  60. One such example is Internet technology, which is rich in protocols and feedback and is beginning to have a rich theory. Even though it is poorly understood by nonexperts and has become a focus of many specious theories, details and enormous data sets are available, and it makes an attractive example to compare with biological networks. See (12).
  61. N. Wiener, *Cybernetics* (Wiley, New York, 1948).
  62. A. R. Mushegian, E. V. Koonin, *Proc. Natl. Acad. Sci. U.S.A.* **93**, 10268 (1996).
  63. H. W. Bode, *Network Analysis and Feedback Amplifier Design* (Krieger, Melbourne, FL, 1945).

## REVIEW

# A Genomic Regulatory Network for Development

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

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

The mechanism causing cats to beget cats and fish to beget fish is hardwired in the genomic DNA, because the species specificity of the body plan is the cardinal heritable property. But despite all the examples of how individual genes affect the developmental process, there is yet no case where the lines of causality can be mapped from the genomic sequence to a major process of bilaterian development. One reason for this is that most of the developmental systems that have been intensively studied produce adult body parts, such as the third instar *Drosophila* wing disc, or the vertebrate hindbrain during rhombomere specification, or the heart anlagen of flies and mice (*I*). These systems

present tough challenges because they go through successive stages of pattern formation in order to generate complex morphologies, and their development is initiated from states that are already complex. Furthermore, traditional molecular, genetic, and developmental biological approaches have focused on determining the functions of one or a few genes at a time, an approach that is not adequate for analysis of large regulatory control systems organized as networks. The heart of such networks consists of genes encoding transcription factors and the cis-regulatory elements that control the expression of those genes. Each of these cis-regulatory elements receives multiple inputs from other

genes in the network; these inputs are the transcription factors for which the element contains the specific target site sequences. The functional linkages of which the network is composed are those between the outputs of regulatory genes and the sets of genomic target sites to which their products bind. Therefore, these linkages can be tested and verified by cis-regulatory analysis. This means identifying the control elements and their key target sites, and experimentally determining their functional significance. The view taken here is that "understanding" why a given developmental process

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

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

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

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

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

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

The most closely examined example of a cis-regulatory information processing system is that which controls developmental expression of the *endo16* gene of the sea urchin embryo. *Endo16* encodes a large polyfunctional protein that is secreted into the lumen of the embryonic and larval midgut. *Endo16* is expressed in the early embryo in the progenitors of the endomesoderm, then throughout the gut, and finally only in the midgut (2–4), a not very elaborate spatial sequence. But its control system turns out to be an elegantly organized and complex information processing device that responds to both positive and negative inputs to set the boundaries of expression. Early and late expression phases are controlled by two different sub-

regions of the regulatory sequence, or modules, each several hundred base pairs long. Together these are serviced by nine different DNA sequence-specific transcription factors. The functional role(s) of each interaction were determined (5, 6), and a computational model was derived to describe how this system responds to its time-varying regulatory inputs and to mutations and combinations of its target sites. The functions that the *endo16* regulatory system performs are conditional on the inputs, and they include linear amplification of these inputs, but also many nonlinear operations such as an intermodule switch that transfers control from the early to the late module, detection of input thresholds, and various logic operations (5, 6). The model affords precise predictions of the responses of this cis-regulatory system under all conditions.

### Uses of a First-Stage Regulatory Network Model

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

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

### Endomesoderm Specification in the Sea Urchin Embryo

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

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

The endomesoderm of the sea urchin embryo forms from cell lineages at the south pole (the “vegetal” pole) of the early embryo (Fig. 1). The endomesodermal constituents of the embryo ultimately consist of the skeletogenic mesenchyme, which arises from the micromere lineage; several other mesodermal cell types; and the gut endoderm. Most of the gut endoderm and all but the skeletogenic mesodermal cell types derive from the progeny of a ring of eight sixth cleavage cells, called “veg<sub>2</sub>”; the remainder of the gut endoderm derives from their eight sister cells, “veg<sub>1</sub>,” which also give rise to some ectoderm. What happens in the specification of the lineages is now reasonably well understood as a result of a long series of experimental studies to which many different labs have contributed [see the compressed summary of major steps in Table 1, and see (10) and (19) for reviews]. The specification of the micromere lineages occurs as soon as these cells are formed at fourth cleavage, because if isolated then and cultured, their progeny will express skeletogenic functions just as they do in their natural situation (10). Their specification depends initially on localized maternal cues.

Specification of the veg<sub>2</sub> lineage in endomesodermal progenitor cells begins immediately as well. There are two inputs required: one a signal passed from the micromeres to the immediate ancestors of the veg<sub>2</sub> ring, at fourth to sixth cleavage (20, 21), and the other the nuclearization of  $\beta$ -catenin (that is, its accumulation in the nuclei of all prospective endomesodermal cells) (22).  $\beta$ -catenin is a cofactor of the Tcf transcription factor, and its initial nuclearization is autonomous rather

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

At seventh through ninth cleavage, the descendants of the micromeres, now located in the center of the disc of *veg<sub>2</sub>* cells (Fig. 1, 10-hour embryo), emit the ligand *Delta* (24, 25), which activates the Notch (N) signal transduction system in the adjacent *veg<sub>2</sub>* cells and is required to specify them as mesoderm [Fig. 1, 15-hour embryo (26–28)]. If we now imagine the specification map from the bottom rather than from the

