HHS Public Access

Author manuscript

Hist Philos Life Sci. Author manuscript; available in PMC 2017 October 29.

Published in final edited form as:

Hist Philos Life Sci. ; 39(4): 37. doi:10.1007/s40656-017-0164-z.

Fitting structure to function in gene regulatory networks

Ellen V. Rothenberg

Division of Biology & Biological Engineering, California Institute of Technology, Pasadena, CA 91125 USA

Abstract

Cascades of transcriptional regulation are the common source of the forward drive in all developmental systems. Increases in complexity and specificity of gene expression at successive stages are based on the collaboration of varied combinations of transcription factors already expressed in the cells to turn on new genes, and the logical relationships between the transcription factors acting and becoming newly expressed from stage to stage are best visualized as gene regulatory networks. However, gene regulatory networks used in different developmental contexts underlie processes that actually operate through different sets of rules, which affect the kinetics, synchronicity, and logical properties of individual network nodes. Contrasting early embryonic development in flies and sea urchins with adult mammalian hematopoietic development from stem cells, major differences are seen in transcription factor dosage dependence, the silencing or damping impacts of repression, and the impact of cellular regulatory history on the parts of the genome that are accessible to transcription factor action in a given cell type. These different features not only affect the kinds of models that can illuminate developmental mechanisms in the respective biological systems, but also reflect the evolutionary needs of these biological systems to optimize different aspects of development.

Keywords

Transcriptional regulation; network topology; Boolean modeling; T lymphocyte development; chromatin accessibility; developmental plasticity

1. General and particular gene regulatory network templates for development

A prophetic insight into the way the genome regulates development began to take shape in a set of papers published by Roy J. Britten and Eric H. Davidson from 1969–1973 (Davidson and Britten 1973; Britten and Davidson 1971, 1969). Recognizing the primacy of differential gene expression for development, they postulated the existence of genomically encoded regulatory molecules that could act combinatorially on specialized regulatory DNA sequences, distinct from protein-coding sequences, to control target gene expression in a cell type-specific way. Because these genomically encoded regulators themselves need to have

For the Eighth International Workshop, Jacques Loeb Centre for the History and Philosophy of the Life Sciences, Dec 2015: “From Genome to Gene: Causality, Synthesis, and Evolution”

their own expression controlled, the progressive diversification of cell types that is integral to development logically implies the existence of a hierarchical causal network for gene expression control with computer-like information processing at each regulatory node (Yuh et al. 1998; Markstein et al. 2004; Stanojevic et al. 1991). Britten and Davidson formulated such a network long before there was technology available to identify its actual components and modes of operation (Davidson and Britten 1973; Britten and Davidson 1969). In its general sense, this network formulation has been confirmed triumphantly by the molecular biology of the succeeding 45 years (Briscoe and Small 2015; Simoes-Costa and Bronner 2015; Clark et al. 2014; Hobert 2014; Peter et al. 2012; Bonzanni et al. 2013; Kueh and Rothenberg 2012; Chen et al. 2012; Jaeger 2011; Laslo et al. 2008; Oliveri et al. 2008; Stathopoulos and Levine 2005a; Meulemans and Bronner-Fraser 2004; Davidson et al. 2002a).

The crucial elements that have made it possible to reveal gene network circuitries that operate in different developmental systems have been (1) high definition of the biological system in terms of cell types present at different timepoints; (2) extensive identification of relevant gene products expressed in these cells at different time points; (3) application of perturbations that can acutely remove or manipulate candidate regulatory molecules at relevant time points; and (4) ways to measure the impact of these perturbations quantitatively on the other genes in the system. These approaches have provided a wealth of data to show that the key regulatory molecules for early embryonic differentiation are proteins that bind DNA sequence-specifically to control gene expression, i.e. transcription factors. To validate how these transcription factors are actually working on the genes they control, it has also been vital (5) to use molecular genomic technology to identify the regulatory DNA of the affected target genes and (6) to use targeted mutation and gene transfer strategies to map the sequences through which the effects of the transcription factors are mediated. These six elements have been powerfully applied to systems as diverse in function and phylogeny as early embryonic axis formation in *Drosophila* (Chen et al. 2012; Jaeger 2011; Stathopoulos and Levine 2005b), the chick and lamprey neural crest (Simoes-Costa and Bronner 2015; Nikitina et al. 2008), the photoreceptor patterning of eyes (Viets et al. 2016; Brzezinski and Reh 2015), and the sea urchin early embryo endomesoderm patterning and differentiation (Peter and Davidson 2011b; Oliveri, Tu, and Davidson 2008). Eloquent summaries of multiple partial networks and their components, together with extensions of the theoretical implications of these networks to unsolved problems in development and evolution have been presented by Davidson and colleagues (Peter and Davidson 2015, 2011a; Davidson 2010; Erwin and Davidson 2009; Davidson 2006).

2. What network models tell us

There is little doubt that ultimately there are gene networks that drive development. To be understood, these can be embodied in logical connectivity models that represent the ways in which each gene activated in the course of development subsequently affects the expression of other genes in the network, and any dependencies of each of those effects on the activities of other genes (AND or OR logic). But the network models that can be constructed by experimental application of the six elements described above most often remain incomplete in important respects. First, the ranges of target genes examined rarely include all the genes

in the genome. Most genes are not relevant to a given developmental choice or progression step and so some of the art of network analysis is to use sufficient prior knowledge to focus on genes that are most likely to be functionally important – but prior knowledge is not infallible or clairvoyant. Second, the responses to perturbation that define one gene as being downstream of another are not always simple or direct in fact. Although an initial network model can show that a perturbation affects a given gene, the actual mechanism through which the effect is seen can go through multiple intermediate steps that are themselves contingent on other genes. In a common problem, a gene that is only expressed in a particular cell type often appears, spuriously, to be regulated by any transcription factor that affects generation of that cell type. Third, it is not always straightforward to determine whether a functional response of a gene is significant. Transcription factors can work on targets via “additive OR” logic, i.e., where they enhance expression even where they are not absolutely required for expression. Such quantitative effects may be important biologically even if they are not required for all target expression, but measurement error can give false negative results. Finally, even when transcription factors work on target genes, they can work through a variety of different biochemistries, and this can complicate the network analysis as discussed in much more detail below.

