Causal Gene Regulatory Network Modeling and Genomics: Second-Generation Challenges

ELLEN V. ROTHENBERG

ABSTRACT

Gene regulatory network modeling has played a major role in advancing the understanding of developmental systems, by crystallizing structures of relevant extant information, by formally posing hypothetical functional relationships between network elements, and by offering clear predictive tests to improve understanding of the mechanisms driving developmental progression. Both ordinary differential equation (ODE)-based and Boolean models have also been highly successful in explaining dynamics within subcircuits of more complex processes. In a very small number of cases, gene regulatory network models of much more global scope have been proposed that successfully predict the dynamics of the processes establishing most of an embryonic body plan. Can such successes be expanded to very different developmental systems, including post-embryonic mammalian systems? This perspective discusses several problems that must be solved in more quantitative and predictive theoretical terms, to make this possible. These problems include: the effects of cellular history on chromatin state and how these affect gene accessibility; the dose dependence of activities of many transcription factors (a problem for Boolean models); stochasticity of some transcriptional outputs (a problem for simple ODE models); response timing delays due to epigenetic remodeling requirements; functionally different kinds of repression; and the regulatory syntax that governs responses of genes with multiple enhancers.

Keywords: developmental kinetics; developmental regulation; epigenetic constraint; gene silencing and de-repression; modeling causality.

1. INTRODUCTION

1.1. Development: A System of Nonlinearities and Irreversibilities

DEVELOPMENTAL BIOLOGY EXPLAINS the irreversible increases in complexity and specialization of distinct cell types within an organism. It encompasses changes in gene expression, gene expression potential, and structural morphology in the organism. Developmental changes are distinguished from physiological response changes by their irreversibility. Although the molecules that control specific transitions in development can have their own expression cease after the transition they regulate, the changes that they have caused to the

system are not reversed when this occurs: hit and run regulation is common. Thus, development is an emergent property of a system that, if not completely irreversible, at least exhibits strong hysteresis.* As a system property, too, the impacts of individual regulators can be highly nonlinear, so that sharp cell fate boundaries emerge at particular thresholds within continuous-valued spatial or temporal gradients of regulator activity. These features are notable because the essential biochemistry of gene expression is normally understood in terms of processes that are intrinsically reversible and dynamic, such as transcription factor binding equilibria to particular DNA sites and the need to re-recruit RNA polymerase to a promoter for synthesis of each mRNA molecule from a given gene (Bolouri and Davidson, 2003; Bintu et al., 2005; Ben Tabou de-Leon and Davidson, 2009). Therefore, any explanation for development has to account explicitly for its irreversibility. In fact, as discussed later, these features emerge from the architecture of developmental gene regulatory networks.

The mechanism of development is based on the ordered activation of gene regulatory networks, in which initial regulators and anisotropically organized intercellular signals turn on cascades of other regulators (Davidson, 2010; Peter and Davidson, 2015). These regulatory inputs operate combinatorially and can activate expression of new signaling molecules as well as new transcription factors. Thus, as the developing cells undergo cycle after cycle of mitosis, new cellular interfaces are formed across which asymmetric signals can extend and further diversify the impacts of the initial signaling asymmetries. Importantly, gene regulatory networks conditionally mobilize specific regulatory antagonists as well as regulatory collaborators, sharpening gene expression differences between cells in different states. This process leads to a progressive organism-wide increase in complexity. Like a kind of thermodynamic entropy term, for the organism as a whole this increase in complexity should contribute to the irreversibility of the process.

1.2. Gene regulatory network models

Gene regulatory network theory, based on these principles, has been extremely successful in explaining early embryonic development, when the initial cleavage cells (or nuclei, in the cases of certain insects) begin to be assigned to different fates. With regard to hysteresis, though, it is worth noting that this is also a very dynamic period in terms of gene expression, and that genes with crucial roles in patterning are turned on in particular locations but may only be expressed transiently. Thus, although the organism is undergoing irreversible change at these stages, individual genes are not necessarily being activated or repressed irreversibly. This kind of regulated gene expression fits well with a textbook biochemical understanding of gene expression, which should make gene regulation a Markov process, that is, one in which history does not matter, and the only factors determining gene expression responses at a given time are those that are present at that time. In a gene regulatory network operating as a chain of Markov processes, the genes expressed at one stage should provide all the regulators needed to explain gene expression at the next stage. For developmental systems with this property, it has been possible to derive formal models for the operation of the underlying gene networks that have high explanatory and predictive power (Balaskas et al., 2012; Peter et al., 2012; Cohen et al., 2014; Mbodj et al., 2016; Clark, 2017).

Several kinds of intellectual constructs are referred to as gene network models, confusingly sharing the name although they seek to achieve different goals. Here, we can distinguish (1) correlation-based "omics" models, whether based on simple clustering or Bayesian inference; (2) deterministic continuous-valued models of ordinary differential equations (ODEs); (3) deterministic Boolean process models of regulatory systems; and (4) mechanistically based "bottom-up" gene network models, which emphasize empirical measurements of protein-DNA interaction. There is a large literature on the strategy, requirements, strengths, and weaknesses of such different types of gene network models, which extend far beyond the scope of this perspective (Bolouri and Davidson, 2002, 2003; Jaeger et al., 2004; Ben Tabou de-Leon and Davidson, 2009; Shmulevich and Aitchison, 2009; Ay and Arnosti, 2011; Wunderlich and DePace, 2011; Peter et al., 2012; Wilczynski et al., 2012; Albert and Thakar, 2014; Briscoe and Small, 2015; Abou-Jaoudé et al., 2016; Mbodj et al., 2016; Olariu and Peterson, 2019). One large division between these types of models is the purpose of

---

*"Hysteresis" describes input-output relationships that depend on a system's history. A typical case is the one alluded to, where some regulator has to increase to a certain threshold value to cause a system to shift to a new state, but then the system can remain in the new state and not return to the previous state even if the regulator drops again below the original threshold value.

the model: whether it is to classify cells on the basis of their different current gene expression patterns (type 1), or whether it is to explain why a dynamic developmental process generates the outputs that it does (types 2, 3, and 4). In this review, the discussion concerns models formulated to explain emergent developmental outcomes on the basis of verifiable causal relationships, to explain the order, conditionality, and timing of different gene expression steps. This is the province of models types 2, 3, and 4; therefore, ''type 1'' models are not discussed. Models formalized by type 2 or type 3 approaches have different strengths and weaknesses. Where it is possible to measure reactant concentrations and rate constants for interactions, ODE models are attractive, but the number of distinct network elements they can handle is normally limited. Where logic relationships can be established clearly, Boolean models are advantageous, but they force inputs as well as outputs to be scored in binary terms. In principle, both ODE-based and Boolean approaches can be used either with black box elements or in versions that draw on significant mechanistic detail, as in type 4.

Multiple modeling approaches are thus available for understanding gene network operations. Why are there still problems with applying these to explain and predict all kinds of development? To address this question, many of the following examples are drawn from mammalian hematopoiesis, a particularly well-studied system of cells developing from long-lasting, self-renewing stem cells rather than from fertilized eggs.

Hematopoiesis encompasses the generation of all types of blood cells, including red blood cells, inflammatory cells, blood-clotting platelets, and classic cells of the immune system. Although hematopoiesis starts in the embryo, it continues throughout life, as a post-embryonic developmental system that is naturally sustained by multipotent stem cells. This system has been highly favorable for mechanistic dissection, since fine-grained differences among cell states in different lineages and stages can be distinguished at the single-cell level in this system, and each of many different specific types of cells can not only be purified but also be interrogated for developmental potential, subjected to deep molecular characterization, and manipulated in vitro. Importantly, in hematopoiesis, the differentiation process works to make decisions that are irreversible for individual cells, cells that must migrate repeatedly throughout the body, and whose identities will be preserved independently of their neighbors. Thus, once they have taken a particular differentiation path, irreversibility of their fates must be accounted for by intrinsic mechanisms in the individual cells, which maintain their distinct states without reference to the overall increase in system complexity.

The gene regulatory mechanisms already identified in this system have been particularly revealing about issues that are challenging to treat predictively via gene regulatory network models; these mechanisms, however, are likely to operate much more broadly and are worth a closer look.