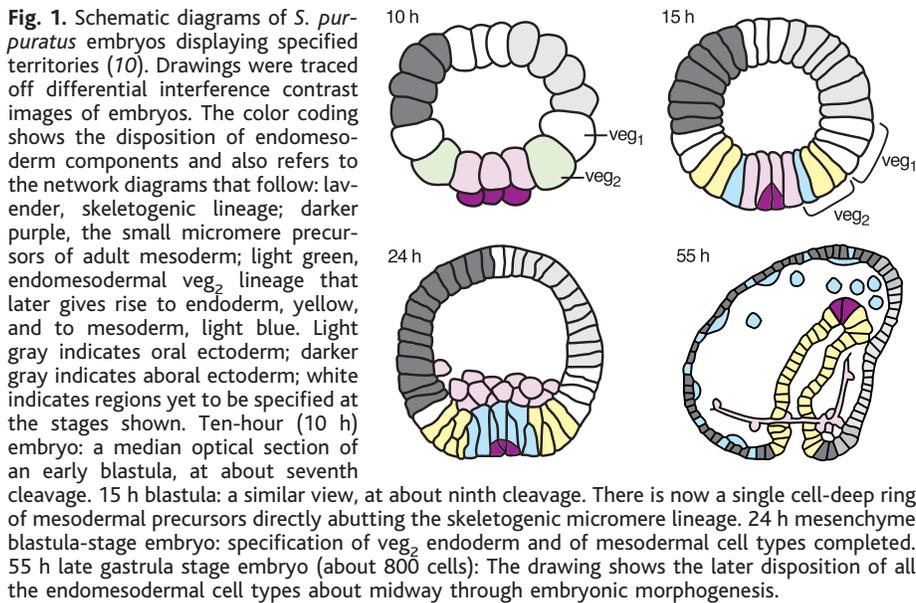
side as in Fig. 1, the pattern of cell fates (and by now of gene expression) would display a concentric arrangement (10): In the center are the “small micromeres,” the fifth-cleavage sister lineage of the skeletogenic micromeres; surrounding them are the skeletogenic precursors; the *veg<sub>2</sub>* mesoderm precursors; and finally the *veg<sub>2</sub>* endoderm precursors. The embryo is still an indifferent-looking hollow ball of cells, but the specification map is well on its way to completion. At 20 to 24 hours, the skeletogenic cells move inside the blastocoel (Fig. 1, 24-hour embryo), leaving behind a now fully specified central disc of prospective mesodermal cell types, and peripheral to them, the endoderm precursors. After this, a late *Wnt8* signal from the *veg<sub>2</sub>* endoderm causes the adjacent *veg<sub>1</sub>* progeny to become specified as endoderm as

well, and gastrular invagination ensues. The problem that we set ourselves was to discover the network of regulatory interactions underlying the events of endomesoderm specification during the first 24 hours, by which point some mesodermal and endodermal differentiation genes are already being expressed in a cell type-specific manner.

**Analyzing the Network**

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

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



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

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

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

In itself, perturbation analysis cannot distinguish between direct and indirect effects: Blockade of the expression of a gene that encodes a transcriptional activator may decrease expression of both immediately and secondarily downstream target genes; and if it encodes a repressor, blockade of its expression may increase expression of both. Direct effects are those in which a perturbation in the expression or function of a transcription factor causes changes in the expression of another gene, because target sites for that factor are included in a cis-regulatory element of the gene. cis-Regulatory analysis can therefore be used to resolve whether effects on a given control element are indeed direct. Another approach that we have used at sev-

eral key nodes of the network is the attempted rescue of a perturbation effect by introduction of appropriate amounts of mRNA encoding a different factor, which might be mediating an indirect effect of the perturbation (33). Where a rescue experiment indicates an indirect effect, or where the effect must be indirect because the affected and the perturbed genes are expressed in different cells or at different times, the implied relationships are omitted from the network models. This is because only direct effects imply specific genomic target site sequences in the cis-regulatory systems of the affected genes, and an object of the network model is to make explicit a testable map of cis-regulatory interrelations.

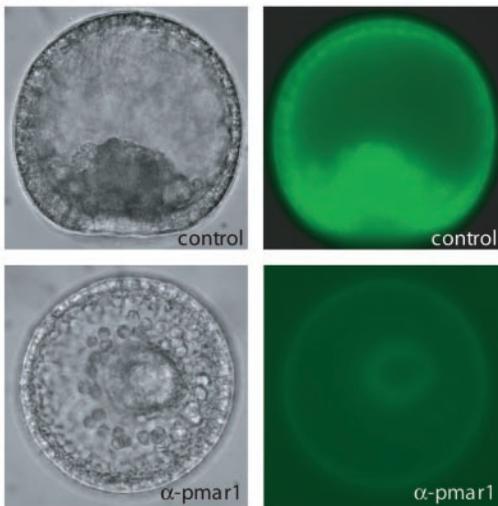
In an iterative process, the inferences from the experimental perturbation results were checked against the network model, further experiments were designed, the model was altered according to their results if necessary, and so

forth. The model was constructed with the program Netbuilder (34), a new tool for the construction of computational models that allows simulations to be performed, so as to test whether its relationships generate the appropriate outputs. But from the start, the model had to conform to the facts from experimental embryology (Table 1).

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

**A Perturbation:** Morpholino antiisense

**Effect:** Prevents translation of mRNA



**B Perturbation:** Engrailed repressor domain fusion

**Effect:** Converts transcription factor into dominant repressor for all target genes



**C Perturbation:** Cadherin mRNA injection

**Effect:** Blocks activation of Wnt/Tcf signaling pathway



**D Perturbation:** Negative Notch mRNA injection

**Effect:** Blocks Notch signaling pathway



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

embryo of the same age expressing an injected mRNA that encodes the intracellular domain of cadherin is shown on the right (image from A. Ransick). The cadherin embryo consists of a hollow ball of ectoderm; endomesodermal specification has been completely wiped out. **(D)** Effect of the introduction of a negatively acting derivative of the N receptor. A control 37-hour late gastrula is shown on the left, and on the right is an embryo of the same age expressing an injected mRNA encoding the extracellular domain of the N receptor (negN) (image from C. Caletani). This embryo has a normal complement of skeletogenic mesenchyme cells and a well-formed gut but only a very few mesodermal cells of  $veg_2$  origin as compared with the control.

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

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

FamilyRelations, was built for the purpose of recognizing short patches of conserved sequence in long stretches of genomic DNA (34). Applied to the *Strongylocentrotus-Lytechinus* species pair, this approach efficiently served to identify cis-regulatory elements that score positively in gene transfer tests (42).

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

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

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

The early cleavage stage events in endomesoderm specification take place in the  $veg_2$  endomesoderm lineage, indicated in light green above the triple line at the top, and in the micromere lineage shown in lavender at

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

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

**Table 2.** Differential gene discovery screens. Macroarray filter screens were carried out with probes prepared by high- $C_0t$  subtractive hybridization, using single-stranded driver and selectate, as described (35). "Selectate" denotes the cDNA preparation that contains the sequences of interest, in contrast to the nucleic acid present in excess in the hybridization reaction: The "Driver," which lacks these sequences. In the subtractive hybridizations, the reactions were carried out to near termination with respect to driver, and nonhybridized selectate sequences were recovered by hydroxyapatite chromatography (35).