How much value, then, can one get from provisional gene network models, such as those that can be constructed with a “reasonable” amount of experimental effort? And how much effort beyond this is worth spending to make networks more complete? Even incomplete network models are indeed useful. They are crucial to capture large bodies of complex experimental results in a comprehensive way. They can clarify understanding of biological relationships that do not fit a preconceived notion: for example, they can show that if several genes failed to fit in a clear linear epistasis pathway, it may be because the genes actually are engaged in a feedback circuit. Provisional gene network models are also operationally powerful to point the way to experiments that improve understanding. Often, a real network circuit will turn out to be simpler than initial results suggest, for example if one of the noted inputs actually works on a target gene by way of another. The experiments needed to indicate the right simplification become evident by looking at the model. By making models and comparing them with observed gene expression patterns, one can also see what kinds of functions are missing, in order to account for the actual expression trends of each target relative to its proposed inputs.

3. Network completeness: success within reach

Incompleteness clearly limits the accuracy of a network model in its ability to explain or predict the actual biological process. If a key positive or negative regulator is not identified, its role will remain either unexplained or inaccurately assigned to another factor or factors that do(es) not actually carry out the needed function. In the case of explaining the development of early embryos, the network model cannot be predictive unless it accounts for the availability of maternally derived molecules pre-loaded into the egg during oogenesis, which can exert extremely powerful functions in many animals (even if not in mammals). In the case of explaining differentiation from long-term adult stem cells as described below, a network model similarly may not explain gene expression dynamics unless initial silenced states, enforced by repressive mechanisms involved in “stem-ness”, are taken into account

(Hodges et al. 2011; Oguro et al. 2010; Pietersen and van Lohuizen 2008; Maier et al. 2004; Tagoh et al. 2004). Another form of incompleteness is a failure to understand the logic of a given input into its target. This is especially important to acknowledge in view of the way transcription factors actually collaborate to regulate target genes. Not all “sufficient” inputs, as defined by gain of function experiments, are also “necessary”; and similarly, not all “necessary” inputs, as defined by loss of function experiments, are “sufficient”. At the extremes, two inputs might work by “AND” logic (both necessary, neither sufficient) or “OR” logic (neither necessary, either sufficient) (Istrail et al. 2007), but in reality the behavior can be intermediate between these extremes, or bimodal.

Completeness can be approached by two complementary strategies. One “building out” strategy is to analyze increasing numbers of network connections until the results that can be computed for the behavior of the known network are predictive of the actual biological system’s behaviors under all conditions (Schutte et al. 2016; Clark and Akam 2016; Bonzanni et al. 2013; Peter, Faure, and Davidson 2012; Jaeger 2011; Davidson et al. 2002b). Another “whole genome catalog” possibility is to start from all the genes expressed in the genome of the cells within the system, and then to define the developmentally relevant inputs to all these genes to the point that their expression is understood (Hughes et al. 2014; May et al. 2013; Zhang et al. 2013; Ciofani et al. 2012; Novershtern et al. 2011). Whereas both strategies would have seemed fantastic to researchers as recently as 15 years ago, they have become feasible in principle due to advances in methods for measuring gene expression and genome-wide transcription factor binding across the whole genome, and methods for inducing targeted mutation or antagonism of nearly any gene in the genome. The closest approach to a success for complete network analysis has been in the sea urchin endomesoderm network, which was solved to a very high criterion of success by the building-out strategy. To measure its approach to completeness, the key was to devise a kind of computational modeling that would determine how well the sum of existing network connectivity knowledge could perform in predicting gene expression. Without fitting model connection parameters to actual gene *expression* data, Peter, Faure and Davidson instead used 15 years’ worth of evidence for individual, measured gene-to-gene regulatory *connections* (activations or repressions), accumulated from many researchers, together with known impacts of anatomically localized signaling events in the embryo, known blastomere cleavage planes, and an absolute time step scale, to create a sophisticated Boolean model (Peter, Faure, and Davidson 2012). When computed from time zero of fertilization to 30 hr of development, this model accurately reproduced the vast majority of known gene expression data in this embryo (Peter, Faure, and Davidson 2012).

This is a triumph for the building-out approach, but its solution has been highly labor-intensive. Furthermore, it retains some reservations. One is that it is not strictly proven to be complete, lacking full genome-wide analysis. Also, there are specific features of the early embryo system that have been important for making this system soluble in terms of a Boolean model, which should be noted.

In life, the operation of the early embryo network is fast and deterministic, and the Boolean model takes advantage of this. In embryos developing by “mode 1” embryogenesis like sea urchin embryos (Peter and Davidson 2015, ch. 3; Davidson 1991), maternal macromolecules

loaded into the egg provide initial polarity and the embryonic fate map is set by canonical cleavage planes. Where gene expression requires triggering by an intercellular signal, the positions of the cell delivering and the cell receiving the signal in each case are invariant due to this reproducible cleavage pattern. In terms of the gene expression linkages themselves, there are three key features that characterize this kind of system. First, gene expression responds rapidly to any new appearance of activators in the system: there is little if any temporal buffering (i.e., as soon as a quorum of needed factors is present). Second, gene expression responds to low doses of activator almost identically as to higher doses of activator: the system saturates its responses at very low levels of positive regulators (Bolouri and Davidson 2003) – although this "rule" is controversial in light of the role of gradients in patterning, e.g. (Briscoe and Small 2015; Reeves et al. 2012). Third, when a trans-acting factor causes a negative regulatory effect, it tends to silence the target gene, not just moderate its expression. This means that transcription factor inputs in the network can be treated in a Boolean fashion and can yield digital all-or-none outputs in terms of target gene expression, and that the time constants for response leave minimal delay between the synthesis of new regulators and the appearance of the response. While these system features of the early sea urchin embryo have made the Boolean modeling approach both powerful and effective, they are not universal in developmental systems.