2. THE PROBLEMS

2.1. Chromatin states and cellular histories

The modeling approaches just noted all provide useful perspectives on how development works. However, both ODE and Boolean models alike are based on the history-free assumption that gene expression should be regulated only by currently available regulatory factors. Specifically, by the assumptions of all of these modeling types, the genes expressed—and the genes not expressed—when the cells commit to a fate should be explicable purely in terms of the factors that are and are not expressed in the cells at that time, working in combinatorial logic. Although this assumption may generally apply in prokaryotic systems, in early embryos, and even in acute inflammation responses of multicellular organisms, it is probably not the rule in the later development of cell types in multicellular organisms, when the cells shift from a state of dynamically changing gene expression to a state of stable, lineage-defining gene expression. Instead, as discussed later, the transcription factors important for maintenance of a particular pattern of gene expression in differentiated cells can fail to include all the factors that have made expression of those genes possible or impossible within those cells, and are thus insufficient to explain the difference between expression and nonexpression in these ''cells with a history.''

The most evident manifestation of this history is the way it installs differential physical accessibility of different regions of genomic DNA within the cell, and makes these differences heritable among all the cells' descendants. Cell type differentiation at the end of a developmental process is clearly influenced by the way these pre-imposed chromatin state differences affect cells' ranges of gene regulatory options (Spitz and Furlong, 2012; Chen and Dent, 2014). Such persistent chromatin states affect local thresholds for

transcription factor activity and alter the quantitative and temporal couplings between transcription factor (TF) binding and target gene responses. The imposition of this layer of regulation creates significant problems for predictive gene regulatory network modeling.

Chromatin states are often called ''epigenetic,'' although this term is not precise. It is taken to include biochemically diverse events embracing DNA methylation, intranuclear localization, post-translational histone protein modification, long non-coding RNA expression, and various measures of histone compaction or decompaction as measured by nuclease accessibility (Kouzarides, 2007; Black et al., 2012; Bock et al., 2012; Hardison and Taylor, 2012; Sashida and Iwama, 2012; Calo and Wysocka, 2013; Mercer et al., 2013; Schwartz and Pirrotta, 2013). As emphasized in an extensive literature beyond the references cited, differences in epigenetic indices, especially in chromatin modifications and DNA accessibility in chromatin at hundreds of thousands of genomic sites, clearly distinguish between various mature differentiated cell types in a very robust way (Calo and Wysocka, 2013). The enzymatic machinery for these modifications is near-universally expressed in cells, but all of these modifications can be established or reversed locally at particular genomic sites as results of changes in local transcription factor activity, thus reflecting a specific cell state.

The excitement about ''epigenetic states'' relative to developmental gene network models comes from evidence that these features may also persist and constrain future transcription factor activity (Arvey et al., 2012; Spitz and Furlong, 2012; Zhang et al., 2012; Chen and Dent, 2014; Bintu et al., 2016; Kribelbauer et al., 2017; Liu et al., 2018; Ungerbäck et al., 2018). However, the key questions about each of these chromatin-modifying mechanisms, with respect to their contributions to developmental gene regulation, have to do with their rates of establishment and their reversibilities or resistances to being reversed, when transcription factor activities change. In practice, at least as measured in post-embryonic cells, chromatin state differences that are established during the cells' prior developmental histories, indeed, appear to mask large portions of the genome from recognition by individual transcription factors (discussed below). Because of these effects, the response of a target gene in the genome cannot be predicted only based on the current presence of the trans-acting transcription factors that normally bind to its regulatory sequences; instead, in ways that are still poorly quantified, the inherited chromatin states of its regulatory sequences matter as well.

The ability of transcription factor binding to trigger extensive chromatin changes has been studied intensively in the context of reprogramming differentiated cells to induced pluripotent cells, by forcing them to co-express the ''Yamanaka factors,'' transcription factors Oct4, Sox2, Myc, and Klf4 (Ho et al., 2011; Watanabe et al., 2013; Takahashi and Yamanaka, 2016; Li et al., 2017; Sardina et al., 2018). Operationally, there is evidence that certain chromatin states in the initial condition exert a restraint against the speed or completeness of reprogramming (Sridharan et al., 2013; Watanabe et al., 2013; Roost et al., 2017). However, over days or weeks of combined Oct4, Sox2, Myc, and Klf4 expression, chromatin accessibility changes, histone marks change, and DNA methylation changes at many sites as a result of the powerful collaboration of these factors.

In principle, then, it should be possible to model changes in chromatin modification as well as the changes in transcription factor binding in normal developmental transitions as well. However, the dynamics of these chromatin changes are complex, some are affected by cell cycle, and the likelihood of changing the status of one nucleosome's marking or one DNA CpG dinucleotide's methylation depends on the current status of the nucleosome or DNA sequence immediately next to it. When chromatin changes spread slowly across a region of the genome, it is not clear whether this reflects a strictly processive mechanism moving along the chromosome, or the stochastic advance of a frontier at which independent gain and loss of modification events compete (Jeong et al., 2014). Thus, although epigenetic states readily distinguish among various differentiated cell states, epigenetic marking dynamics and the impacts of different regulatory inputs on them are formidably difficult to model predictively over large regions of the genome.

2.2. Non-Boolean inputs

Hematopoietic cell fates are determined by overlaps among about a score of potent transcription factors, each of which may be necessary for a group of cell fates and/or exclusionary for another group of cell fates. They cross-regulate each other, causing certain combinations to stabilize and other combinations to be extinguished. Some of the combinations and oppositions have inspired highly influential ODE-based models of ''developmental switches,'' based on mutual repression between two regulators, intended to explain the mutual exclusivities of particular cell fates. One of these models, the GATA-1:

PU.1 antagonism, has been believed to explain why erythroid cells and megakaryocytes, using GATA-1, diverge in fate from myeloid cells, including macrophages and neutrophils, which use PU.1 (Swiers et al., 2006; Graf and Enver, 2009). By formulating the relationship as an ODE model with continuous valued regulator levels, the model purports to explain why two mutually exclusive branches of development emerge from a bipotent precursor that may transiently coexpress PU.1 and GATA-1 (or the related factor, GATA-2) (Huang et al., 2007).

This model has been very influential but has glaring shortcomings in relation to the details of the biology. Most importantly, it seems to ignore the generation of several lineages of cells that depend on both GATA factors and PU.1 permanently through numerous cell divisions, for example, mast cells. Examples such as mast cells (Walsh et al., 2002) show that the mutual repression of these factors cannot be hard-wired and must at least be conditional, with possible controlling influences coming from other factors in the cells (Chickarmane et al., 2009; Monteiro et al., 2011). This kind of network could, in principle, be accommodated by a Boolean model, in which the mutual repression actually depended on the other factors working in ''and'' logic with PU.1 and GATA-1 themselves.

However, Singh and coworkers early recognized another crucial feature of hematopoietic transcription factor action, namely its high dosage sensitivity (DeKoter and Singh, 2000). PU.1, in particular, has qualitatively different effects on gene expression when it is expressed at higher levels than when it is expressed at lower levels, at least in part because of its triggering different shifts in gene network states. Modest levels of both PU.1 and GATA-1 establish a stable *convivencia*, as in mast cells (Walsh et al., 2002), whereas higher levels of either one will tip the balance. In contrast to most versions of the mutual repression ODE model, this coexpression state is not unstable, but rather is highly sustainable in actual mast cells, which can be grown to large cell cultures in a seemingly unchanging developmental state. In practice, the ability of hematopoietic lineage-determining factors to collaborate in a wide range of non-exclusive combinations is used, at least in mammals, to diversify hematopoietic developmental pathways (Rothenberg et al., 2016), but many of these combinations may only be sustainable at moderate expression levels of the key factors.

Two additional kinds of experimental systems have shown possible switches that are activated differentially by high-dose and lower-dose PU.1, to yield qualitatively different PU.1-dependent fates. One switch controls the relative activities of Egr family transcription factors compared with the progenitor factor Gfi1 (Laslo et al., 2006), and the other switch controls the activity of cell cycle-promoting factors Myc and Myb (Kueh et al., 2013).

In the first of these switch models, a mutual repression of Egr factors and Gfi1 factor leave uncertainty as to whether a myeloid progenitor will activate a macrophage gene expression program or a neutrophil program (Laslo et al., 2006). Both of these programs depend on both PU.1 and another factor, C/EBPα, and so neither one can be excluded by Boolean relationships between these two factors; however, these fates are sensitive to relative doses of these transcription factors (Dahl et al., 2003). At higher effective doses of PU.1 than C/EBPα, Egr family factors are more highly activated than Gfi1, causing repression of Gfi1 and directing differentiation to the macrophage fate (Laslo et al., 2006). Another macrophage-promoting factor, Irf8 (Becker et al., 2012; Olsson et al., 2016), also antagonizes Gfi1 in this case. At higher effective doses of C/EBPα than PU.1, Gfi1 is expressed more highly than Egr factors, keeping Egr factors and Irf8 reduced, thus directing differentiation to the neutrophil fate instead (Laslo et al., 2006; Olsson et al., 2016). The ability to overcome an excess level of one factor with an excess level of the other suggests strongly that it is not the crossing of a particular concentration threshold for each factor in itself but the relative levels of two factors that determine the outcome. To be accommodated in a Boolean model type, then, at a minimum, two thresholds (e.g., ''present'' and ''high'') would have to applied for both PU.1 and C/EBPα, and some function would still have to be added to allow the levels of the two factors to be compared as well.

