Nucleus 4:2, 89–91; March/April 2013; © 2013 Landes Bioscience

Networking development by Boolean logic

Shikui Tu,[1,3] Thoru Pederson[2,3] and Zhiping Weng[1,3,*]

[1]Program in Bioinformatics and Integrative Biology; University of Massachusetts Medical School; Worcester, MA USA; [2]Program in Cell and Developmental Dynamics; University of Massachusetts Medical School; Worcester, MA USA; [3]Department of Biochemistry and Molecular Pharmacology; University of Massachusetts Medical School; Worcester, MA USA

Eric Davidson at Caltech has spent several decades investigating the molecular basis of animal development using the sea urchin embryo as an experimental system[1,2] although his scholarship extends to all of embryology as embodied in several editions of his landmark book.[3] In recent years his laboratory has become a leading force in constructing gene regulatory networks (GRNs) operating in sea urchin development.[4] This axis of his work has its roots in this laboratory's cDNA cloning of an actin mRNA from the sea urchin embryo (for the timeline, see ref. 1)—one of the first eukaryotic mRNAs to be cloned as it turned out. From that point of departure, the Davidson lab has drilled down into other genes and gene families and the factors that regulate their coordinated regulation, leading them into the GRN era (a field they helped to define) and the development of the computational tools needed to consolidate and advance the GRN field.

The nodes in a GRN represent regulatory genes (genes that encode transcription factors and signaling molecules) and the edges represent regulatory logic and gene interactions encoded in genomic cis-regulatory binding sites. GRNs are constructed using four types of experimental data: temporal expression patterns of regulatory genes, spatial expression patterns of regulatory genes, trans-perturbation data (the change in gene expression patterns when one or more regulatory genes are perturbed) and cis-perturbation data (the change in gene expression patterns when one or more cis-regulatory binding sites are mutated). In principle, GRNs provide a system-level, causal model for the spatial and temporal patterns of gene expression and mechanistic understanding of how a zygote develops into an embryo.

The decades of effort by Davidson and colleagues has resulted in a GRN for endomesoderm speciation for the sea urchin embryo, one of the most complete GRNs currently available for animal development. Yet, this GRN is a static conglomeration of heterogeneous experimental data. It is unknown whether this GRN is complete; more importantly, it is unknown whether it can predict the progression of the large-scale developmental process. In a recent PNAS publication[5] Peter, Faure and Davidson built a Boolean computation model based on this GRN in order to address these two questions. They analyzed the latest version of the endomesoderm GRN which contained 50 regulatory genes plus the regulatory interactions controlling the specification of endoderm and mesoderm from early cleavage stages (six hours post-fertilization) to the onset of gastrulation (30 h). They argued that spatial expression was discrete in nature and could be captured by "on" or "off" states in a Boolean model. Another key assumption in their model was that the time interval between the activation of a regulatory gene and its immediate downstream regulatory gene in the GRN is three hours for all the genes in the GRN, with this assumption resting on the notion that this is the amount of time required for transcription of the upstream gene, synthesis of the protein (together with a protein half-life much longer) and binding

Keywords: sea urchin, development, specification, gene regulatory networks, computational embryology

Submitted: 01/18/13

Accepted: 02/12/13

http://dx.doi.org/10.4161/nucl.23986

*Correspondence to: Zhiping Weng; Email: zhiping.weng@umassmed.edu

Commentary to: Peter IS, Faure E, Davidson EH. Predictive computation of genomic logic processing functions in embryonic development. Proc Natl Acad Sci U S A 2012; 109:16434-42; PMID:22927416; http://dx.doi.org/10.1073/pnas.1207852109

of the protein to the cis-regulatory region of the downstream gene. They converted the GRN to a set of 75 vector equations (e.g., if gene A in domain 1 is on and gene B in domain 2 is off, then gene C in domain 1 is on; otherwise gene C in domain 1 is off, etc.) and tested whether, starting with non-zygotic (i.e., maternal) inputs as the initial state the model, it could run autonomously and reproduce the temporal and spatial expression patterns during the developmental processes of four cellular domains of the sea urchin embryo.

Implications and Speculations

Encouragingly, the Boolean model was able to make highly accurate predictions on spatial and temporal gene expression patterns. Indeed, only several predictions among the 2,772 time-space-gene combinations were at odds with experimental data. This may not seem surprising because the GRN is based on interpretations of huge masses of expression data, perturbation data and cis-regulatory data. Peter et al. proceeded to perform more stringent tests of the model by asking how it would respond to four perturbations: extinction of the expression of the *delta* gene, global expression of the *pmar1* gene, extinction of *hox11/13b* expression, and most challengingly, the transplantation of four cleavage skeletogenic micromeres into the animal pole of an otherwise normal embryo that possessed its own set of four micromeres at the vegetal pole. (The intra-embryo transposition of blastomeres harkens back to the classical era of experimental embryology). The authors emphasized that except for the *hox11/13b* test, the perturbation results they sought to reproduce were not used in building the GRN. (Transplanted micromeres could of course not have been part of the GRN, which is for a normal embryo). The results of these perturbation tests were in nearly perfect agreement with the experimental data, which led the authors to two conclusions: the GRN contained sufficient information to provide a system-level causal explanation for sea urchin development, and the Boolean computational model was a useful tool for in silico testing of GRN and making predictions upon perturbations.