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

\*Cad, intracellular domain of cadherin (Fig. 2C). This domain sequesters  $\beta$ -catenin, which is thereby localized at the inner surface of the cell membrane. An excess of the cadherin intracellular domain severely decreases the availability of  $\beta$ -catenin for transit into the nucleus. †LiCl-treated embryos produce excess endomesoderm (12, 62).

‡The extracellular domain of N acts as a repressor of N function in mesoderm specification (27) (Fig. 2D).

§The *brachyury* (*bra*) gene is active by about 18 hours. Driver mRNA was extracted from normal 15-hour embryos.

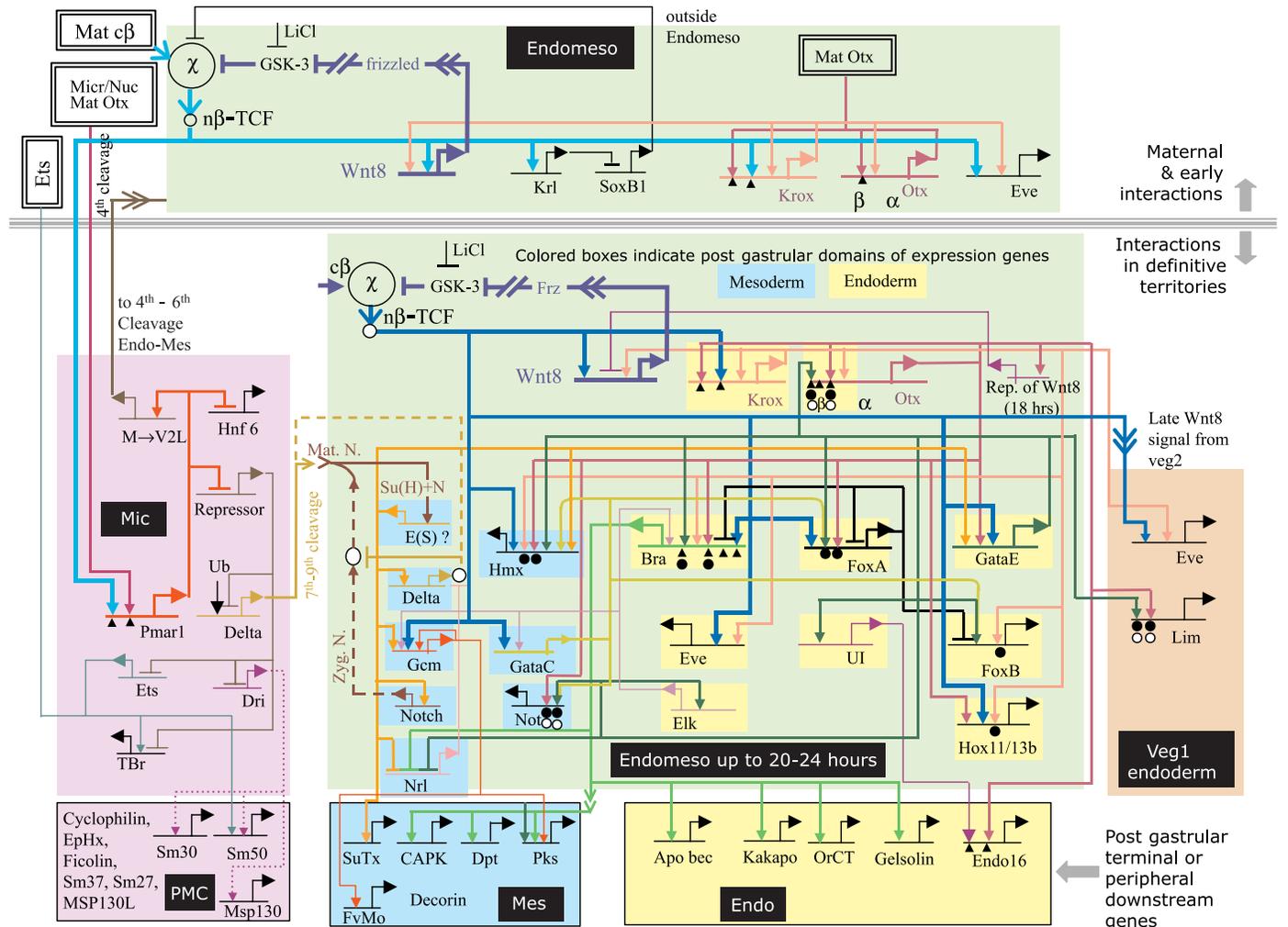
||Ectopic *bra*-expressing cells were obtained by disaggregating 18-hour embryos expressing genetic constructs that produce *bra* mRNA under the control of a ubiquitously active cis-regulatory element. The transgenic cells were tagged with green fluorescent protein (GFP) and isolated by fluorescence-activated cell sorting (FACS) (37).

¶MASO embryos were collected at 24 to 27 hours (late blastula stage). #Cells expressing *bra* were obtained by FACS as above, but at 24 to 27 hours (37).

The network explains some of the phenotypes observed when given processes are perturbed, in terms of its consequential regulatory logic. For example, as shown in Fig. 2C, if  $\beta$ -catenin nuclearization is prevented by intro-

duction of mRNA encoding the intracellular domain of cadherin, neither endodermal nor mesodermal cell types and structures appear. In default of  $\beta$ -catenin/Tcf inputs, the embryo becomes a hollow ball of ectoderm. Note, howev-

er, that all the perturbation data underlying the network in Fig. 3 were obtained between 6 and 24 hours, long before any gastrulation phenotypes can be seen (30). Initiation of  $\beta$ -catenin nuclearization produces such a catastrophic re-