4. Stem-cell based development: model for stochastic fate determination systems?

In postnatal animals, it is now known that many tissues stockpile cells whose differentiation is suspended, but which remain available as a reservoir of precursors for renewal and/or repair. These stem cells have been studied longest, and remain best characterized, in the hematopoietic system, where it is clear that the stem cells retain both pluripotency (>10 potential cell-type derivatives, many of them further subspecialized) and an impressive self-renewal potential. The stem cell differentiation clock is by definition uncoupled from the organism's overall ontogenic clock, in that any given stem cell makes a series of independent decisions to further postpone or to undergo differentiation, as long as it is still a stem cell, repeatedly and out of synchrony with other stem cells. In the case of hematopoietic cells, not only timing but also differentiation pathway choices and differentiated cell outputs can be variable. The potential for extensive but variable proliferation can further be propagated to lineage-committed blood-cell precursors descended from stem cells, like those determined to give rise to dendritic cells, erythroid cells, or T lymphocytes (Naik et al. 2013; Perié et al. 2015; Manesso et al. 2013; Yui and Rothenberg 2014). Importantly, in between stem cells and committed precursors, a variety of highly proliferative developmental intermediates exhibit stereotyped patterns of partial lineage restriction, implying that lineage choice occurs by stepwise subtraction of options even while fate indeterminacy persists, rather than by coordinated initiation of a particular developmental program. All of these features are markedly different from those of mode 1 developing embryos, and they are associated with specific differences in the operating principles of the relevant gene networks.

These differences emerge from gene networks with important roles for regulator dosage sensitivity, conditionality of inter-regulator repression, and a substantial influence of history on the range of effects a transcription factor can exert. The combinations of fates that can remain open in particular precursors can be traced to expression of transcription factors that are necessary for particular fates but which can support other fates, too, at different levels. Dosage sensitivity calls for dosage control mechanisms, and it turns out that many of the regulatory factors used in hematopoietic differentiation can exert partial damping effects on other regulatory genes without silencing them. Finally, the response to a particular transcription factor even within a given lineage can be delayed substantially with respect to the factor's synthesis, both due to synthesis and turnover kinetics and due to interactions of the factor with slowly changing chromatin conformation states around its target sites. These effects add considerable complexity to modeling of causality in gene expression dynamics.

5. Regulator dose dependence in early T-cell development

As multipotent blood cell precursors begin to differentiate toward a T-lymphocyte fate, they use a constellation of transcription factors that are partially held over from more pluripotent precursors and partially induced de novo by signals presented in the microenvironment of the body's T cell "nursery", the thymus (Rothenberg et al. 2016; Yui and Rothenberg 2014). Among the transcription factors that are crucial for enabling cells to become T cells are a relatively T-cell specific factor, GATA-3, a factor used to make T and B lymphocytes, E2A, and a factor used for various non-T cell fates as well as the early steps of T-cell development, PU.1. In addition, to promote the T-cell fate, the cell needs to be able to use the Notch signal response system (the bifunctional Notch1 molecule, plus signal processing molecules) to sense signals from the microenvironment. Although the cells express all these factors simultaneously, and require all these factors simultaneously, at least two of them are capable of poisoning the T-cell program if expressed at somewhat higher levels. PU.1 expressed at too high a level kills the cells or diverts them to a non-lymphocyte fate, in accord with its key role in the same non-lymphoid fates. More surprisingly, the "T-cell specific" factor GATA-3 at too high a level also kills the cells or diverts them to a different non-lymphocyte fate, converting them to mast cells (Taghon et al. 2007). The reason that PU.1 can block T-cell development is in large part due to the ability of high-level PU.1 to shut down the effectiveness of the Notch signaling system (Del Real and Rothenberg 2013; Champhekar et al. 2015). The reason that high-level GATA-3 can block T-cell development appears in part to be due to its ability to shut down PU.1 activity, and in part to its ability to mimic roles of a related GATA factor that is normally expressed in the mast cell fate, GATA-2 (Scripture-Adams et al. 2014; Taghon, Yui, and Rothenberg 2007).

Dose dependence implies that each factor is titrating some interaction partner or constraint. Both PU.1 and GATA-3 have constraints that are normally embedded in the same T-cell program in which they work themselves. PU.1 cannot block Notch signaling unless it is expressed at a high level. As long as Notch signaling continues, PU.1 can promote expression of numerous target genes without extinguishing the T-cell program genes and without activating key parts of the non-T myeloid–cell program (Del Real and Rothenberg 2013). Preliminary results (H. Hirokawa, J. Ungerbäck, and E.V. R., unpublished data) suggest that PU.1 protein is expressed at similar levels and binds to similar or identical

genomic sites with or without Notch signaling, but that Notch signaling may affect its ability to interact with other transcription factors on the DNA to modulate its activity. Thus the quantitative balance between PU.1 level and environment-driven Notch signaling intensity likely works as a bistable switch to determine whether PU.1 works within the T-cell pathway or as a diversion from it (Del Real and Rothenberg 2013). It is a balance between continuous-valued regulatory inputs that yields a Boolean output – but it is a mechanism that is difficult to use to predict the output of a particular level of PU.1 expression.