In the second dose-dependent switching system, PU.1 is shown to be able to induce a highly nonlinear developmental response by exploitation of its own high stability as a protein (Kueh et al., 2013), a strong contrast to C/EBPα that is very short-lived. If its levels rise beyond a certain threshold of expression, PU.1 can slow the cell cycle, in part through downregulation of the cell cycle genes Myc and Myb. Then, pre-existing PU.1 protein is more slowly diluted by cell division, PU.1 protein builds up in the cell even without an increase in new synthesis rate, cell cycle is slowed even further, PU.1 levels build up more, and the cells differentiate into macrophages (Kueh et al., 2013).

Both of these cases emphasize that the presence or absence of a factor such as PU.1 is not sufficient to specify which rule set it will follow in regulating its various target genes, except when there is additional

information about its level of activity. Importantly, the high-level expression effects are not simple ''overexpression toxicity,'' but rather a shift in developmental preference of the cell to a different, but coherent, cell fate. PU.1 binding to genomic sites in vivo appears to follow an orderly relationship of factor concentration to site-recognition affinity, and it even occupies sites with low accessibilities in chromatin if the site sequences are more optimal or if its own factor levels are higher (Ungerbäck et al., 2018). Thus, increasing levels of this factor should directly result in occupancy of an expanded set of genomic target sites, but a well-defined set, and these would be expected to account for the appearance of new responses to PU.1 when its levels rise above certain thresholds. Interestingly, the GATA family transcription factor that is famous for being required for T cell development, GATA-3, itself causes an analogous developmental effect when it is expressed in the T cell precursors at an elevated level; there, its high-level expression blocks the T cell fate and redirects the cells to a mast-cell developmental pathway (Taghon et al., 2007). Thus, the direction of precursors to different fates by the same factor at different levels is not unique to PU.1 and may be a property of multiple important cell lineage-determining regulatory factors.

Dose dependency per se does not necessarily depend on a chromatin configuration feature. As just summarized, it is likely to be a network property based on activation of different targets at higher factor doses that alter the expression or impacts targets activated at lower doses. However, the closely analyzed example of PU.1 indicates that a ''closed'' state of chromatin can also raise the (affinity $\times$ concentration) threshold required for binding (Ungerbäck et al., 2018), whereas histone modifications such as H3K27 trimethylation may exclude its local binding entirely (Zhang et al., 2012; Ungerbäck et al., 2018). Thus, a chromatin configuration inherited from a prior regulatory state is an effective way to change the dose range over which a newly expressed transcription factor can gain access to a particular subset of its potential target sites, as compared with other sites that would otherwise be recognized with the same affinity. Local chromatin state consequences of developmental history thus change the ''playing field'' over which transcription factor dose-response effects operate.

2.3. Non-continuous gene regulatory outputs

The combinatorial aspect of transcription factor action is believed to involve assembly of an ensemble of collaborating transcription factors, simultaneously bound to an important enhancer (cis-regulatory element) of the gene (Panne et al., 2007; Crocker et al., 2015; Farley et al., 2015; Khoueiry et al., 2017). The initiation of RNA synthesis that this triggers might be expected to reflect an initiation frequency distribution with a continuous-valued relationship between effective transcription factor binding and RNA output in a given cell. Measurements at the population level appear to agree with this expectation in most cases: Lower positive regulator inputs or mutationally weakened cis-regulatory DNA sequences often yield lower gene expression, as averaged over the population (Goncalves et al., 2012; Ulirsch et al., 2014; Wong et al., 2017). However, an important insight into gene regulation at the single-cell level is that transcription states of a gene are often stochastically distributed in a population of nominally equivalent cells (Raj and van Oudenaarden, 2008). Partly this is due to intermittent transcriptional ''bursting'' which occurs with frequencies that directly reflect enhancer activity (Fukaya et al., 2016). Recent evidence even demonstrates that the functional output from transcription factor binding to cis-regulatory elements can be discontinuous and probabilistic at the single-allele level (Ng et al., 2018). Although bursting can cause instantaneous transcription rates to fluctuate over time scales of minutes (Bothma et al., 2014), diversity in expression versus nonexpression of a given gene relative to its trans-regulators can also be sustained by some cells for much longer. This kind of regulation is problematic for the assumptions underlying standard, deterministic ODE models of gene expression.

Stochastic gene expression outputs may seem incompatible with a functional developmental gene network. Gene networks with highly interlinked positive and negative feedbacks are assumed to be able to rein in the developmental consequences of this kind of stochastic response of individual genes at the single-cell level. In development of tissues of uniform cell type, indeed, ''community effect'' paracrine signaling between each cell and its tissue neighbors plays a substantial role in homogenizing developmental programs to make cell fates converge (Gurdon, 1988; Bolouri and Davidson, 2010). However, in developmental systems such as hematopoiesis, cell fates are determined individually, and there is no requirement that each cell adopt either the same or a different fate as compared with its sister or neighbor. Net outputs of different cell types in hematopoiesis are controlled downstream of cell lineage specification, by differential proliferation and death in response to environmental signals. It is important to note, therefore, that the transcription factor combinations that specify cell fate in such systems need not always yield a single,

homogeneous gene expression profile in all developmental intermediates of a similar regulatory state to be effective in development.

Stochastic gene expression has only been recognized through single-cell examination. Single-cell measurements of RNA expression are possible not only by single-cell RNA-seq (deep reverse-transcriptase-dependent RNA sequencing), which gives a snapshot of the transcript abundances in a cell at one fixed moment in time, but also by fluorescent protein reporter knock-in genes, which allow cells to be distinguished based on their individual transcriptional activity levels from a locus of interest, integrated over hours (Rieger et al., 2009; Hoppe et al., 2016; Kueh et al., 2016). When the fluorescent protein reporter is relatively stable, like many derivatives of Green Fluorescent Protein, then any short-term noise in transcription rate is evened out, and any differences in fluorescence levels in different cells must represent cell-to-cell differences that persist over a period of hours or days. These assays all reveal that as the intensity of an activating signal or positive regulator expression is varied over a continuous range, cells may respond not by shifting all the target gene expression to a different level, but rather by altering the percentage of cells expressing at all, while keeping the level per expressing cell the same as when signal is maximal (Kueh et al., 2016). Similarly, genes that are used in acute immune responses, but held in a default ''off'' state in resting cells, have long been known to turn on only in a fraction of cells that varies according to the strength of the activating signal, and sometimes are expressed only monoallelically (Bix and Locksley, 1998; Holländer et al., 1998; Chiodetti et al., 2000). Importantly, stochastic activation control at an intermediate developmental stage need not average out over time; instead, it can lead to long-term mosaicism of gene expression in a population.

The long-term maintenance of these stochastic gene expression decisions probably requires heritable modifications of chromatin state. Recent evidence for the activation of the *Bcl11b* gene in early T cells shed light on several ways that epigenetic states affect developmental progression (Ng et al., 2018). The *Bcl11b* gene encodes a crucial regulator used throughout T cell function, but the locus is completely silent in all but one other hematopoietic cell type, and it is known exactly when the gene normally is activated for the first time in early T cell precursors (Liu et al., 2010; Zhang et al., 2012; Avram and Califano, 2014; Kueh et al., 2016; Longabaugh et al., 2017). A particular element within the large *Bcl11b* superenhancer complex was required for the earliest stage of its expression (Li et al., 2013; Ng et al., 2018): When this element was deleted in vivo, all developing T cells showed retarded activation of the mutant allele, although most of the cells recovered ~ 2 days later (Ng et al., 2018). The cells that recovered expression, presumably using other enhancer elements, went on to express *Bcl11b* as highly from the mutant allele as from the control allele. However, a subset of each cohort of mature T cells in mice with the mutant allele never activated the gene from that allele, even weeks after they should initially have turned the gene on (Ng et al., 2018).