**Fig. 3.** Regulatory gene network for endomesoderm specification: the view from the genome. The current version of the model in this figure and the perturbation data on which it is based are available on a Web site (30); for additional details and discussion, see (19). At the top, above the triple line, are the earliest interactions; in the middle tier, the spatial domains are color-coded (Fig. 1), and genes are placed therein according to their final loci of expression. As indicated (black background labels), the lavender area at the left represents the skeletogenic micromere (mic) domain before ingress; the light green area indicates the veg<sub>2</sub> endomesoderm domain, with genes eventually expressed in endoderm on yellow backgrounds and genes eventually expressed in mesoderm on blue backgrounds; the tan box at right represents the veg<sub>1</sub> endoderm domain. Many genes are initially expressed over broader ranges, and their expression later resolves to the definitive domains. The rectangles in the lower tier of the diagram show downstream differentiation genes (PMC, "primary" or skeletogenic mesenchyme). Short horizontal lines from which bent arrows extend represent cis-regulatory elements responsible for expression of the genes named beneath the line. Embryonic gene expression was perturbed in specific ways as in Fig. 2. The arrows and barred lines indicate the inferred normal function of the input (activation or repression), as deduced from changes in transcript levels due to the perturbations. Each input arrow constitutes a prediction of specific transcription factor target site sequence(s) in the cis-regulatory control element. In some cases, the predicted target sites have been identified in experimentally defined cis-regulatory elements that generate the correct

spatial pattern of expression (solid triangles). At the upper left, the light blue arrow represents the maternal  $\beta$ -catenin ( $c\beta$ ) nuclearization system ( $\chi$ ). 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 (50, 51). Data for the role of Ets are from (52, 65). "Micr/Nuc Mat Otx" refers to the early localization of maternal Otx in micromere nuclei at fourth cleavage (56). Genes labeled "Repressor" are inferred; all other genes shown are being studied at the DNA sequence level and by multiplexed QPCR. "Ub" indicates a ubiquitously active positive input inferred on the basis of ubiquitous expression seen by whole-mount in situ hybridization, under conditions in which a spatial repression system that normally confines expression has been disarmed. Dotted lines in the diagram indicate inferred but indirect relationships. Arrows inserted in arrow tails indicate intercellular signaling interactions. Small open or closed circles indicate perturbation effects that resist rescue by the introduction of mRNA where there is a possibility that the effect seen is actually an indirect result of an upstream interaction; that is, this possibility of such an indirect effect has been experimentally excluded, and both sites are shown as probable direct inputs (19). Large open ovals represent cytoplasmic biochemical interactions at the protein level, such as those responsible for nuclearization of  $\beta$ -catenin, for the effect of Delta on N (66); or for the effect of Neuralized, an E3 ubiquitin ligase with specificity for Delta (67, 68).

sult because multiple endodermal and mesoderm regulatory genes depend on a  $\beta$ -catenin/Tcf input. For these genes, only a few percent of control transcript levels survive cadherin mRNA injection (19, 30). Another interesting phenotype is obtained when embryos are treated with  $\alpha$ -gcm MASO. The result is albino larvae (49). The gene *gcm* is ultimately expressed in pigment cells (36), and a downstream target of *gcm* is the *pks* (polyketide synthase) gene, which is also expressed in pigment cells (38, 39). This product (and other pigment cell genes under *gcm* control, not shown) is likely to be required for synthesis of the red quinone pigment these cells produce. Upstream, the network shows *gcm* to be a target of the N signaling system, because its expression is severely depressed by the introduction of a negatively acting N derivative (19) (Fig. 2D). In fact, *gcm* expression begins in the single ring of mesoderm progenitor cells that directly receives the Delta micromere signal (36). So we now have a sequence of DNA-based interactions that leads from the initial specification to the terminal differentiation of pigment cells and that explains the albino phenotype. Similarly, the network explains the

$\alpha$ -*gatae* MASO phenotype. This treatment produces a severe interference with endoderm specification and gut development (43), which is no less than would be expected from the branching regulatory effects of *gatae* expression indicated in the network.

The network explains the role of the signaling interactions required in endomesodermal specification in terms of their inputs into cis-regulatory systems (except for the early micromere-to-veg<sub>2</sub> signal, the targets of which remain unknown). The gene encoding Wnt8 is itself a target of a  $\beta$ -catenin/Tcf input and it is, in addition, under the control of the early endomesoderm regulator *krox*. These inputs show how the autonomous nuclearization of  $\beta$ -catenin soon causes the Wnt8 loop to start up in all endomesoderm cells, strengthening the set of regulatory relationships indicated by the blue lines in Fig. 3.

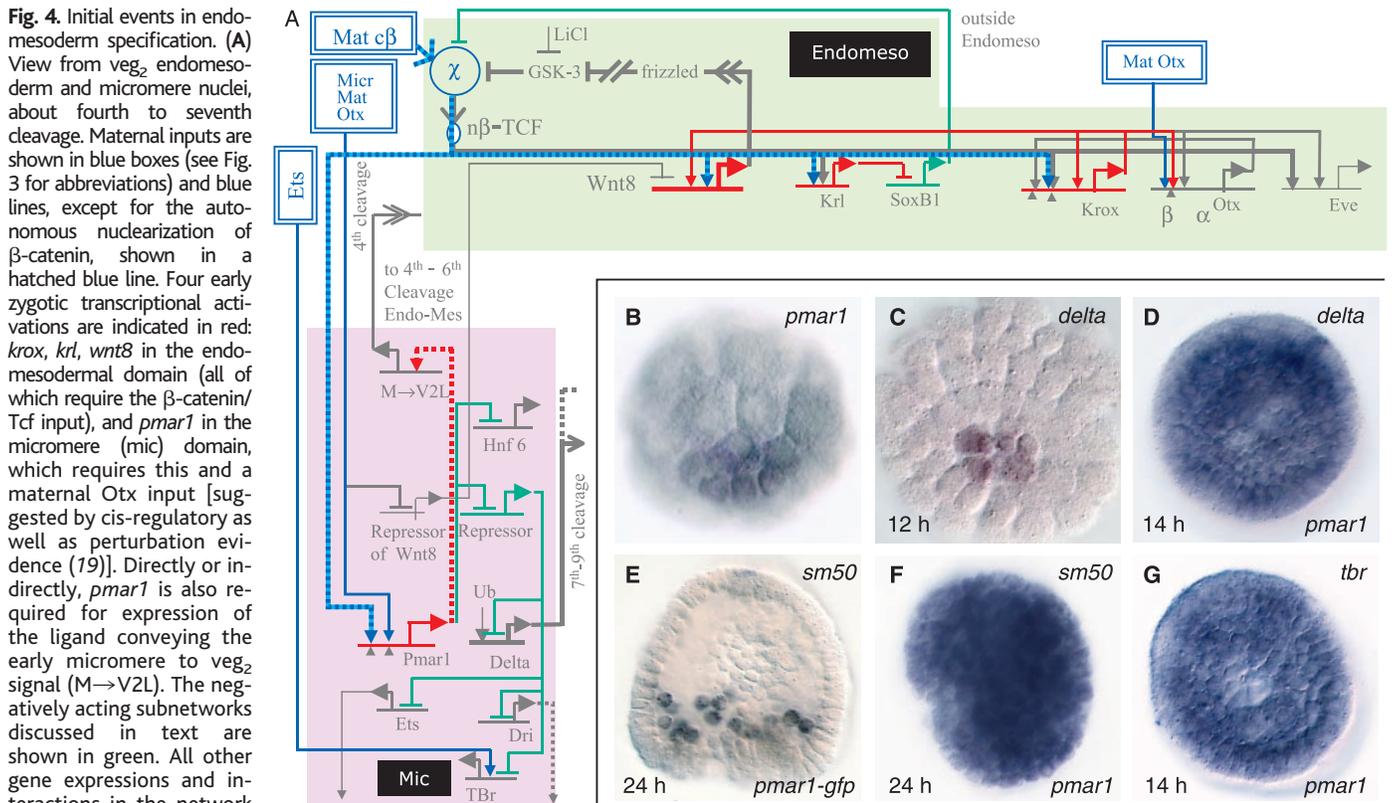
The view from the genome provides a qualitative DNA-level explanation for the spatial domains of expression of many endomesodermal regulatory genes. No two of these genes have identical inputs: Each cis-regulatory information processing system has its own job to do.