The reasons for the extreme toxicity of high-level GATA-3 are not completely defined. However, there are several clues. GATA-3 has a non-DNA binding interaction partner, FOG-1 (Zfpm1), that is normally expressed at about the same level as GATA-3 in early T cells. FOG-1 interaction with GATA family factors usually affects their functionality, either to promote activation or to promote repression (Miccio et al. 2010; Letting et al. 2004). If this complex has different effects from GATA-3 without FOG-1, then excess GATA-3 could be functionally competing with GATA-3/FOG-1 heterodimers for activity at important DNA target sites. Alternatively, GATA-3 in excess could bind to an inappropriate spectrum of DNA sites for the stage of development. Unlike PU.1, GATA-3 protein is normally recruited to different subsets of its target sites from one stage of T-cell development to the next (Zhang et al. 2012), even though the levels of GATA-3 protein change very little (Scripture-Adams et al. 2014). This probably reflects weak DNA-binding interaction that is stabilized normally by contacts with other, locally bound transcription factors, and avoidance of stable binding at sites where other factors are not present. Excess GATA-3 could make site occupancy less dependent on partners, thus causing incoherent gene regulation responses. One of these effects is the aberrant downregulation of PU.1, but excess GATA-3 could be poisoning the T-cell program through other effects as well.

6. Soft repression in early T cells

The T-cell differentiation program has evolved to be protracted through multiple cell cycles before resolving into T-cell lineage commitment, apparently in order to enable each precursor to generate large clones of cells that can develop into mature T cells with very diverse antigen recognition specificities at later stages of development (Lu et al. 2005). This program thus keeps alternative differentiation at bay through 10–12 cell cycles (Manesso et al. 2013) despite continuing to use factors, like PU.1, that preserve some possible access to other developmental options. A prominent mechanism that may hold the key to this slow progression toward commitment is the use of cross-regulation to limit but not to silence expression of other regulators. A striking case of this is the way E2A is used. E2A works in part as a positive regulator of Notch pathway signaling competence (Yashiro-Ohtani et al. 2009), and thus sustains the cells' ability to stay on course for the T-cell program despite PU.1 expression. This role fits well with E2A's continued expression even after PU.1 is finally turned off and the cells are committed to T-cell development. More surprisingly, E2A turns out to be a significant negative regulator of GATA-3 (Xu et al. 2013), the factor that is co-expressed and co-sustained with E2A in the T-cell program long after commitment. It turns out that E2A activity is needed to prevent GATA-3 overexpression, and either E2A mutation or expression of an E2A antagonist (Id2) can lead to overexpression of GATA-3 (Xu et al. 2013; Del Real and Rothenberg 2013). If E2A were to silence GATA-3, as

generally happens between regulators and antagonistic regulatory genes in embryonic boundary formation, it would destroy T-cell development because GATA-3 is so acutely required by developing T cells. However, instead it sustains T-cell development by preventing GATA-3 levels from rising too high.

Soft repression like this is also seen in responses to high-level PU.1 by a variety of regulatory genes that are normally expressed in the same stages as PU.1 (Champhekar et al. 2015). Single-cell multiplex gene expression analysis is ongoing to verify whether these soft repression targets are actually co-expressed with PU.1 in individual cells (W. Zhou, M. A. Yui and E. V. R., unpublished data). But the population-level evidence strongly suggests that multipotency, while it lasts, is also a kind of mutual nonaggression state enforced by cross-regulatory damping of expression among sets of co-expressed regulatory genes.

These effects are not easily captured by a predictive gene regulatory network model formulation with purely Boolean logic, even though the relationships can be laid out in such a topological format (Longabaugh et al. 2017), because the impacts of one regulator on a target node can depend on the level of that regulator at a given time relative to the level of another regulator. This forces effects like protein half-life regulation and cell cycle time (Kueh et al. 2013) to become relevant for predictiveness.

7. History and “memory”: contributions of chromatin state to regulatory dynamics

Underlying the gene regulatory network models of early embryo development is an assumption that gene expression should be a fairly immediate response to the “regulatory state” as defined by the transcription factors that are currently present in the cell. The actual time scale defined as “fairly immediate” is variable depending on features of different embryos, for example the temperature at which the organism normally develops. Response times (“step times” from transcription factor gene activation to target gene activation) can range from ~15 min in the case of the fruit fly, *Drosophila melanogaster* (Clark and Akam 2016), to ~3 hr in the case of the sea urchin, *Strongylocentrotus purpuratus* (Bolouri and Davidson 2003; Peter, Faure, and Davidson 2012). However, the validation of a network architecture is defined by the ability to account for changes in target gene expression from one time point to the next in terms of changes in the transcription factors that are present at those times.

Development of hematopoietic cells from pluripotent precursors in mammals is shockingly slow in this context, requiring days to make transitions from one step to the next recognizable step. This could have different explanations. The step times could in fact be shorter, but hard to recognize if the right indicator genes were not being followed. Alternatively, the cross-regulatory damping influences among different regulators could keep the cells in a kind of gene network “Brownian motion” with only occasional, random cells escaping into forward developmental progression. However, a third possibility is that the initial state of the cells is fundamentally different if they emerge from a dedicated stem-cell progenitor than if they emerge from a fertilized egg. The hematopoietic stem cell is known to be in itself a specialized cell type that is only produced at a relatively late stage in

the embryo, not a naïve cell type set aside from the earliest stages of development (Dzierzak and Speck 2008; Boïers et al. 2013). It is prevented from differentiating while the rest of the embryo differentiates. Thus, it is possible that the stem cell history of late-fetal and postnatal hematopoietic cells interposes the task of dismantling the stem cell state before differentiation can be unleashed.

This kind of generic possibility could be reflected in chromatin state modifications, for example, the methylation of DNA and/or the compaction or polycomb repression of chromatin at key regulatory sites for differentiation genes. However, even if hematopoietic stem cells could be shown to have more “repressive marks” on their chromatin than early cleavage-stage embryonic cells, that would not in itself predict quantitatively how much of a delaying effect such repression mechanisms would be expected to have on differentiation. It is clear that “epigenetic marks” like this can be removed as well as installed at specific genomic sites as a result of transcription factor action (e.g. Ji et al. 2010; Zhang et al. 2012; rev. in Rothenberg 2013). Could a cell’s history actually affect gene network dynamics?