The test of blastomere transplantation demonstrated the critical role of intercellular signaling between the different spatial domains in development. The cells in these domains obviously need to work cooperatively to ensure precise and robust developmental progression and thus information on the regulatory state of each cell must be able to diffuse spatially. Peter et al. defined a Janus factor for every inductive signaling interaction in their Boolean model. It would be interesting to investigate the molecular basis of intercellular signaling here. We enjoy so many examples of such events in development, e.g., the Wnt pathway in Drosophila development to mention just one of many examples, but we have no case in which a paracrine signaling pathway amidst a developmentally determinative cluster of cells can be put into the context of a GRN as detailed as the one Peter et al. have defined. From the information theory perspective, intercellular signaling can be regarded as communicating messages between two or more intelligent agents. Thus, the developmental process is also a diffusion process of the genomic regulatory information. Related domains of this emerging field include "molecular information theory"[6] and information networks in the data mining field[7] from both of which the modeling of intercellular signaling may benefit.

Peter et al. went so far as to suggest that the gene regulatory models could sufficiently explain all the gene expression patterns in sea urchin development, without considering non-coding RNAs. For example, a recent study identified long noncoding RNAs (lncRNA) in zebrafish embryogenesis[8] and revealed that lncRNAs were specifically enriched in early-stage embryos. lncRNAs of course must be integrated into all current GRN research. Ironically, Davidson and his longtime partner Roy Britten were the first to postulate such a role.[9,10]

Peter et al. concluded that Boolean modeling is a useful implementation of the original GRN concept. Is there any information loss from the continuous data to Boolean data, from gene regulatory logic to Boolean logic? What would be a good cutoff for converting the expression level to "on" or "off"? Does the cutoff depend on the gene or the spatial domain? If a GRN is not available in a given case, can a Boolean model be directly inferred and tested using raw experimental data? If so, what kinds of experimental data are most suitable and in what way can the causal structure be best captured in the Boolean model? Theoretical frameworks of causal inference from observational data have been studied in machine learning for decades.[11] Possibly the established causal inference algorithms can expand this newest chapter in GRN-ology from the pioneering Davidson lab into more general settings. Can this work provide insights into other animal development, e.g., mouse, human or into human embryonic stem (ES) cell differentiation? What are the conserved parts and new features across species? Is it possible to apply the general computational framework to test GRN models derived from biological processes in addition to development, such as immunology and postnatal neurogenesis in the subventricular zone of the brain to mention only two of many frontiers before us that beckon for the GRN approach? With the advances of next generation sequencing technology and its cost now rocketing downward, a large amount of data will soon be in hand for various biological processes across diverse phyla. Much of this momentum now comes from the sea urchin embryo and from the scientific mind of Eric Davidson.

Disclosure of Potential Conflicts of Interest

No potential conflicts of interest were disclosed.

References

1. Pederson T. The sea urchin's siren. Dev Biol 2006; 300:9-14; PMID:17113576; http://dx.doi.org/10.1016/j.ydbio.2006.10.006.
2. Ernst SG. Offerings from an urchin. Dev Biol 2011; 358:285-94; PMID:21723860; http://dx.doi.org/10.1016/j.ydbio.2011.06.021.
3. Davidson EH. Gene Activity in Early Development. 1968. Academic Press.
4. Davdison EH. Genomic Regulatory Systems. Development and Evolution. 2001. Academic Press.
5. Peter IS, Faure E, Davidson EH. Predictive computation of genomic logic processing functions in embryonic development. Proc Natl Acad Sci U S A 2012; 109:16434-42; PMID:22927416; http://dx.doi.org/10.1073/pnas.1207852109.
6. Schneider TD. A brief review of molecular information theory. Nano Commun Netw 2010; 1:173-80; PMID:22110566; http://dx.doi.org/10.1016/j.nancom.2010.09.002.
7. Sun Y, Han J. Mining heterogeneous information networks: principles and methodologies. Synth Lect Data Mining Knowl Disc 2012; 3:1-159; http://dx.doi.org/10.2200/S00433ED1V01Y201207DMK005.
8. Pauli A, Valen E, Lin MF, Garber M, Vastenhouw NL, Levin JZ, et al. Systematic identification of long noncoding RNAs expressed during zebrafish embryogenesis. Genome Res 2012; 22:577-91; PMID:22110045; http://dx.doi.org/10.1101/gr.133009.111.
9. Britten RJ, Davidson EH. Gene regulation for higher cells: a theory. Science 1969; 165:349-57; PMID:5789433; http://dx.doi.org/10.1126/science.165.3891.349.
10. Pederson T. The discovery of eukaryotic genome design and its forgotten corollary--the postulate of gene regulation by nuclear RNA. FASEB J 2009; 23:2019-21; PMID:19567373; http://dx.doi.org/10.1096/fj.09-0702.
11. Pearl J. An introduction to causal inference. Int J Biostat 2010; 6:7; PMID:20305706; http://dx.doi.org/10.2202/1557-4679.1203.