The network shows that the downstream targets of a few of these regulatory genes, such as *bra* (37), include differentiation proteins that were discovered in our differential screens, but for many of the regulatory genes the downstream targets are still unknown.

**System-Level Insights into the Developmental Process**

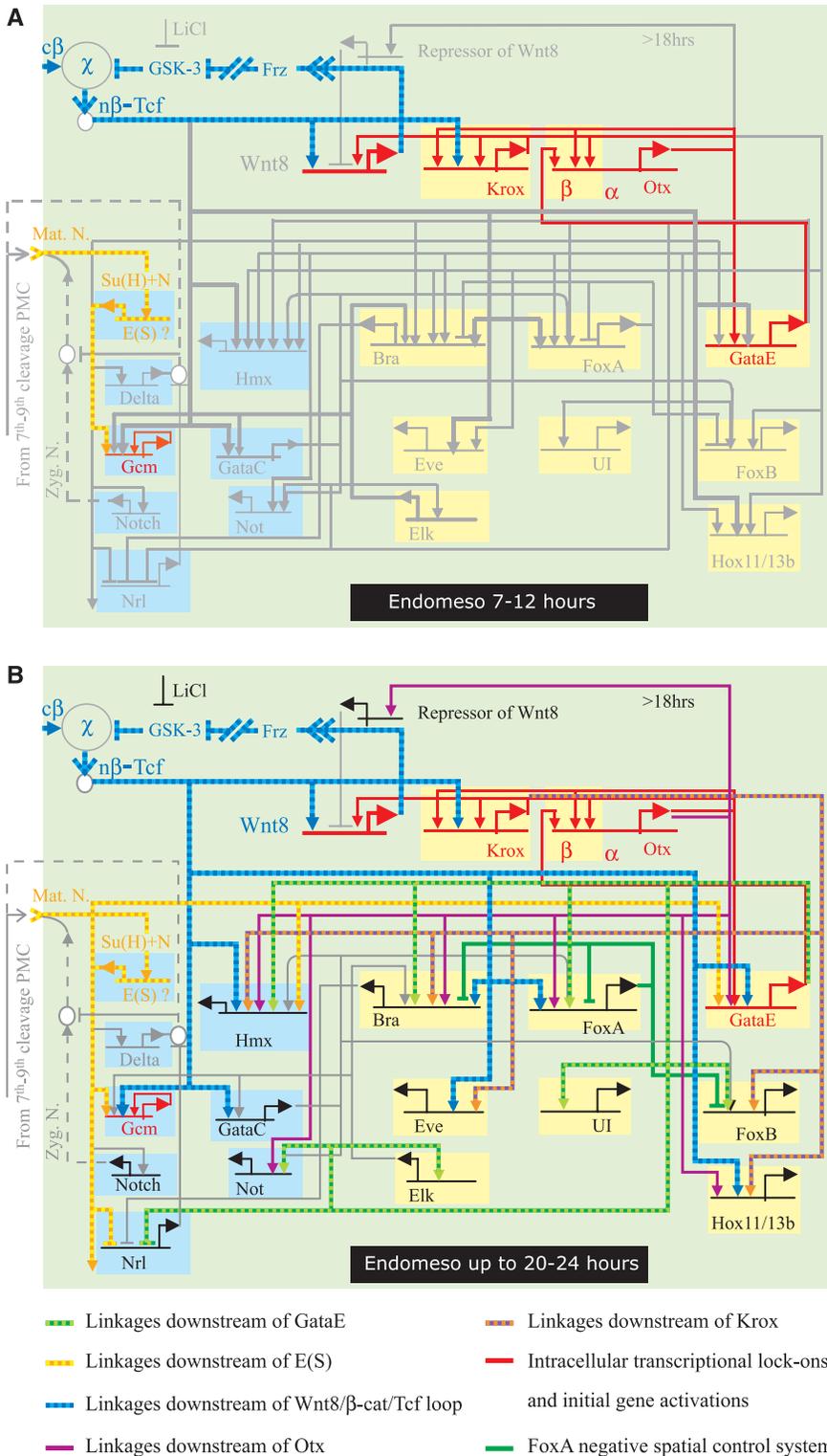
Physiological transcriptional responses flicker on after the advent of stimuli, then return to their ground state; for example, after changes in the level of nutrients or the advent of toxins in the bloodstream, or after the appearance of pathogens. In contrast, the fundamental feature of developmental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

The initial events in endomesoderm specification occur in the micromeres and in the veg<sub>2</sub>