Evidence to support this interpretation has come from a close analysis of the activation of a regulatory gene, *Bcl11b*, that occurs precisely as pro-T cells undergo commitment (Kueh et al. 2016). The Bcl11b gene product is needed for the commitment process to occur normally (Ikawa et al. 2010; Li, Leid, et al. 2010; Li, Burke, et al. 2010), and its expression is delayed for several days after Notch signaling induces expression of other regulators, including GATA-3 and another T-cell-associated factor, TCF-1. By causing the *Bcl11b* gene to drive the expression of a fluorescent protein as well as the normal Bcl11b protein, it has been possible to use live-cell imaging to examine what it takes for individual cells to cross the threshold to activate this gene. Once expressed in a given cell, the *Bcl11b* gene is overwhelmingly likely to remain expressed, and its expression is found to depend on at least four positive regulatory inputs: Notch signaling, the T-cell factors GATA-3 and TCF-1, and a factor with gently increasing expression from the stem cell stage, Runx1/CBFβ (Kueh et al. 2016). Knocking down any of these factors blocks the ability of the cell to activate *Bcl11b*. However, the kinetics of Bcl11b gene activation do not follow the expression patterns of these inputs together, and nor do they follow the time window of sensitivity to loss of activity of these factors as a whole. All four factors are expressed days and multiple cell cycles before *Bcl11b* turns on, and if they are removed at that early stage, the activation of *Bcl11b* indeed is blocked. But if cells at the stage immediately prior to *Bcl11b* activation are deprived of these inputs, three of four are partially or completely dispensable for the response, including Notch signaling and the two most T-cell specific inputs. Instead, these factors appear to be involved in a licensing process that slowly makes the *Bcl11b* gene eligible for expression. The factor that is immediately engaged in controlling *Bcl11b* expression and expression amplitude, both during activation and long afterwards in mature T cells, is none of these lineage-specific inputs. Instead, it is the multilineage, ancestrally inherited factor, Runx1/CBFβ (Kueh et al. 2016).

The key to the delay is that *Bcl11b* in stem cells and in the earliest T-cell precursors is completely repressed, with its DNA methylated and its chromatin marked with repressive histone marks. A complex process involving a far-distal enhancer, nearly a million base pairs away in the genome (Li et al. 2013), is required to activate it, by removing these repressive

mechanisms in a way that only finishes at the time the gene is activated (Li et al. 2013). Presumably for this reason, the Runx1/CBFβ factor, although present from the start, apparently cannot work on the *Bcl11b* locus until after the licensing event (dependent on Notch, GATA-3, and TCF-1) and a slow gene-accessibility process have become complete, over a period of ~4 days. Separate evidence for the nature of this slow process comes from further analysis using two different colored fluorescent reporters to track the two alleles in the same cells (K. K. H. Ng, M. A. Yui, S. S. Pease, S. Siu, A. Mehta, E. V. Rothenberg, H. Y. Kueh, unpublished). If the delay were only due to the lack of some other, as yet unidentified transcription factor, then both alleles would be activated in parallel. However, in fact they can be activated quite asynchronously in many of the cells. This shows that the slow step is of a kind that can work on each chromosome separately, even in the same cell where both chromosomes are exposed to the same transcription factors. The need for this slow process in the *Bcl11b* case is thus a measure of the resilience of the silencing mechanism inherited from the stem cell ancestor that has to be undone in order for the cells to complete T-cell commitment. This is quite different from the behavior of differentiation genes in an embryonic cell undergoing initial specification in most well-studied systems.

8. Operating principles of a T-cell specification network

The distinctive features of regulatory gene product interactions with other regulatory genes and with the chromatin state of key genomic loci in mammalian hematopoiesis do not supplant central features of gene network architecture, including the roles of combinatorial action of transcription factors on key targets. However, they change the relationship between the activation of an “upstream regulator” gene and a “downstream target” gene, in temporal response and in conditionality upon dosage. We suggest that part of the delay, at least, is due to the pre-existing chromatin states of genes in the dedicated stem and progenitor cell populations. Opening or closing a gene can require transcription factor actions on multiple different genomic regions that gain or lose accessibility through a slow process. The system makes use of this delay to enable cells to proliferate extensively, even after key transcription factors have been activated, before completing commitment to a particular developmental pathway. Because of the dosage control, the same transcription factor can participate in multiple different programs, where its effect is modulated by levels of other factors. Most importantly, in this system the repression of one set of factors by another often occurs only at high doses of the repressing factor, either due to biochemistry of factor/site interaction or due to gene network circuitry. Thus, in these early T precursors, coexpression of two factors that are supposed to be mutual antagonists, like PU.1 and GATA factors (Huang et al. 2007; Nerlov et al. 2000; Rekhtman et al. 1999; Zhang et al. 1999), can persist through many cell cycles before their “opposition” is manifest.[1] The de-emphasis on silencing-type repression significantly increases the number of transcription factors that are able to work combinatorially to define distinct developmental stages and/or developmental pathways.

[1]The relationship between GATA-3 and PU.1 in early T cell precursors resembles the relationship between GATA-2 and PU.1 in mast cells, and contrasts with the early programming for divergent regulation of GATA-1 and PU.1 seen in other multipotent progenitor subsets (Hoppe et al. 2016)

This set of properties fits the job that mammalian hematopoiesis has evolved to do. This is a system in which a dazzling variety of cell types is generated. The T-cell subtypes alone run to at least a dozen in mice, and their close cousins the Innate Lymphoid Cells have only recently begun to have their own diversity revealed. Dendritic cells, macrophages, and a whole range of cell types that share key regulators in violation of a “myeloid” vs. “erythroid” dichotomy are being discovered to have comparable diversity that is important physiologically. The diversity of possible outcomes is matched by a relaxation of temporal constraints and population size constraints, both of which are “outsourced” to physiological cues from the body’s environment. Prolonged multipotentiality, indeterminacy of commitment timing, and flexible use and re-use of the same transcription factors in divergent but related programs even at the expense of crispness of specification, thus become virtues in this system.