The results imply that this cis-regulatory element was normally working to enhance the likelihood of all-or-none activation from a silent state within a particular developmental time window, with the choice made by a particular cell inherited stably by its descendants thereafter. The cells thus appeared to make a stochastic decision of whether to express a gene or not, with the input transcription factors and the cis-regulatory target sites primarily controlling the likelihood of activation in a given developmental context, rather than the amplitude of expression.

This strongly suggests that the cis-regulatory systems of genes not only integrate multiple inputs but also convert analog inputs into digital, stochastic ones. The ''Analogue-to-Digital'' conversion fits a Boolean model better than a simple ODE model. On the other hand, its probabilistic nature and the high likelihood that an ''equivalent'' starting population will actually become mosaic with respect to expression of the affected gene when regulatory input levels are manipulated are features not normally accommodated in standard Boolean models.

2.4. Timing

Gene network construction is usually based on the assumption that each node processes inputs and that this results immediately in a change in the output. In successful gene network models of embryos, simulations have matched observed developmental time courses with step times from regulator node to target node of 15 minutes (fruit fly) (Clark and Akam, 2016) to 3 hours (sea urchin) (Peter et al., 2012). Some immediate-early gene expression in acute signal responses can also occur within less than 30 minutes in mammalian cells (Bhatt et al., 2012), but a common observation is that using transcription factors to induce developmental state changes in postnatal mammalian cell populations seems to take days (Dionne et al.,

2005; Laiosa et al., 2006; Bussmann et al., 2009; Del Real and Rothenberg, 2013). Is this because of a need for regulator cascades to generate the full complement of trans-acting factors, or because of chromatin state resistance? Single-cell analysis has now made it possible in mammalian postembryonic development to examine the kinetic coupling between the provision of regulatory inputs and the appearance of one direct regulatory output.

In analysis of *Bcl11b* activation in early T cell precursors in mice, it was recently possible to separate the roles of trans-acting factors from those of any other state variables that might be affecting each allele in *cis* (Ng et al., 2018), presumably via chromatin or intranuclear location (Isoda et al., 2017). The two mechanisms could be clearly distinguished by tagging the two *Bcl11b* alleles in the same cells with different fluorescent reporters (Ng et al., 2018). Activation of one allele would show when the cell contained a full complement of permissive trans-regulatory factors and had turned off any trans-acting repressors; then, the question was whether both alleles would always be activated coordinately as T cell precursors developed. The two alleles were, in fact, activated at similar times in the population, but non-coordinately in single cells (Ng et al., 2018). Analysis of clones from single cells revealed that one allele could be activated more than 2 days earlier (2–4 cell divisions earlier) than the other within the same cells. The preference for one allele often continued through several cell divisions within an individual clone, even though within the cell population as a whole both alleles were equally likely to be first. The full analysis strongly indicated that a slow, cis-acting mechanism on each individual allele must operate to allow gene expression, even after the cell contained the full transcription factor complement needed to activate another, equivalent allele (Ng et al., 2018).

The *Bcl11b* case is clearly one extreme of behavior. It does appear to begin in a "repressed" chromatin state in the stem-cell precursors of T and other hematopoietic cells, based on DNA methylation and histone modification (Ji et al., 2010; Li et al., 2013), and so its slow, stochastic activation of each allele independently can be interpreted as a measure of the tenacity of that locally repressed state. It is quite possible that the embryonic systems previously modeled are faster in their responses to trans-activating factors because they do not exhibit an equally deep starting level of repression; this remains to be studied systematically. Nevertheless, the slow release of the *Bcl11b* gene from a closed, inactive chromatin configuration proved to be a major contributor to its developmental activation timing. Thus, an attempt to predict its response from the presence of its trans-acting factors alone would be inaccurate in this system, and possibly in other systems exhibiting stochastic activation control, or even more generally in postembryonic developmental systems.

2.5. Repression

The "dark matter" of epigenetic mechanisms is repression. It is not clear how many kinds of repression there are, or what determines the severity of the barrier they pose to gene expression (Filion et al., 2010; Jones, 2012; Schwartz and Pirrotta, 2013; Tchasovnikarova et al., 2015; Bogdanovic and Lister, 2017). In early embryos, where major functions are the orderly activation of gene expression in appropriate combinations, repression is used to establish borders between future cell types. In later development, repression shows additional facets, which strongly implies that repression comprises a variety of distinct mechanisms with distinct properties of reversibility.

Silence may not always reflect repression. First, there are physiological activation response genes that are non-expressed as a default, but are, in fact, primed for rapid expression when the correct combination of stimulatory factors is activated (Ramirez-Carrozzi et al., 2009). Second, there are genes associated with closely related developmental lineages or with developmental states that cells may achieve conditionally, which are not expressed normally by the cells but which can be activated under stress conditions or at a late stage of development (Rothenberg, 2013). These genes are facultatively repressed, offering cells the option of functional plasticity. Third, there are many genes in the genome of any differentiated cell that are no longer eligible to be expressed under any condition, short of wholesale artificial reprogramming and recapitulating embryonic development from the start. In a simple gene network assumption, it is not clear why these genes should be so hard to activate.

Early views, conditioned by some reprogramming successes, made it possible to imagine that these stably silent genes simply awaited the presence of a particular "master regulator," which would normally be expressed only in a given cell context. However, this view ignores the fact that many cells do express the "master regulators" of unrelated cell types, and yet the two cell types may successfully put those factors to

work at controlling the expression of completely different gene sets. The B cell transcription factor EBF1, used to define the B cell identity, has completely different roles in adipose cells and in the olfactory epithelium (Liberg et al., 2002), where it is also important. EBF1 was one of the first factors shown to bind completely dissimilar genomic sites in B cell and stromal cell contexts (Treiber et al., 2010). In fact, the regulatory factors used in any given type of differentiated cell, or close family members with highly similar DNA recognition specificities, are usually recycled in other tissues of the body, working in completely different gene networks.

Indeed, this transcription factor multitasking could be viewed as part of the evolutionary pressure for establishing deep repression mechanisms that effectively mask off parts of the genome. Both myogenic and neurogenic basic helix-loop-helix factors, for example, bind very similar regulatory sites that overlap enough to threaten havoc if, for example, they were actually to activate both muscle and neuron target genes in the same cells (Fong et al., 2012). Evidence has now accumulated for other potent lineage-determining factors besides EBF1, showing that their genomic occupancy patterns are quite divergent in different cell types that happen to share the expression of that factor, as for PU.1 (myeloid cells, B cells, and early T cells) and Tal1 (leukemia and erythroid cells), or even within the same lineage at different developmental stages, as for Pax5 (B cells), GATA-3, and Runx1 (T cells) (Heinz et al., 2010; Palii et al., 2011; Revilla-i-Domingo et al., 2012; Zhang et al., 2012; Hosokawa et al., 2018b). Thus, despite the potency of these factors in permitting or driving particular transitions in development, they are not autonomous in their ability to select their target sites; they follow as well as lead. The coherence of the programs they activate may depend, at least in part, on the masking of many potential sites from their activity, as well as on their ability to engage and work at others.

How trans- and cis-mechanisms interact in these various forms of repression is not well defined. Transcription factor binding sites can be mapped efficiently, but this has been disappointingly weak as a way to predict where transcription factors are functionally effective in general (McManus et al., 2011; Fisher et al., 2012; Kok et al., 2015; Barakat et al., 2018; Hosokawa et al., 2018a; Ungerbäck et al., 2018), and repression is much less well understood than activation. Open chromatin, indicating favored sites of transcription factor binding, has a strong bias toward correlation with sites of positive transcriptional regulation (Buenrostro et al., 2013; Vierstra and Stamatoyannopoulos, 2016). The great majority of the known histone modifications that can give evidence for the effect of a transcription factor's binding at a given site are also biased to be indicators of local positive regulation (Schones and Zhao, 2008; Sashida and Iwama, 2012; Calo and Wysocka, 2013). Three histone modifications currently known indicate repression: H3K9me3, H3K27me3, and the more rarely studied H2AK119Ub, and these are not found at all silent loci. Although most genes that have H3K27me3 marks on their promoters are repressed, the majority of repressed genes in a given cell type cannot be identified by this mark, and those that do are often facultatively rather than deeply repressed (see discussion in Rothenberg, 2013). Qualitatively different mechanisms may need to be invoked. Sometimes, developmentally deeply repressed genes in postembryonic hematopoietic cells may be found in a separate nuclear compartment, associated with the nuclear lamina (Lin et al., 2012). Little or nothing is known about how deeply repressed autosomal genes (including Bcl11b in non-T blood cells and EBF1 in non-B blood cells) are brought there, and only a little about how they may be released (Kueh et al., 2016; Isoda et al., 2017; Ng et al., 2018).