**Fig. 4.** Initial events in endomesoderm specification. (A) View from veg<sub>2</sub> endomesoderm and micromere nuclei, about fourth to seventh cleavage. Maternal inputs are shown in blue boxes (see Fig. 3 for abbreviations) and blue lines, except for the autonomous nuclearization of  $\beta$ -catenin, shown in a hatched blue line. Four early zygotic transcriptional activations are indicated in red: *krox*, *krl*, *wnt8* in the endomesodermal domain (all of which require the  $\beta$ -catenin/Tcf input), and *pmar1* in the micromere (mic) domain, which requires this and a maternal Otx input [suggested by cis-regulatory as well as perturbation evidence (19)]. Directly or indirectly, *pmar1* is also required for expression of the ligand conveying the early micromere to veg<sub>2</sub> signal (M  $\rightarrow$  V2L). The negatively acting subnetworks discussed in text are shown in green. All other gene expressions and interactions in the network are indicated in gray. (B through G) Whole-mount in situ hybridization displays, from (25). The gene, expression of which is being displayed, is shown at the upper right, and the mRNA injected into the egg at the lower right; the age of the embryo is at lower left. (B) Expression of *pmar1* specifically in micromeres. (C) Expression of *delta* specifically in micromeres. (D) Expression of *delta* in all embryonic cells when *pmar1* mRNA is translated everywhere, after injection into the egg. Exactly the same result is obtained if an Engrailed domain fusion is instead expressed (25); because the Engrailed fusion acts as an obligate repressor of *pmar1* target genes, *pmar1*

must normally act as a repressor. (E) Expression of *sm50*, a skeletogenic differentiation gene exclusively in skeletogenic mesenchyme cells (69). (F) Global expression of *sm50* in embryos expressing *pmar1* globally. (G) Expression of the skeletogenic regulator *tbr* in embryos expressing *pmar1* mRNA globally. (F) and (G) show that the whole embryo has been converted to a state of skeletogenic mesenchyme differentiation. Note the rounded form of the cells at 24 hours in (F), as compared to the control in (E), due to their tendency to behave mesenchymally.



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

lineage, as summarized above. The maternal inputs provide the initial state, with respect to regulatory transactions. There are two consequences of the initial zygotic transcriptional responses (Fig. 4A, shown in red). The first is to begin the activation of the endomesodermal zygotic control apparatus; here, by turning on the *krox* (35) and *krl* [krüppel-like (50)] genes in the *veg*<sub>2</sub> endomesoderm and the *pmar1* gene in the micromeres. The second is a surprise: An immediate sequel, in both domains, is to engage repressive subnetworks (shown in green) of interactions that have the effect of stabilizing the initial definition of the endomesodermal and mesomere territories by cutting off the possibility of similar transcriptional activations elsewhere. The *krl* gene encodes a repressor that prevents expression of *soxB1* in the endomesoderm, though it is expressed everywhere else (50, 51). The SoxB1 protein antagonizes nuclearization of  $\beta$ -catenin. The *krl/soxB1* loop is an early lock-down device to keep the endomesodermal cells endomesodermal (because they have elevated nuclear  $\beta$ -catenin from the start) and to prevent other cells from going the same way. The *pmar1* gene active in the micromeres also encodes a repressor. Its target is an unknown gene that produces another repressor of key regulators of micromere-specific function. Like *soxB1*, it too is potentially active everywhere, except where it itself is repressed, which is the role accomplished by *pmar1* in the micromeres. Micromere regulators that are micromere-specific only because of the *pmar1* repression system include the gene that produces the Delta signal to the surrounding *veg*<sub>2</sub> cells and the regulatory genes that are responsible for installing the skeletogenic state of differentiation in the micromere progeny [the *t-brain* (*tbr*) gene, the *ets* gene, and the *deadringer* (*dri*) gene (19, 25, 52)]. Some evidence for the *pmar1* repression system is reproduced in Fig. 4, B through G. Expression of the *delta* gene, the *tbr* skeletogenic control gene, and *sm50*, a skeletogenic differentiation gene, all occur globally if *pmar1* mRNA is expressed globally (25) (Fig. 4). Almost the first thing accomplished by zygotic genes activated in both the *veg*<sub>2</sub> endomesoderm and the micromeres is to activate local negative control of otherwise global repressors of the respective states of specification. The network reveals active repression of these endomesodermal regulatory states in all the cells of the embryo, except those where *krl* and *pmar1* are respectively activated.

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

tions. In the future mesodermal domain, the *gcm* gene autoregulates after its initial activation through the N pathway (49). Similarly, the *krox* gene positively autoregulates, in addition to stimulating expression of the *wnt8* gene, which locks *wnt8* and *krox* in a positive regulatory embrace. The *krox* gene product also activates one of the transcription units of the *otx* gene (19, 30, 42). In turn, Otx stimulates the *krox* gene. The *otx* gene now provides an input into the *gatae* gene, the importance of which was discussed above; but note that the  $\beta$ -*otx* cis-regulatory system in turn responds positively to GataE input (30, 43). This is a further positive feedback that links the *gatae* gene, a dedicated endomesodermal activator, into the stabilization circuitry. As illustrated by the color coding in Fig. 5B, the regulatory state illustrated in Fig. 5A suffices to provide inputs to every one of the known transcriptional regulatory genes in the endomesodermal domain. The drivers are Krox, Otx, GataE, Tcf, and whatever Enhancer of Split-like factor operates in this embryo downstream of N signal transduction. After this, the expression of the *wnt8* gene falls off [probably the gene is repressed by one of the Otx isoforms (19, 30, 42, 53)]; and during the late blastula stage,  $\beta$ -catenin disappears from the  $\text{veg}_2$  endomesoderm nuclei (22). By now, the regulatory system is locked in and has no further need of this input, which was so important in the initial phases of the specification process.

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

## Conclusions

Developmental regulatory network analysis can be done in any organism where the necessary genomics, a high-throughput method of gene transfer, and the ancillary molecular methods are available. But it requires a new mix of technologies and a new level of close interactions between system-minded biologists and computational scientists. It seems no more possible to understand development from an informational point of view without unraveling the underlying regulatory networks than to understand where protein sequence comes from without knowing about the triplet code. To understand the operation of whole systems of regulatory interactions, computational models are essential: for orga-

nizing experimental extensions and tests at each stage of construction of the model, to check on consistency, and to integrate experimental results with the current network architecture by means of simulation. The cis-regulatory systems at the nodes of the network in reality each process kinetic input information: the rise and fall of the activities of the transcription factors to which they respond. But even from the first-stage model, which just states the interactions that occur at each node, there emerge system properties that can only be perceived at the network level. Examples are the features of the system treated in Figs. 4 and 5. These features explain the means by which maternal spatial cues are used to activate the zygotic transcriptional network, the progressivity of the developmental process, and its lock-down mechanisms. The network model relates these and other developmental features of the process of endomesoderm specification (19) directly to the genome, because it is couched in terms of cis-regulatory interactions at the DNA level. The model thus represents an outline of the heritable developmental program, but the program is not the machine. The DNA regulatory network coexists with many other multicomponent systems that constitute the machine. These systems execute biochemical functions, produce signal transduction pathways, and cause cell biological changes to occur. They sum to the majority of the working parts of the cell. Their mobilization is controlled by the transcriptional switches that hook them into the genomic regulatory control system.