Concluding Remarks

Development is the most obvious output of genomic regulatory network operation, yet the analytical and predictive models that capture essential features of these networks need to be considered separately for different modes of development. The importance of quantitative effects on target gene expression as opposed to qualitative ones, the precision of timing of state changes, and the nature of negative regulation can all differ considerably between biological systems. Here we have focused on two developmental systems that have evolved to optimize different desiderata: the sea urchin embryo system and the mammalian postnatal hematopoietic system. The sea urchin embryo is highly precise and accurate about both cell fate determination and timing of differentiation of every cell type in the embryo, and has evolved to accomplish this by relative insensitivity to absolute levels of expression of key regulators. The mechanisms that make this possible are embedded in the prevalence of positive feed-forward circuits in its network architecture and in the all-or-none form that repression takes in this system. In contrast, the mammalian hematopoietic system has evolved to emphasize plasticity and environmental responsiveness at the expense of precision in both timing and cell fate determinism. The networks involved in this system are remarkably sensitive to the levels of different regulators, slow to resolve into end states, and permissive of coexpression of mutually antagonistic regulators as long as neither one is too highly expressed. These features arise not only from a strong role of chromatin structural constraint but also from incoherent feedback architecture of genetic network circuit elements in hematopoiesis and a prevalence of incomplete, amplitude-damping type repression. In this review, these contrasts have been used to illustrate both the richness of possible gene network behaviors, and the challenge to biologists to understand the operation of developmental networks in terms of models that capture the operational essences of the individual biological systems.

Acknowledgments

I thank members of my group for the use of our unpublished data, and the late Eric H. Davidson for years of inspiring and challenging discussions about gene regulatory logic, gene network inference, and network operating principles. Gene network research in the Rothenberg group was supported by grants from the USPHS, especially R33 HL89123, RC2 CA148278, R01 AI95943, and R01 HL119102. I gratefully acknowledge support from the Albert Billings Ruddock Professorship at Caltech.

References

Boïers C, et al. Lymphomyeloid contribution of an immune-restricted progenitor emerging prior to definitive hematopoietic stem cells. Cell Stem Cell. 2013; 13(5):535–548. [PubMed: 24054998]

Bolouri H, Davidson EH. Transcriptional regulatory cascades in development: initial rates, not steady state, determine network kinetics. Proceedings of the National Academy of Sciences of the United States of America. 2003; 100(16):9371–9376. [PubMed: 12883007]

Bonzanni N, et al. Hard-wired heterogeneity in blood stem cells revealed using a dynamic regulatory network model. Bioinformatics. 2013; 29(13):i80–8. [PubMed: 23813012]

Briscoe J, Small S. Morphogen rules: design principles of gradient-mediated embryo patterning. Development. 2015; 142(23):3996–4009. [PubMed: 26628090]

Britten RJ, Davidson EH. Gene regulation for higher cells: a theory. Science. 1969; 165(3891):349–357. [PubMed: 5789433]

Britten RJ, Davidson EH. Repetitive and non-repetitive DNA sequences and a speculation on the origins of evolutionary novelty. Quarterly Review of Biology. 1971; 46(2):111–138. [PubMed: 5160087]

Brzezinski JA, Reh TA. Photoreceptor cell fate specification in vertebrates. Development. 2015; 142(19):3263–3273. [PubMed: 26443631]

Champhekar A, et al. Regulation of early T-lineage gene expression and developmental progression by the progenitor cell transcription factor PU.1. Genes & Development. 2015; 29(8):832–848. [PubMed: 25846797]

Chen H, et al. A system of repressor gradients spatially organizes the boundaries of Bicoid-dependent target genes. Cell. 2012; 149(3):618–629. [PubMed: 22541432]

Ciofani M, et al. A validated regulatory network for Th17 cell specification. Cell. 2012; 151(2):289–303. [PubMed: 23021777]

Clark E, Akam M. Odd-paired controls frequency doubling in Drosophila segmentation by altering the pair-rule gene regulatory network. eLife. 2016; 5:e18215. [PubMed: 27525481]

Clark MR, et al. Orchestrating B cell lymphopoiesis through interplay of IL-7 receptor and pre-B cell receptor signalling. Nature Reviews Immunology. 2014; 14(2):69–80.

Davidson EH. Spatial mechanisms of gene regulation in metazoan embryos. Development. 1991; 113(1):1–26. [PubMed: 1684931]

Davidson, EH. The Regulatory Genome: Gene Regulatory Networks in Development and Evolution. San Diego: Academic Press; 2006.

Davidson EH. Emerging properties of animal gene regulatory networks. Nature. 2010; 468(7326):911–920. [PubMed: 21164479]

Davidson EH, Britten RJ. Organization, transcription, and regulation in the animal genome. Quarterly Review of Biology. 1973; 48(4):565–613. [PubMed: 4592421]

Davidson EH, Rast JP, Oliveri P, Ransick A, Calestani C, Yuh CH, Minokawa T, Amore G, Hinman V, Arenas-Mena C, Otim O, Brown CT, Livi CB, Lee PY, Revilla R, Rust AG, et al. A genomic regulatory network for development. Science. 2002a; 295(5560):1669–1678. [PubMed: 11872831]

Davidson EH, Rast JP, Oliveri P, Ransick A, Calestani C, Yuh CH, Minokawa T, Amore G, Hinman V, Arenas-Mena C, Otim O, Brown CT, Livi CB, Lee PY, Revilla R, Schilstra MJ, et al. A provisional regulatory gene network for specification of endomesoderm in the sea urchin embryo. Developmental Biology. 2002b; 246(1):162–190. [PubMed: 12027441]

Del Real MM, Rothenberg EV. Architecture of a lymphomyeloid developmental switch controlled by PU.1, Notch and Gata3. Development. 2013; 140(6):1207–1219. [PubMed: 23444353]

Dzierzak E, Speck NA. Of lineage and legacy: the development of mammalian hematopoietic stem cells. Nature Immunology. 2008; 9(2):129–136. [PubMed: 18204427]

Erwin DH, Davidson EH. The evolution of hierarchical gene regulatory networks. Nature Reviews Genetics. 2009; 10(2):141–148.