For the purpose of modeling gene network behavior, the real question is not just whether a blockade of expression can be predicted, but whether it is possible to predict how rapidly a given mechanism can establish repression and/or allow the repressed state to be lifted. Downregulation of transcription at a locus frequently precedes appearance of ''repressive'' chromatin marks (Zhang et al., 2012; Hosogane et al., 2016). Recent work in minimal artificial systems has been extremely useful, showing notable differences in these respects between repression based on recruitment of histone deacetylase (HDAC), of DNA methylation, of Polycomb repressive complex 2, and of Krüppel-associated box (KRAB)-domain recruited effectors, respectively (Bintu et al., 2016). Some of these repression mechanisms silence genes rapidly but are rapidly reversed; whereas others, taking longer to become established, are more permanent. Even when the same co-repression mechanism seems to be used, the vernalization system in Arabidopsis plants shows a convincing distinction between the mechanism initiating repression (rapid, nucleating Polycomb repressive complex 2 at a specific site) and the mechanism rendering it stable (extension of repressive H3K27me3 marks) (Yang et al., 2017). Thus, to be able to model accurately the speed and conditions under which positive trans-acting factors can activate a silenced gene, much needs to be learned about the nature of the repressive mechanism and possibly even about the duration of that mechanism's application, as well.

2.6. Regulatory information mapping and syntax

Finally, the predictiveness of modeling depends on accurate definition of the relationship between regulatory inputs and responding genes, a problem that has turned out to be considerably more challenging than was previously believed. The strong, explanatory models for developmental gene regulatory network operation have been supported by identification of regulatory elements that directly mediate the effects of specific upstream regulators on particular downstream targets (Jaeger et al., 2012; Peter et al., 2012). The power of the analysis has been supported by the apparently discrete roles of compact regulatory elements in directing particular domains of the target genes' expression, so that such regulatory elements can act as the templates for the response logic at each of the network connections (Small et al., 1991, 1993; Bockamp et al., 1997; Fujioka et al., 1999; Sanchez et al., 1999; Markstein et al., 2004). In this case, the specific combination of transcription factor binding sites at the responsible cis-regulatory enhancer can be seen as the ''ground truth'' for identifying the direct inputs to the target gene's expression pattern.

But are these regulatory elements complete templates? The genes that represent important nodes in developmental gene networks often have complex regulatory systems, not only with multiple enhancers driving expression in discrete contexts but also with multiple regulatory elements that work in overlapping contexts (Perry et al., 2011; Barolo, 2012; Gonzalez et al., 2015; Staller et al., 2015; Huang et al., 2016; Cui et al., 2017; Berthelot et al., 2018; Vincent et al., 2018). Activity involving multiple cis-regulatory elements at the same time is a genomically widespread feature of regulatory genes, sometimes spreading hundreds of kilobases beyond the genes themselves. This is seen through the complexity of histone mark and DNA accessibility features around such genes (Gonzalez et al., 2015) and the highly complex DNA looping interactions that contact their promoters, apparently simultaneously, in various developmental contexts (Kieffer-Kwon et al., 2013; Yashiro-Ohtani et al., 2014; Fabre et al., 2017; Bahr et al., 2018; Hu et al., 2018). Individual cis-regulatory sequences, recognized specifically by particular transcription factors, may, indeed, set a ''rule'' for the collaboration of those factors to cause activity of that particular element. However, if multiple regulatory elements are working simultaneously to regulate a given gene, it becomes difficult to predict how much of a contribution to the total expression of that gene any particular enhancer can make.

As genome-scale enhancer deletion tests have become feasible, it is clear that individual elements may have only weak non-negative influences on overall target gene expression (Kieffer-Kwon et al., 2013; Bothma et al., 2015; Diao et al., 2016; Huang et al., 2016; Gonen et al., 2017; Moorthy et al., 2017). Further, if the same transcription factor binds in parallel to different regulatory elements of the same target, it can become part of several different rule sets, and its own impact on the target gene may not be accounted for by any of the regulatory elements on its own (Staller et al., 2015). Finally, whether these regulatory elements are working via probabilistic or continuous-valued outputs, their collaboration matters. However, current knowledge does not tell us what syntax governs their relationships, or how to recognize a case in which repression at one element may be allowed to dominate over activation at other elements.

For phenomenological gene networks, this may not matter. Phenomenologically, it is possible simply to carry out massive perturbation screens and score all the genes in the system for their responses to all the perturbations. For machine learning, possible network configurations can probably be proposed based on optimization of Boolean or ODE-based frameworks. Thus, descriptively rendering all these responses in a formal pattern could be viewed as a gene network model, a useful summary of existing observations. However, such a model would not necessarily identify true causal linkages, and would not, in fact, reveal how best to re-engineer the network with the correct kinetic response parameters and with the minimum of off-target effects. Not only better kinetic measurements but also substantive scientific advances in explaining repression modalities and large-scale genomic regulatory element syntax will be needed to convert a descriptive system picture into an explanatory gene network solution.

3. CONCLUDING REMARKS

Building gene network models is a project that greatly enhances understanding of developmental processes. Models that are formulated capture currently available knowledge and lay out the best available guesses as to the rule sets that explain developmental causality. In so doing, they provide a strong hypothetical framework in which specific relationships need to be tested, to measure quantitative or kinetic effects or in some cases to

be disproven. However, there are postembryonic systems where well-studied developmental mechanisms are currently very difficult to model with high predictiveness, not only because of difficulties in measuring parameter values but also because of theoretical gaps in understanding the component processes. If epigenetic modifications, indeed, play a role in regulatory state inertia, then they need to be studied not only in static snapshots but also in the context of their relationships with transcription factor dynamics and the outputs of transcription factor change. If different repression mechanisms make the distinction between conditionally activatable and permanently inaccessible genes, then these differences are fundamental to the irreversibility of differentiation, and they need to be characterized in a serious comparison. Finally, the rules that make some transcription factor/DNA complexes work as enhancers whereas others fail to do so must be elucidated with much greater generality, for until this is understood, we will only glean the lowest-hanging fraction of fruits from the abundant orchards of genomics. Development, much richer in real organisms than most artificial models, provides an impetus as well as a laboratory for these explorations.

ACKNOWLEDGMENTS

AUTHOR DISCLOSURE STATEMENT

The author declares there are no competing financial interests.

REFERENCES

Abou-Jaoudé, W., Traynard, P., Monteiro, P.T., et al. 2016. Logical modeling and dynamical analysis of cellular networks. *Front. Genet.* 7, 94.

Albert, R., and Thakar, J. 2014. Boolean modeling: A logic-based dynamic approach for understanding signaling and regulatory networks and for making useful predictions. *Wiley Interdiscip. Rev. Syst. Biol. Med.* 6, 353–369.

Arvey, A., Agius, P., Noble, W.S., et al. 2012. Sequence and chromatin determinants of cell-type-specific transcription factor binding. *Genome Res.* 22, 1723–1734.

Avram, D., and Califano, D. 2014. The multifaceted roles of Bcl11b in thymic and peripheral T cells: Impact on immune diseases. *J. Immunol.* 193, 2059–2065.

Ay, A., and Arnosti, D.N. 2011. Mathematical modeling of gene expression: A guide for the perplexed biologist. *Crit. Rev. Biochem. Mol. Biol.* 46, 137–151.

Bahr, C., von Paleske, L., Uslu, V.V., et al. 2018. A Myc enhancer cluster regulates normal and leukaemic haematopoietic stem cell hierarchies. *Nature* 553, 515–520.

Balaskas, N., Ribeiro, A., Panovska, J., et al. 2012. Gene regulatory logic for reading the Sonic Hedgehog signaling gradient in the vertebrate neural tube. *Cell* 148, 273–284.

Barakat, T.S., Halbritter, F., Zhang, M., et al. 2018. Functional dissection of the enhancer repertoire in human embryonic stem cells. *Cell Stem Cell* 23, 276–288 e278.

Barolo, S. 2012. Shadow enhancers: Frequently asked questions about distributed cis-regulatory information and enhancer redundancy. *Bioessays* 34, 135–141.

Becker, A.M., Michael, D.G., Satpathy, A.T., et al. 2012. IRF-8 extinguishes neutrophil production and promotes dendritic cell lineage commitment in both myeloid and lymphoid mouse progenitors. *Blood* 119, 2003–2012.