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

what other forms share them, which are new, and which are ancient.

## References and Notes

1. E. H. Davidson, *Genomic Regulatory Systems. Development and Evolution* (Academic Press, San Diego, CA, 2001).
2. C. Nocente McGrath, C. A. Brenner, S. G. Ernst, *Dev. Biol.* **136**, 264 (1989).
3. A. Ransick, S. Ernst, R. J. Britten, E. H. Davidson, *Mech. Dev.* **42**, 117 (1993).
4. M. Soltysik-Espanola, D. C. Klinzing, K. Pfarr, R. D. Burke, S. G. Ernst, *Dev. Biol.* **165**, 73 (1994).
5. C.-H. Yuh, H. Bolouri, E. H. Davidson, *Science* **279**, 1896 (1998).
6. ———, *Development* **128**, 617 (2001).
7. M. Arnone, E. H. Davidson, *Development* **124**, 1851 (1997).
8. H. Bolouri, E. H. Davidson, *Dev. Biol.*, in press.
9. E. H. Davidson, *Development* **113**, 1 (1991).
10. E. H. Davidson, R. A. Cameron, A. Ransick, *Development* **125**, 3269 (1998).
11. L. M. Angerer, R. C. Angerer, *Dev. Biol.* **218**, 1 (2000).
12. S. Hörstadius, *Biol. Rev. Camb. Philos. Soc.* **14**, 132 (1939).
13. R. A. Cameron et al., *Proc. Natl. Acad. Sci. U.S.A.* **97**, 9514 (2000).
14. X. Zhu, G. Mahairas, R. A. Cameron, E. H. Davidson, C. A. Etensohn, *Development* **128**, 2615 (2001).
15. A. Poustka et al., *Genomics* **59**, 122 (1999).
16. C. V. Kirchhamer, E. H. Davidson, *Development* **122**, 333 (1996).
17. A. P. McMahon et al., *Dev. Biol.* **108**, 420 (1985).
18. C. N. Flytzanis, R. J. Britten, E. H. Davidson, *Proc. Natl. Acad. Sci. U.S.A.* **84**, 151 (1987).
19. E. H. Davidson et al., *Dev. Biol.*, in press.
20. A. Ransick, E. H. Davidson, *Science* **259**, 1134 (1993).
21. ———, *Development* **121**, 3215 (1995).
22. C. Y. Logan, J. R. Miller, M. J. Ferkowicz, D. R. McClay, *Development* **126**, 345 (1999).
23. A. Wikramanayake et al., in preparation.
24. H. C. Sweet, M. Gehring, C. A. Etensohn, *Development*, in press.
25. P. Oliveri, D. M. Carrick, E. H. Davidson, *Dev. Biol.*, in press.
26. D. R. Sherwood, D. R. McClay, *Development* **124**, 3363 (1997).
27. ———, *Development* **126**, 1703 (1999).
28. ———, *Development* **128**, 2221 (2001).
29. For perturbation measurements by QPCR, RNA was extracted from embryos grown from eggs injected with the respective perturbation reagents and converted to cDNA. This was used in multiplexed QPCR experiments to assess the quantitative effects of each perturbation on the other genes indicated in Fig. 3, usually at several different stages. The effects for data included in the model ranged from threefold increases or decreases in the amount of transcript as a result of the perturbation up to several hundred fold increases or decreases. Results were obtained on at least two different independent batches of cDNA for each perturbation, sometimes more, and often several repeats were performed on each batch. For numerical data, see the URL in (26).
30. For the Gene Network Update Web site, see [www.its.caltech.edu/~mirsky/endomes.htm](http://www.its.caltech.edu/~mirsky/endomes.htm); for the QPCR data underlying the network, see [www.its.caltech.edu/~mirsky/qpcr.htm](http://www.its.caltech.edu/~mirsky/qpcr.htm).
31. X. Li, A. H. Wikramanayake, W. H. Klein, *Dev. Biol.* **212**, 425 (1999).
32. The introduction of mRNA encoding sequence-specific transcription factors at levels grossly beyond those naturally present per cell of course breeds artifacts, by engendering interactions with weak target sites. The only mRNA encoding a natural transcription factor used in this work was *pmar1* mRNA, which produces a repressor. The level injected into the egg was only two to three times the normal level per cell (discounting mRNA decay), although after injection it was globally distributed. The identical phenotypes (shown in Fig. 4) were obtained after injection of mRNA encoding a Pmar1-Engrailed domain fusion. This result justifies both experiments (25). Three other mRNAs encoding Engrailed fusions were used. The specificity of the first, Otx-Engrailed, was thoroughly studied earlier (37). The