Hobert O. Development of left/right asymmetry in the Caenorhabditis elegans nervous system: from zygote to postmitotic neuron. Genesis. 2014; 52(6):528–543. [PubMed: 24510690]

Hodges E, et al. Directional DNA methylation changes and complex intermediate states accompany lineage specificity in the adult hematopoietic compartment. Molecular Cell. 2011; 44(1):17–28. [PubMed: 21924933]

Hoppe PS, et al. Early myeloid lineage choice is not initiated by random PU.1 to GATA1 protein ratios. Nature. 2016; 535(7611):299–302. [PubMed: 27411635]

Huang S, et al. Bifurcation dynamics in lineage-commitment in bipotent progenitor cells. Developmental Biology. 2007; 305(2):695–713. [PubMed: 17412320]

Hughes JR, et al. Analysis of hundreds of cis-regulatory landscapes at high resolution in a single, high-throughput experiment. Nature Genetics. 2014; 46(2):205–212. [PubMed: 24413732]

Ikawa T, et al. An essential developmental checkpoint for production of the T cell lineage. Science. 2010; 329(5987):93–96. [PubMed: 20595615]

Istrail S, De-Leon SB, Davidson EH. The regulatory genome and the computer. Developmental Biology. 2007; 310(2):187–195. [PubMed: 17822690]

Jaeger J. The gap gene network. Cellular and Molecular Life Sciences : CMLS. 2011; 68(2):243–274. [PubMed: 20927566]

Ji H, et al. Comprehensive methylome map of lineage commitment from haematopoietic progenitors. Nature. 2010; 467(7313):338–342. [PubMed: 20720541]

Kueh HY, et al. Positive feedback between PU.1 and the cell cycle controls myeloid differentiation. Science. 2013; 341(6146):670–673. [PubMed: 23868921]

Kueh HY, Rothenberg EV. Regulatory gene network circuits underlying T cell development from multipotent progenitors. Wiley Interdisciplinary Reviews in Systems Biology and Medicine. 2012; 4(1):79–102. [PubMed: 21976153]

Kueh HY, et al. Asynchronous combinatorial action of four regulatory factors activates *Bcl11b* for T cell commitment. Nature Immunology. 2016; 17(8):956–965. [PubMed: 27376470]

Laslo P, et al. Gene regulatory networks directing myeloid and lymphoid cell fates within the immune system. Seminars in Immunology. 2008; 20(4):228–235. [PubMed: 18771937]

Letting DL, et al. Context-dependent regulation of GATA-1 by friend of GATA-1. Proceedings of the National Academy of Sciences of the United States of America. 2004; 101(2):476–481. [PubMed: 14695898]

Li L, Leid M, Rothenberg EV. An early T cell lineage commitment checkpoint dependent on the transcription factor Bcl11b. Science. 2010; 329(5987):89–93. [PubMed: 20595614]

Li L, et al. A far downstream enhancer for murine Bcl11b controls its T-cell specific expression. Blood. 2013; 122(6):902–911. [PubMed: 23741008]

Li P, et al. Reprogramming of T cells to natural killer-like cells upon Bcl11b deletion. Science. 2010; 329(5987):85–89. [PubMed: 20538915]

Longabaugh, WJR., et al. Bcl11b and combinatorial resolution of cell fate in the T-cell gene regulatory network. Proceedings of the National Academy of Sciences of the United States of America; 2017. p. 114in press

Lu M, et al. The earliest thymic progenitors in adults are restricted to T, NK, and dendritic cell lineage and have a potential to form more diverse TCRβ chains than fetal progenitors. Journal of Immunology. 2005; 175(9):5848–5856.

Maier H, et al. Early B cell factor cooperates with Runx1 and mediates epigenetic changes associated with mb-1 transcription. Nature Immunology. 2004; 5(10):1069–1077. [PubMed: 15361869]

Manesso E, et al. Computational modelling of T-cell formation kinetics: output regulated by initial proliferation-linked deferral of developmental competence. Journal of the Royal Society Interface. 2013; 10(78):20120774.

Markstein M, et al. A regulatory code for neurogenic gene expression in the Drosophila embryo. Development. 2004; 131(10):2387–94. [PubMed: 15128669]

May G, et al. Dynamic analysis of gene expression and genome-wide transcription factor binding during lineage specification of multipotent progenitors. Cell Stem Cell. 2013; 13(6):754–768. [PubMed: 24120743]

Meulemans D, Bronner-Fraser M. Gene-regulatory interactions in neural crest evolution and development. Developmental Cell. 2004; 7(3):291–299. [PubMed: 15363405]

Miccio A, et al. NuRD mediates activating and repressive functions of GATA-1 and FOG-1 during blood development. EMBO Journal. 2010; 29(2):442–456. [PubMed: 19927129]

Naik SH, et al. Diverse and heritable lineage imprinting of early haematopoietic progenitors. Nature. 2013; 496(7444):229–232. [PubMed: 23552896]

Nerlov C, et al. GATA-1 interacts with the myeloid PU.1 transcription factor and represses PU.1-dependent transcription. Blood. 2000; 95:2543–2551. [PubMed: 10753833]