Ben Tabou de-Leon, S., and Davidson, E.H. 2009. Modeling the dynamics of transcriptional gene regulatory networks for animal development. *Dev. Biol.* 325, 317–328.

Berthelot, C., Villar, D., Horvath, J.E., et al. 2018. Complexity and conservation of regulatory landscapes underlie evolutionary resilience of mammalian gene expression. *Nat. Ecol. Evol.* 2, 152–163.

Bhatt, D.M., Pandya-Jones, A., Tong, A.J., et al. 2012. Transcript dynamics of proinflammatory genes revealed by sequence analysis of subcellular RNA fractions. *Cell* 150, 279–290.

Bintu, L., Buchler, N.E., Garcia, H.G., et al. 2005. Transcriptional regulation by the numbers: Models. *Curr. Opin. Genet. Dev.* 15, 116–124.

Bintu, L., Yong, J., Antebi, Y.E., et al. 2016. Dynamics of epigenetic regulation at the single-cell level. *Science* 351, 720–724.

Bix, M., and Locksley, R.M. 1998. Independent and epigenetic regulation of the interleukin-4 alleles in CD4 + T cells. *Science* 281, 1352–1354.

Black, J.C., Van Rechem, C., and Whetstine, J.R. 2012. Histone lysine methylation dynamics: Establishment, regulation, and biological impact. *Mol. Cell.* 48, 491–507.

Bock, C., Beerman, I., Lien, W.H., et al. 2012. DNA methylation dynamics during in vivo differentiation of blood and skin stem cells. *Mol. Cell.* 47, 633–647.

Bockamp, E.O., McLaughlin, F., Gottgens, B., et al. 1997. Distinct mechanisms direct SCL/tal-1 expression in erythroid cells and CD34 positive primitive myeloid cells. *J. Biol. Chem.* 272, 8781–8790.

Bogdanovic, O., and Lister, R. 2017. DNA methylation and the preservation of cell identity. *Curr. Opin. Genet. Dev.* 46, 9–14.

Bolouri, H., and Davidson, E.H. 2002. Modeling transcriptional regulatory networks. *Bioessays* 24, 1118–1129.

Bolouri, H., and Davidson, E.H. 2003. Transcriptional regulatory cascades in development: Initial rates, not steady state, determine network kinetics. *Proc. Natl. Acad. Sci. U. S. A.* 100, 9371–9376.

Bolouri, H., and Davidson, E.H. 2010. The gene regulatory network basis of the ''community effect,'' and analysis of a sea urchin embryo example. *Dev. Biol.* 340, 170–178.

Bothma, J.P., Garcia, H.G., Esposito, E., et al. 2014. Dynamic regulation of eve stripe 2 expression reveals transcriptional bursts in living Drosophila embryos. *Proc. Natl. Acad. Sci. U. S. A.* 111, 10598–10603.

Bothma, J.P., Garcia, H.G., Ng, S., et al. 2015. Enhancer additivity and non-additivity are determined by enhancer strength in the Drosophila embryo. *Elife* 4.

Briscoe, J., and Small, S. 2015. Morphogen rules: Design principles of gradient-mediated embryo patterning. *Development* 142, 3996–4009.

Buenrostro, J.D., Giresi, P.G., Zaba, L.C., et al. 2013. Transposition of native chromatin for fast and sensitive epigenomic profiling of open chromatin, DNA-binding proteins and nucleosome position. *Nat. Methods.* 10, 1213–1218.

Bussmann, L.H., Schubert, A., Vu Manh, T.P., et al. 2009. A robust and highly efficient immune cell reprogramming system. *Cell Stem Cell* 5, 554–566.

Calo, E., and Wysocka, J. 2013. Modification of enhancer chromatin: What, how, and why? *Mol. Cell.* 49, 825–837.

Chen, T., and Dent, S.Y. 2014. Chromatin modifiers and remodellers: Regulators of cellular differentiation. *Nat. Rev. Genet.* 15, 93–106.

Chickarmane, V., Enver, T., and Peterson, C. 2009. Computational modeling of the hematopoietic erythroid-myeloid switch reveals insights into cooperativity, priming, and irreversibility. *PLoS Comput. Biol.* 5, e1000268.

Chiodetti, L., Barber, D.L., and Schwartz, R.H. 2000. Biallelic expression of the IL-2 locus under optimal stimulation conditions. *Eur. J. Immunol.* 30, 2157–2163.

Clark, E. 2017. Dynamic patterning by the Drosophila pair-rule network reconciles long-germ and short-germ segmentation. *PLoS Biol.* 15, e2002439.

Clark, E., and Akam, M. 2016. Odd-paired controls frequency doubling in Drosophila segmentation by altering the pair-rule gene regulatory network. *Elife* 5, e18215.

Cohen, M., Page, K.M., Perez-Carrasco, R., et al. 2014. A theoretical framework for the regulation of Shh morphogen-controlled gene expression. *Development* 141, 3868–3878.

Crocker, J., Abe, N., Rinaldi, L., et al. 2015. Low affinity binding site clusters confer hox specificity and regulatory robustness. *Cell* 160, 191–203.

Cui, M., Vielmas, E., Davidson, E.H., et al. 2017. Sequential response to multiple developmental network circuits encoded in an intronic cis-regulatory module of sea urchin *hox11/13b. Cell Rep.* 19, 364–374.

Dahl, R., Walsh, J.C., Lancki, D., et al. 2003. Regulation of macrophage and neutrophil cell fates by the PU.1:C/EBPα ratio and granulocyte colony-stimulating factor. *Nat. Immunol.* 4, 1029–1036.

Davidson, E.H. 2010. Emerging properties of animal gene regulatory networks. *Nature* 468, 911–920.

DeKoter, R.P., and Singh, H. 2000. Regulation of B lymphocyte and macrophage development by graded expression of PU.1. *Science* 288, 1439–1441.

Del Real, M.M., and Rothenberg, E.V. 2013. Architecture of a lymphomyeloid developmental switch controlled by PU.1, Notch and Gata3. *Development* 140, 1207–1219.

Diao, Y., Li, B., Meng, Z., et al. 2016. A new class of temporarily phenotypic enhancers identified by CRISPR/Cas9-mediated genetic screening. *Genome Res.* 26, 397–405.

Dionne, C.J., Tse, K.Y., Weiss, A.H., et al. 2005. Subversion of T lineage commitment by PU.1 in a clonal cell line system. *Dev. Biol.* 280, 448–466.

Fabre, P.J., Leleu, M., Mormann, B.H., et al. 2017. Large scale genomic reorganization of topological domains at the HoxD locus. *Genome. Biol.* 18, 149.

Farley, E.K., Olson, K.M., Zhang, W., et al. 2015. Suboptimization of developmental enhancers. *Science* 350, 325–328.

Filion, G.J., van Bemmel, J.G., Braunschweig, U., et al. 2010. Systematic protein location mapping reveals five principal chromatin types in Drosophila cells. *Cell* 143, 212–224.

Fisher, W.W., Li, J.J., Hammonds, A.S., et al. 2012. DNA regions bound at low occupancy by transcription factors do not drive patterned reporter gene expression in Drosophila. *Proc. Natl. Acad. Sci. U. S. A.* 109, 21330–21335.

Fong, A.P., Yao, Z., Zhong, J.W., et al. 2012. Genetic and epigenetic determinants of neurogenesis and myogenesis. *Dev. Cell.* 22, 721–735.

Fujioka, M., Emi-Sarker, Y., Yusibova, G.L., et al. 1999. Analysis of an even-skipped rescue transgene reveals both composite and discrete neuronal and early blastoderm enhancers, and multi-stripe positioning by gap gene repressor gradients. *Development* 126, 2527–2538.

Fukaya, T., Lim, B., and Levine, M. 2016. Enhancer control of transcriptional bursting. *Cell* 166, 358–368.

Goncalves, A., Leigh-Brown, S., Thybert, D., et al. 2012. Extensive compensatory cis-trans regulation in the evolution of mouse gene expression. *Genome Res.* 22, 2376–2384.

Gonen, N., Quinn, A., O'Neill, H.C., et al. 2017. Normal levels of Sox9 expression in the developing mouse testis depend on the TES/TESCO enhancer, but this does not act alone. *PLoS Genet.* 13, e1006520.

Gonzalez, A.J., Setty, M., and Leslie, C.S. 2015. Early enhancer establishment and regulatory locus complexity shape transcriptional programs in hematopoietic differentiation. *Nat. Genet.* 47, 1249–1259.