- specificity of the effects of the second, encoding *Krox-Engrailed*, is attested to by the minor fraction of tested genes on which it had any effect whatsoever (19) and the fact that the only genes that were affected are required just where *krox* is expressed, even though the exogenous mRNA is present globally. The last, encoding an Elk-Engrailed construct, affects only three other genes out of all those tested (19) (Fig. 3). Elk plays a peripheral role in the network up to 24 hours, and its main importance may be for later events in development.
33. In such rescue experiments, if the effect is indirect via a second gene, then the introduction of mRNA generated from the second gene will suffice to correct the perturbation effect; but if it is direct, no rescue can be obtained by this route. For example, if gene A activates gene B, which in turn activates gene C, then the effect of a knockout of gene A expression is direct for gene B but indirect for gene C, and the effect of the gene A knockout on gene C would be rescued by the introduction of B mRNA. If, on the other hand, there are necessary target sites for the gene A product in the cis-regulatory elements of both genes B and C, then the effect on gene C of a gene A knockout cannot be rescued by the introduction of B mRNA.
  34. C. T. Brown *et al.*, *Dev. Biol.*, in press.
  35. J. P. Rast *et al.*, *Dev. Biol.* **228**, 270 (2000).
  36. A. Ransick, J. P. Rast, T. Minokawa, C. Calestani, E. H. Davidson, *Dev. Biol.*, in press.
  37. J. P. Rast, R. A. Cameron, A. J. Poustka, E. H. Davidson, in preparation.
  38. See *Science* Online for supplementary Web material ([www.sciencemag.org/cgi/content/full/295/5560/1669/DC1](http://www.sciencemag.org/cgi/content/full/295/5560/1669/DC1)).
  39. C. Calestani, E. Davidson, unpublished data. All QPCR measurements cited are posted on the QPCR Web site (30). The phenotype of the negatively acting N embryo is shown in Fig. 2D. The expression of the *pks* gene is posted on the *Science* Web site (38).
  40. P. Gonzalez, H. A. Lessios, *Mol. Biol. Evol.* **16**, 938 (1999).
  41. C. R. C. Paul, A. B. Smith, *Biol. Rev.* **59**, 443 (1984).
  42. C.-H. Yuh *et al.*, *Dev. Biol.*, in press.
  43. P.Y. Lee, E. Davidson, unpublished data. All QPCR data cited are posted on the QPCR Web site (30). An image of the  $\alpha$ -GataE MASO phenotype is posted on the *Science* Web site (38).
  44. N. H. Luke, C. E. Killian, B. T. Livingston, *Dev. Growth Differ.* **39**, 285 (1997).
  45. D. R. McClay, D. R. Gross, *Dev. Biol.* **239**, 132 (2001).
  46. R. Cameron, E. Davidson, unpublished data. The expanded expression of *bra* in sea urchin embryos bearing  $\alpha$ -*foxa* MASO is illustrated on the *Science* Web site (38).
  47. V. Hinman, E. Davidson, unpublished data. The expanded expression of *bra* in a starfish embryo bearing  $\alpha$ -*foxa* MASO is illustrated on the *Science* Web site (38).
  48. P. Oliveri, E. Davidson, unpublished data. All *foxa* QPCR data cited are posted on the QPCR Web site (30). Other data are cited in (19).
  49. A. Ransick, E. Davidson, unpublished data. An image of the albino phenotype produced by  $\alpha$ -*gcm* MASO is posted on the *Science* Web site (38). Evidence for *gcm* autoregulation comes from QPCR data posted on the web site indicated in (30).
  50. E. W. Howard, L. A. Newman, D. W. Oleksyn, R. C. Angerer, L. M. Angerer, *Development* **128**, 365 (2001).
  51. A. P. Kenny, D. Kozlowski, D. W. Oleksyn, L. M. Angerer, R. C. Angerer, *Development* **126**, 5473 (1999).
  52. K. Akasaki, personal communication.
  53. X. Li, C. K. Chuang, C. A. Mao, L. M. Angerer, W. H. Klein, *Dev. Biol.* **187**, 253 (1997).
  54. R. J. Britten, E. H. Davidson, *Quart. Rev. Biol.* **46**, 111 (1971).
  55. This description relates specifically to indirectly developing euechinoid species, such as *S. purpuratus*, *L. variegatus*, *Hemicentrotus pulcherrimus*, and *Paracentrotus lividus*.
  56. C. Chuang, A. H. Wikramanayake, C. Mao, X. Li, W. Klein, *Dev. Genet.* **19**, 231 (1996).
  57. H. C. Sweet, P. G. Hodor, C. A. Ettensohn, *Development* **126**, 5255 (1999).
  58. S. W. Ruffins, C. A. Ettensohn, *Dev. Biol.* **160**, 285 (1993).
  59. ———, *Development* **122**, 253 (1996).
  60. A. Ransick, E. H. Davidson, *Dev. Biol.* **195**, 38 (1998).
  61. C. Y. Logan, D. R. McClay, *Development* **124**, 2213 (1997).
  62. R. A. Cameron, E. H. Davidson, *Dev. Biol.* **187**, 236 (1997).
  63. W. Wang *et al.*, *Mech. Dev.* **60**, 185 (1996).
  64. C. Livi, E. Davidson, unpublished data. QPCR data cited are posted on the QPCR Web site (30); the *Krox-Engrailed* phenotype is illustrated in Fig. 2B.
  65. D. Kurokawa *et al.*, *Mech. Dev.* **80**, 41 (1999).
  66. T. L. Jacobsen, K. Brennan, A. Martinez-Arias, M. A. T. Muskavitch, *Development* **125**, 4531 (1998).
  67. E. Yeh, L. Zhou, N. Rudzik, G. L. Boulianne, *EMBO J.* **19**, 4827 (2000).
  68. E. Yeh *et al.*, *Curr. Biol.* **11**, 1675 (2001).
  69. H. M. Sucof, B. R. Hough-Evans, R. J. Britten, E. H. Davidson, *Genes Dev.* **2**, 1238 (1988).
  70. We are particularly grateful to E. Branscombe, who arranged for the U.S. Department of Energy's Joint Genome Institute to provide many of the BAC sequences included in this work (19); the remainder were obtained at the Institute for Systems Biology, Seattle, WA. We thank E. Rothenberg of Caltech for perspicacious and very helpful comments on the manuscript. The major embryological aspects of this work were supported by grants from NIH (HD-37105 and RR-06591); the computational aspects were supported by a grant from NIH (GM-61005); the comparative aspects were supported by a grant from the NASA/Ames Fundamental Space Biology program (NAG2-1368); the arrayed cDNA and BAC libraries on which the project depended were produced with the support of a grant from NIH (RR-15044); and the work on the *brachyury* gene was supported by a Human Frontiers grant (RG0290) to R.A.C. Other support was provided by the Lucille P. Markey Charitable Trust, the Stowers Institute for Medical Research, the Beckman Institute of Caltech, and a grant from NSF (IBN99882875) to R.A.C.

## REVIEW

# Modeling the Heart—From Genes to Cells to the Whole Organ

Denis Noble

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

The amount of biological data generated over the past decade by new technologies has completely overwhelmed our ability to understand it. Genomics has provided us with a massive “parts catalog” for the human body; proteomics seeks to define these individual

“parts” and the structures they form in detail. But there is as yet no “user’s guide” describing how these parts are put together to allow those interactions that sustain life or cause disease. In many cases, the cellular, organ, and system functions of genes and proteins

are unknown, although clues often come from similarity in the gene sequences. Moreover, even when we understand function at the protein level, successful intervention, for example, in drug therapy, depends on knowing how a protein behaves in context, as it interacts with the rest of the relevant cellular machinery to generate function at a higher level. Without this integrative knowledge, we may not even know in which disease states a receptor, enzyme, or transporter is relevant, and we will certainly encounter side effects that are unpredictable from molecular information alone.

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

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