Gene regulatory networks and embryonic specification

Leroy Hood*

Institute for Systems Biology, 1441 North 34th Street, Seattle, WA 98103

Eric Davidson embarked on the study of sea urchin development more than 40 years ago, with his initial efforts focused on studying the animal's nucleic acid properties through hybridization—studies that, amazingly enough, began to delineate the early logic of how to think about development (1). However, Davidson was always driven by biology: in his case, by a desire to understand the molecular basis of sea urchin development. In the intervening years, Davidson and his colleagues came to understand the biology of sea urchin development at a uniquely deep level, and this knowledge was the driver and integrator of subsequent developmental studies. Over the past 15 years or so, Davidson arrived at the realization that development was hardwired in the gene regulatory logic of the genome and that determination of this logic would require a systems approach (i.e., global analyses, rather than the study of one or a few genes or proteins at a time). He also came to understand that the systems approach to development required (i) determination of the complete sea urchin genome sequence, to compile a complete “parts list” of sea urchin genes, especially those encoding transcription factors; (ii) an understanding of the deep logic of how individual genes are regulated by transcription factors through their *cis*-regulatory elements; and (iii) a delineation of the dynamics of complete gene regulatory networks (GRNs) specifying the distinct stages of sea urchin development. Complete GRNs—the dynamic network relationships of all relevant transcription factors—are essential because the genomic regulatory logic is so complex that inferences from incomplete networks can be misleading or inadequate. For each of these challenges, Davidson and his colleagues carried out groundbreaking work. Davidson was almost single-handedly responsible for initiating the sea urchin genome project, and he certainly inspired what was perhaps the most thorough annotation of any genome at that time (2). His group specified, in detail, the *cis* regulation of the developmentally regulated endo-16 gene and demonstrated the fascinating logic and complexity of this regulatory circuit (3). Finally, Davidson and his associates pioneered the preliminary delineation of a variety of GRNs and their specification of sea urchin larval development (see ref. 4).

Skeletal Cell Development

In this issue of PNAS, the article by Oliveri, Tu, and Davidson (5) is the most powerful demonstration to date of how a dynamically changing, complete GRN can explain virtually every aspect of skeletal-cell development in the sea urchin larva over its first 48 h. The authors focus on specification of the skeletogenic micromeres that ultimately differentiate into skeletal cells exhibiting mineralization. The idea is to determine the dynamic architecture of the GRN specifying this process, that is, the nature of the interactions of the transcription factors that mediate development in

Sophisticated computational studies must be driven by good biology.

these cells over the 48-h time period studied. The general approach was as follows: (i) a list of all identified DNA-binding domains (proteins) was determined from the complete genome sequence (it is believed that this list is relatively complete); (ii) dynamic quantitative PCR (QPCR) was used to ascertain which of these regulatory genes were present during embryogenesis; (iii) the spatial localization of these regulatory genes was determined by whole-mount *in situ* hybridization for up to 48 h (and was believed to be sufficiently sensitive to observe low-abundance transcription factors); (iv) all of the micromere-expressed transcription factors were suppressed individually by morpholinos, and the resulting perturbations (network linkages) in gene expression were recorded dynamically across the 48-h time period by QPCR; and (v) a delineation of the *cis*-regulatory structures for several genes in the network was used to check the architecture of the GRN. In addition, key nodal points in the GRN were perturbed by clever biological experiments, to check and extend the network architecture. Moreover, the signal transduction factors initiating this developmental process were identified, and their transcript expression was perturbed by morpholinos and appropriate biological experiments.

From these experiments, a complete and dynamically changing GRN was constructed (complete in that it contained all of the transcription factors expressed in the micromeres), and this network could be used to explain virtually every known phenotypic feature of skeletogenic micromere development. This GRN also begins to reveal the underlying modular logic of the GRNs (e.g., the double-negative gate and lockdown of GRN subnetworks). These approaches, the GRN, and its developmental logic and conclusions are elegantly described by Oliveri *et al.* (5), and I will not discuss them further here. Rather, I will examine some of the broader implications and questions arising from these elegant studies.

Implications and Speculations

The logic of sea urchin development does, indeed, appear to be hardwired in the genome through its specification of all transcription factors and through the logic of their assembly into GRNs by the *cis* binding of the transcription factors to one another and to their batteries of developmental effector genes. It may be surprising to some that epigenetic, microRNA, and protein regulatory phenomena do not play a larger role in this process, but it is logical that development—a process that must build all individuals of a given species in a relatively invariant manner—is largely controlled by the species-invariant digital genome (apart from polymorphisms). This understanding opens the possibility for us to begin thinking about all of the GRNs that regulate development in the adult and larval sea urchins—indeed, in the entire organism—and ultimately we can dream of a similar accomplishment for humans.

On the basis of studies by Davidson and others taking similar approaches, a critical opportunity is emerging to begin evolutionary comparisons of GRNs specifying similar aspects of phenotype. With the next generation of sequencing technologies and their promise to dramatically speed DNA sequencing and significantly reduce its cost comes the

Author contributions: L.H. wrote the paper.

The author declares no conflict of interest.

See companion article on page 5955.

*To whom correspondence should be addressed. E-mail: lhood@systemsbiology.org.