Graf, T., and Enver, T. 2009. Forcing cells to change lineages. *Nature* 462, 587–594.

Gurdon, J.B. 1988. A community effect in animal development. *Nature* 336, 772–774.

Hardison, R.C., and Taylor, J. 2012. Genomic approaches towards finding cis-regulatory modules in animals. *Nat. Rev. Genet.* 13, 469–483.

Heinz, S., Benner, C., Spann, N., et al. 2010. Simple combinations of lineage-determining transcription factors prime cis-regulatory elements required for macrophage and B cell identities. *Mol. Cell.* 38, 576–589.

Ho, R., Chronis, C., and Plath, K. 2011. Mechanistic insights into reprogramming to induced pluripotency. *J. Cell. Physiol.* 226, 868–878.

Holländer, G.A., Zuklys, S., Morel, C., et al. 1998. Monoallelic expression of the Interleukin-2 locus. *Science* 279, 2118–2121.

Hoppe, P.S., Schwarzfischer, M., Loeffler, D., et al. 2016. Early myeloid lineage choice is not initiated by random PU.1 to GATA1 protein ratios. *Nature* 535, 299–302.

Hosogane, M., Funayama, R., Shirota, M., et al. 2016. Lack of transcription triggers H3K27me3 accumulation in the gene body. *Cell Rep.* 16, 696–706.

Hosokawa, H., Romero-Wolf, M., Yui, M.A., et al. 2018a. Bcl11b sets pro-T cell fate by site-specific cofactor recruitment and by repressing Id2 and Zbtb16. *Nat. Immunol.* 19, 1427–1440.

Hosokawa, H., Ungerbäck, J., Wang, X., et al. 2018b. Transcription factor PU.1 represses and activates gene expression in early T cells by redirecting partner transcription factor binding. *Immunity* 48, 1119–1134.e1117.

Hu, G., Cui, K., Fang, D., et al. 2018. Transformation of accessible chromatin and 3D nucleome underlies lineage commitment of early T cells. *Immunity* 48, 227–242 e228.

Huang, J., Liu, X., Li, D., et al. 2016. Dynamic control of enhancer repertoires drives lineage and stage-specific transcription during hematopoiesis. *Dev. Cell.* 36, 9–23.

Huang, S., Guo, Y.P., May, G., et al. 2007. Bifurcation dynamics in lineage-commitment in bipotent progenitor cells. *Dev. Biol.* 305, 695–713.

Isoda, T., Moore, A.J., He, Z., et al. 2017. Non-coding transcription instructs chromatin folding and compartmentalization to dictate enhancer-promoter communication and T cell fate. *Cell* 171, 103–119 e118.

Jaeger, J., Manu, and Reinitz, J. 2012. Drosophila blastoderm patterning. *Curr. Opin. Genet. Dev.* 22, 533–541.

Jaeger, J., Surkova, S., Blagov, M., et al. 2004. Dynamic control of positional information in the early Drosophila embryo. *Nature* 430, 368–371.

Jeong, M., Sun, D., Luo, M., et al. 2014. Large conserved domains of low DNA methylation maintained by Dnmt3a. *Nat. Genet.* 46, 17–23.

Ji, H., Ehrlich, L.I., Seita, J., et al. 2010. Comprehensive methylome map of lineage commitment from haematopoietic progenitors. *Nature* 467, 338–342.

Jones, P.A. 2012. Functions of DNA methylation: Islands, start sites, gene bodies and beyond. *Nat. Rev. Genet.* 13, 484–492.

Khoueiry, P., Girardot, C., Ciglar, L., et al. 2017. Uncoupling evolutionary changes in DNA sequence, transcription factor occupancy and enhancer activity. *Elife* 6.

Kieffer-Kwon, K.R., Tang, Z., Mathe, E., et al. 2013. Interactome maps of mouse gene regulatory domains reveal basic principles of transcriptional regulation. *Cell* 155, 1507–1520.

Kok, K., Ay, A., Li, L.M., et al. 2015. Genome-wide errant targeting by Hairy. *Elife* 4.

Kouzarides, T. 2007. Chromatin modifications and their function. *Cell* 128, 693–705.

Kribelbauer, J.F., Laptenko, O., Chen, S., et al. 2017. Quantitative analysis of the DNA methylation sensitivity of transcription factor complexes. *Cell Rep.* 19, 2383–2395.

Kueh, H.Y., Champhekar, A., Nutt, S.L., et al. 2013. Positive feedback between PU.1 and the cell cycle controls myeloid differentiation. *Science* 341, 670–673.

Kueh, H.Y., Yui, M.A., Ng, K.K.H., et al. 2016. Asynchronous combinatorial action of four regulatory factors activates *Bcl11b* for T cell commitment. *Nat. Immunol.* 17, 956–965.

Laiosa, C.V., Stadtfeld, M., Xie, H., et al. 2006. Reprogramming of committed T cell progenitors to macrophages and dendritic cells by C/EBPα and PU.1 transcription factors. *Immunity* 25, 731–744.

Laslo, P., Spooner, C.J., Warmflash, A., et al. 2006. Multilineage transcriptional priming and determination of alternate hematopoietic cell fates. *Cell* 126, 755–766.

Li, D., Liu, J., Yang, X., et al. 2017. Chromatin accessibility dynamics during iPSC reprogramming. *Cell Stem Cell* 21, 819–833 e816.

Li, L., Zhang, J.A., Dose, M., et al. 2013. A far downstream enhancer for murine Bcl11b controls its T-cell specific expression. *Blood* 122, 902–911.

Liberg, D., Sigvardsson, M., and Kerblad, P. 2002. The EBF/Olf/Collier family of transcription factors: Regulators of differentiation in cells originating from all three embryonal germ layers. *Mol. Cell Biol.* 22, 8389–8397.

Lin, Y.C., Benner, C., Mansson, R., et al. 2012. Global changes in the nuclear positioning of genes and intra- and interdomain genomic interactions that orchestrate B cell fate. *Nat. Immunol.* 13, 1196–1204.

Liu, G., Wang, W., Hu, S., et al. 2018. Inherited DNA methylation primes the establishment of accessible chromatin during genome activation. *Genome. Res.* 28, 998–1007.

Liu, P., Li, P., and Burke, S. 2010. Critical roles of Bcl11b in T-cell development and maintenance of T-cell identity. *Immunol. Rev.* 238, 138–149.

Longabaugh, W.J.R., Zeng, W., Zhang, J.A., et al. 2017. Bcl11b and combinatorial resolution of cell fate in the T-cell gene regulatory network. *Proc. Natl. Acad. Sci. U. S. A.* 114, 5800–5807.

Markstein, M., Zinzen, R., Markstein, P., et al. 2004. A regulatory code for neurogenic gene expression in the Drosophila embryo. *Development* 131, 2387–2394.

Mbodj, A., Gustafson, E.H., Ciglar, L., et al. 2016. Qualitative dynamical modelling can formally explain mesoderm specification and predict novel developmental phenotypes. *PLoS Comput. Biol.* 12, e1005073.

McManus, S., Ebert, A., Salvagiotto, G., et al. 2011. The transcription factor Pax5 regulates its target genes by recruiting chromatin-modifying proteins in committed B cells. *EMBO J.* 30, 2388–2404.

Mercer, T.R., Edwards, S.L., Clark, M.B., et al. 2013. DNase I-hypersensitive exons colocalize with promoters and distal regulatory elements. *Nat. Genet.* 45, 852–859.

Monteiro, R., Pouget, C., and Patient, R. 2011. The *gata1/pu.1* lineage fate paradigm varies between blood populations and is modulated by *tif1γ*. *EMBO J.* 30, 1093–1103.

Moorthy, S.D., Davidson, S., Shchuka, V.M., et al. 2017. Enhancers and super-enhancers have an equivalent regulatory role in embryonic stem cells through regulation of single or multiple genes. *Genome Res.* 27, 246–258.

Ng, K.K.H., Yui, M.A., Mehta, A., et al. 2018. A stochastic epigenetic switch controls the dynamics of T-cell lineage commitment. *Elife* 7, e37851.

Olariu, V., and Peterson, C. 2019. Kinetic models of hematopoietic differentiation. *Wiley Interdiscip. Rev. Syst. Biol. Med.* 11, e1424.

Olsson, A., Venkatasubramanian, M., Chaudhri, V.K., et al. 2016. Single-cell analysis of mixed-lineage states leading to a binary cell fate choice. *Nature* 537, 698–702.