Nikitina N, Sauka-Spengler T, Bronner-Fraser M. Dissecting early regulatory relationships in the lamprey neural crest gene network. Proceedings of the National Academy of Sciences of the United States of America. 2008; 105(51):20083–20088. [PubMed: 19104059]

Novershtern N, et al. Densely interconnected transcriptional circuits control cell states in human hematopoiesis. Cell. 2011; 144(2):296–309. [PubMed: 21241896]

Oguro H, et al. Poised lineage specification in multipotential hematopoietic stem and progenitor cells by the polycomb protein Bmi1. Cell Stem Cell. 2010; 6(3):279–286. [PubMed: 20207230]

Oliveri P, Tu Q, Davidson EH. Global regulatory logic for specification of an embryonic cell lineage. Proceedings of the National Academy of Sciences of the United States of America. 2008; 105(16): 5955–5962. [PubMed: 18413610]

Perié L, et al. The branching point in erythro-myeloid differentiation. Cell. 2015; 163(7):1655–1662. [PubMed: 26687356]

Peter IS, Davidson EH. Evolution of gene regulatory networks controlling body plan development. Cell. 2011a; 144(6):970–985. [PubMed: 21414487]

Peter IS, Davidson EH. A gene regulatory network controlling the embryonic specification of endoderm. Nature. 2011b; 474(7353):635–639. [PubMed: 21623371]

Peter, IS., Davidson, EH. Genomic Control Process: Development and Evolution. San Diego, CA: Academic Press, Elsevier; 2015.

Peter IS, Faure E, Davidson EH. Predictive computation of genomic logic processing functions in embryonic development. Proceedings of the National Academy of Sciences of the United States of America. 2012; 109(41):16434–16442. [PubMed: 22927416]

Pietersen AM, van Lohuizen M. Stem cell regulation by polycomb repressors: postponing commitment. Current Opinion in Cell Biology. 2008; 20(2):201–207. [PubMed: 18291635]

Reeves GT, et al. Dorsal-ventral gene expression in the Drosophila embryo reflects the dynamics and precision of the dorsal nuclear gradient. Developmental Cell. 2012; 22(3):544–557. [PubMed: 22342544]

Rekhtman N, et al. Direct interaction of hematopoietic transcription factors PU.1 and GATA-1: functional antagonism in erythroid cells. Genes & Development. 1999; 13:1398–1411. [PubMed: 10364157]

Rothenberg EV. Epigenetic mechanisms and developmental choice hierarchies in T-lymphocyte development. Briefings in Functional Genomics. 2013; 12(6):512–524. [PubMed: 23922132]

Rothenberg EV, Ungerback J, Champhekar A. Forging T-Lymphocyte Identity: Intersecting Networks of Transcriptional Control. Advances in Immunology. 2016; 129:109–174. [PubMed: 26791859]

Schutte J, et al. An experimentally validated network of nine haematopoietic transcription factors reveals mechanisms of cell state stability. eLife. 2016; 5:e11469. [PubMed: 26901438]

Scripture-Adams DD, et al. GATA-3 dose-dependent checkpoints in early T cell commitment. Journal of Immunology. 2014; 193(7):3470–3491.

Simoes-Costa M, Bronner ME. Establishing neural crest identity: a gene regulatory recipe. Development. 2015; 142(2):242–257. [PubMed: 25564621]

Stanojevic D, Small S, Levine M. Regulation of a segmentation stripe by overlapping activators and repressors in the Drosophila embryo. Science. 1991; 254:1385–1387. [PubMed: 1683715]

Stathopoulos A, Levine M. Genomic regulatory networks and animal development. Developmental Cell. 2005a; 9(4):449–462. [PubMed: 16198288]

Stathopoulos A, Levine M. Localized repressors delineate the neurogenic ectoderm in the early Drosophila embryo. Developmental Biology. 2005b; 280(2):482–493. [PubMed: 15882587]

Taghon T, Yui MA, Rothenberg EV. Mast cell lineage diversion of T lineage precursors by the essential T cell transcription factor GATA-3. Nature Immunology. 2007; 8(8):845–855. [PubMed: 17603486]

Tagoh H, et al. Dynamic reorganization of chromatin structure and selective DNA demethylation prior to stable enhancer complex formation during differentiation of primary hematopoietic cells in vitro. Blood. 2004; 103(8):2950–2955. [PubMed: 15070670]

Viets K, Eldred KC, Johnston RJ Jr. Mechanisms of Photoreceptor Patterning in Vertebrates and Invertebrates. Trends in Genetics : TIG. 2016; 32(10):638–659. [PubMed: 27615122]

Xu W, et al. E2A transcription factors limit expression of Gata3 to facilitate T lymphocyte lineage commitment. Blood. 2013; 121(9):1534–1542. [PubMed: 23297135]

Yashiro-Ohtani Y, et al. Pre-TCR signaling inactivates Notch1 transcription by antagonizing E2A. Genes & Development. 2009; 23(14):1665–1676. [PubMed: 19605688]

Yuh CH, Bolouri H, Davidson EH. Genomic cis-regulatory logic: experimental and computational analysis of a sea urchin gene. Science. 1998; 279(5358):1896–1902. [PubMed: 9506933]

Yui MA, Rothenberg EV. Developmental gene networks: a triathlon on the course to T cell identity. Nature Reviews Immunology. 2014; 14(8):529–545.

Zhang JA, et al. Dynamic transformations of genome-wide epigenetic marking and transcriptional control establish T cell identity. Cell. 2012; 149(2):467–482. [PubMed: 22500808]

Zhang P, et al. Negative cross-talk between hematopoietic regulators: GATA proteins repress PU.1. Proceedings of the National Academy of Sciences of the United States of America. 1999; 96:8705–8710. [PubMed: 10411939]

Zhang Y, et al. Chromatin connectivity maps reveal dynamic promoter-enhancer long-range associations. Nature. 2013; 504(7479):306–310. [PubMed: 24213634]