© 2008 by The National Academy of Sciences of the USA

possibility of generating thousands of developmentally relevant genome sequences in the next 5–10 years. Such data will require a revolution in our techniques for comparative sequence analysis and will allow us to begin answering a variety of interesting questions. The ultimate aspiration is to be able to deduce the GRN directly from the genome sequence, by delineating the network connections through an accurate computational folding of the transcription factors (and other proteins) into their 3-dimensional structures, coupled with the capacity to describe their interactions with other macromolecules and small molecules. Are there core logical strategies or subnetworks in the GRNs that are highly conserved across species? How does the architecture of the GRN evolve across closely related and more distantly related species? Will the strategies and logic of deciphering sea urchin development apply to mice and humans? Can we use stem cells, and the potential for their guided development, to generate interesting cell lineages and create more experimentally accessible developmental systems for higher organisms? Indeed, with the emergence of microfluidic and nanotechnology methodologies to separate, study, and characterize single cells and to mix any combination of two or three cells together to analyze their interactions, stem cells may provide a fascinating window into development.

How do the GRNs that specify physiological processes differ from their developmental counterparts? Will the architectures of the physiological networks differ significantly? Will epigenetic, microRNA, and protein regulatory events play a more important role for physiology than for development? What will be the relationships between the signal transduction networks and the physiological GRNs?

Many of the studies involving systems approaches to development and physiol-

ogy employ high-throughput data-gathering (gene expression, protein expression, protein/protein and protein/DNA interactions, localization, etc.). Davidson's work elegantly demonstrates the power of a deep knowledge of biology for generating systems perturbations that elucidate the subtle and complete details of a more circumscribed aspect of biological phenotype, as well as the power of quantitative, more-circumscribed measurements. Do the high-throughput approaches have sufficient signal-to-noise discrimination to actually build accurate networks? Can they discriminate among the subtleties of the regulatory logic (e.g., repression vs. enhancement)? These methods certainly have potential for use in global studies, but without a true understanding of the biology, will the studies lead to deeper insights? Davidson's approach argues powerfully that sophisticated computational studies must be driven by good biology. Indeed, Davidson and his colleagues have made significant contributions to the creation of computational tools for developmental studies—contributions driven by the biological needs themselves (see ref. 6).

The findings of Oliveri *et al.* (5) suggest that, in time, we will be capable of delineating the GRNs that regulate the complex processes of human development and physiology. Presumably we will also be able to determine the dynamics of the disease-perturbed GRNs that mediate a particular disease. We can truly envision a time when, through a multilayered understanding of network function, we will be able to reengineer networks by using drugs to manipulate the expression or behavior of key nodal points, so that the network behaves in a more normal manner or abrogates some of the deleterious consequences of the disease-perturbed network function. Perhaps in the future, drugs can be designed to prevent networks from becoming disease-perturbed.

It is interesting to note that the use of the sea urchin to study molecular development was originally viewed with skepticism because the animal could not be subjected to genetic analysis. The sea urchin today beautifully illustrates the power of a model organism selected because of its special and accessible features and because of emerging technologies that make relevant measurements more accurately (and globally). Development in the sea urchin can be readily visualized in both its early and late stages. One can effectively transfect large numbers of sea urchin eggs and easily follow their development. Very large numbers of eggs are readily available. The rich history of sea urchin developmental studies, extending back more than 100 years, even today provides useful phenotypic insights. Add to this the dedication of Eric Davidson and his many students and colleagues in developing the sea urchin as a model and in applying emerging molecular approaches to the elucidation of its features, and one has a uniquely powerful model for development. With appropriate biology and technology, the lack of genetics is not a fatal flaw. Indeed, with the emerging technologies, biologists and physicians will have ready access to other “genetically intractable” organisms, including humans.

As one who has pondered systems biology and its approaches to development and physiology for the past 20 years, I judge Oliveri *et al.* (5) a tour de force. It represents a brilliant integration of biology, technology, computational approaches, and powerful logic to achieve a real first—the determination of a dynamic GRN that specifies the skeletal lineage in the sea urchin larva and can explain virtually every aspect of its relevant phenotypes. This paper will be the model for many more that will undoubtedly follow, transforming the landscape of developmental biology and ultimately elucidating the molecular systems that drive development.

1. Britten RJ, Davidson EH (1969) Gene regulation for higher cells: A theory. *Science* 165:349–357.

2. Yuh CH, Bolouri H, Davidson EH (2001) *Cis*-regulatory logic in the endo16 gene: Switching from a specification to a differentiation mode of control. *Development* 128:617–629.

3. Sodergren E, *et al.* (2006) The genome of the sea urchin *Strongylocentrotus purpuratus*. *Science* 314:941–952.

4. Davidson EH (2006) *The Regulatory Genome: Gene Regulatory Networks in Development and Evolution* (Academic, San Diego).

5. Oliveri P, Tu Q, Davidson EH (2008) Global regulatory logic for specification of an embryonic cell lineage. *Proc Natl Acad Sci USA* 105:5955–5962.

6. Longabaugh WJR, Davidson EH, Bolouri H (2005) Computational representation of developmental genetic regulatory networks. *Dev Biol* 283:1–16.