Palii, C.G., Perez-Iratxeta, C., Yao, Z., et al. 2011. Differential genomic targeting of the transcription factor TAL1 in alternate haematopoietic lineages. *EMBO J.* 30, 494–509.

Panne, D., Maniatis, T., and Harrison, S.C. 2007. An atomic model of the interferon-β enhanceosome. *Cell* 129, 1111–1123.

Perry, M.W., Boettiger, A.N., and Levine, M. 2011. Multiple enhancers ensure precision of gap gene-expression patterns in the Drosophila embryo. *Proc. Natl. Acad. Sci. U. S. A.* 108, 13570–13575.

Peter, I.S., and Davidson, E.H., 2015. *Genomic Control Process: Development and Evolution*. Academic Press, Elsevier, San Diego.

Peter, I.S., Faure, E., and Davidson, E.H. 2012. Predictive computation of genomic logic processing functions in embryonic development. *Proc. Natl. Acad. Sci. U. S. A.* 109, 16434–16442.

Raj, A., and van Oudenaarden, A. 2008. Nature, nurture, or chance: Stochastic gene expression and its consequences. *Cell* 135, 216–226.

Ramirez-Carrozzi, V.R., Braas, D., Bhatt, D.M., et al. 2009. A unifying model for the selective regulation of inducible transcription by CpG islands and nucleosome remodeling. *Cell* 138, 114–128.

Revilla-i-Domingo, R., Bilic, I., Vilagos, B., et al. 2012. The B-cell identity factor Pax5 regulates distinct transcriptional programmes in early and late B lymphopoiesis. *EMBO J.* 31, 3130–3146.

Rieger, M.A., Hoppe, P.S., Smejkal, B.M., et al. 2009. Hematopoietic cytokines can instruct lineage choice. *Science* 325, 217–218.

Roost, M.S., Slieker, R.C., Bialecka, M., et al. 2017. DNA methylation and transcriptional trajectories during human development and reprogramming of isogenic pluripotent stem cells. *Nat. Commun.* 8, 908.

Rothenberg, E.V. 2013. Epigenetic mechanisms and developmental choice hierarchies in T-lymphocyte development. *Brief. Funct. Genomics.* 12, 512–524.

Rothenberg, E.V., Kueh, H.Y., Yui, M.A., et al. 2016. Hematopoiesis and T-cell specification as a model developmental system. *Immunol. Rev.* 271, 72–97.

Sanchez, M.J., Göttgens, B., Sinclair, A.M., et al. 1999. An SCL 3' enhancer targets developing endothelium together with embryonic and adult haematopoietic regulators. *Development* 126, 3891–3904.

Sardina, J.L., Collombet, S., Tian, T.V., et al. 2018. Transcription factors drive Tet2-mediated enhancer demethylation to reprogram cell fate. *Cell Stem Cell* 23, 727–741 e729.

Sashida, G., and Iwama, A. 2012. Epigenetic regulation of hematopoiesis. *Int. J. Hematol.* 96, 405–412.

Schones, D.E., and Zhao, K. 2008. Genome-wide approaches to studying chromatin modifications. *Nat. Rev. Genet.* 9, 179–191.

Schwartz, Y.B., and Pirrotta, V. 2013. A new world of Polycombs: Unexpected partnerships and emerging functions. *Nat. Rev. Genet.* 14, 853–864.

Shmulevich, I., and Aitchison, J.D. 2009. Deterministic and stochastic models of genetic regulatory networks. *Methods Enzymol.* 467, 335–356.

Small, S., Arnosti, D.N., and Levine, M. 1993. Spacing ensures autonomous expression of different stripe enhancers in the even-skipped promoter. *Development* 19, 762–772.

Small, S., Kraut, R., Hoey, T., et al. 1991. Transcriptional regulation of a pair-rule stripe in Drosophila. *Genes. Dev.* 5, 827–839.

Spitz, F., and Furlong, E.E.M. 2012. Transcription factors: From enhancer binding to developmental control. *Nat. Rev. Genet.* 13, 613–626.

Sridharan, R., Gonzales-Cope, M., Chronis, C., et al. 2013. Proteomic and genomic approaches reveal critical functions of H3K9 methylation and heterochromatin protein-1gamma in reprogramming to pluripotency. *Nat. Cell Biol.* 15, 872–882.

Staller, M.V., Vincent, B.J., Bragdon, M.D., et al. 2015. Shadow enhancers enable Hunchback bifunctionality in the Drosophila embryo. *Proc. Natl. Acad. Sci. U. S. A.* 112, 785–790.

Swiers, G., Patient, R., and Loose, M. 2006. Genetic regulatory networks programming hematopoietic stem cells and erythroid lineage specification. *Dev. Biol.* 294, 525–540.

Taghon, T., Yui, M.A., and Rothenberg, E.V. 2007. Mast cell lineage diversion of T lineage precursors by the essential T cell transcription factor GATA-3. *Nat. Immunol.* 8, 845–855.

Takahashi, K., and Yamanaka, S. 2016. A decade of transcription factor-mediated reprogramming to pluripotency. *Nat. Rev. Mol. Cell. Biol.* 17, 183–193.

Tchasovnikarova, I.A., Timms, R.T., Matheson, N.J., et al. 2015. GENE SILENCING. Epigenetic silencing by the HUSH complex mediates position-effect variegation in human cells. *Science* 348, 1481–1485.

Treiber, T., Mandel, E.M., Pott, S., et al. 2010. Early B cell Factor 1 regulates B cell gene networks by activation, repression, and transcription- independent poising of chromatin. *Immunity* 32, 714–725.

Ulirsch, J.C., Lacy, J.N., An, X., et al. 2014. Altered chromatin occupancy of master regulators underlies evolutionary divergence in the transcriptional landscape of erythroid differentiation. *PLoS Genet.* 10, e1004890.

Ungerbäck, J., Hosokawa, H., Wang, X., et al. 2018. Pioneering, chromatin remodeling, and epigenetic constraint in early T-cell gene regulation by SPI1 (PU.1). *Genome Res.* 28, 1508–1519.

Vierstra, J., and Stamatoyannopoulos, J.A. 2016. Genomic footprinting. *Nat. Methods.* 13, 213–221.

Vincent, B.J., Staller, M.V., Lopez-Rivera, F., et al. 2018. Hunchback is counter-repressed to regulate even-skipped stripe 2 expression in Drosophila embryos. *PLoS Genet.* 14, e1007644.

Walsh, J.C., DeKoter, R.P., Lee, H.J., et al. 2002. Cooperative and antagonistic interplay between PU.1 and GATA-2 in the specification of myeloid cell fates. *Immunity* 17, 665–676.

Watanabe, A., Yamada, Y., and Yamanaka, S. 2013. Epigenetic regulation in pluripotent stem cells: A key to breaking the epigenetic barrier. *Philos. Trans. R. Soc. Lond. B. Biol. Sci.* 368, 20120292.

Wilczynski, B., Liu, Y.H., Yeo, Z.X., et al. 2012. Predicting spatial and temporal gene expression using an integrative model of transcription factor occupancy and chromatin state. *PLoS Comput. Biol.* 8, e1002798.

Wong, E.S., Schmitt, B.M., Kazachenka, A., et al. 2017. Interplay of *cis* and *trans* mechanisms driving transcription factor binding and gene expression evolution. *Nat. Commun.* 8, 1092.

Wunderlich, Z., and DePace, A.H. 2011. Modeling transcriptional networks in Drosophila development at multiple scales. *Curr. Opin. Genet. Dev.* 21, 711–718.

Yang, H., Berry, S., Olsson, T.S.G., et al. 2017. Distinct phases of Polycomb silencing to hold epigenetic memory of cold in Arabidopsis. *Science.* 357, 1142–1145.

Yashiro-Ohtani, Y., Wang, H., Zang, C., et al. 2014. Long-range enhancer activity determines Myc sensitivity to Notch inhibitors in T cell leukemia. *Proc. Natl. Acad. Sci. U. S. A.* 111, E4946–E4953.

Zhang, J.A., Mortazavi, A., Williams, B.A., et al. 2012. Dynamic transformations of genome-wide epigenetic marking and transcriptional control establish T cell identity. *Cell* 149, 467–482.

Address correspondence to:
Prof. Ellen V. Rothenberg
Division of Biology and Biological Engineering
California Institute of Technology
1200 E. California Boulevard, Mailcode 156-29
Pasadena, CA 91125

E-mail: evroth@its.caltech.edu