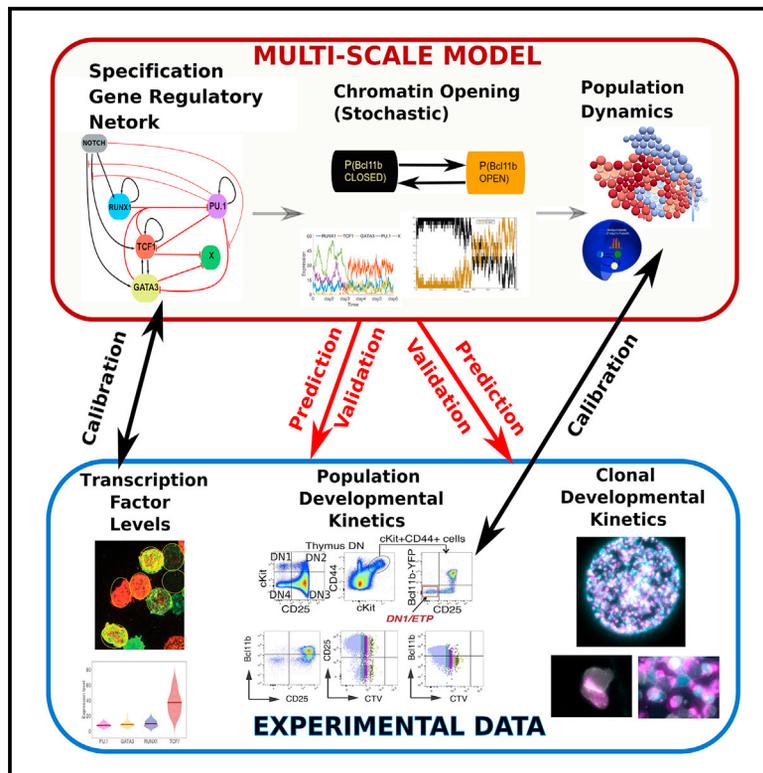


Multi-scale Dynamical Modeling of T Cell Development from an Early Thymic Progenitor State to Lineage Commitment

Graphical Abstract



Authors

Victor Olariu, Mary A. Yui, Pawel Krupinski, ..., Emil Andersson, Ellen V. Rothenberg, Carsten Peterson

Correspondence

evroth@its.caltech.edu (E.V.R.), carsten@thep.lu.se (C.P.)

In Brief

Olariu et al. use computational modeling and live-cell developmental imaging to explain the kinetics of early T cell lineage commitment. An integrated computational multi-scale model incorporating gene network architecture, single-cell RNA levels, chromatin state shifts, and proliferation is developed, explored, and validated.

Highlights

- A multi-level dynamical model is developed for the commitment of T cell precursors
- It links gene networks, single-cell RNA analysis, chromatin changes, and cell division
- It provides quantitative understanding of commitment kinetic requirements
- The model predictions are verified against new clonal and real-time imaging data



Article

Multi-scale Dynamical Modeling of T Cell Development from an Early Thymic Progenitor State to Lineage Commitment

Victor Olariu,^{1,3} Mary A. Yui,^{2,3} Pawel Krupinski,¹ Wen Zhou,² Julia Deichmann,¹ Emil Andersson,¹ Ellen V. Rothenberg,^{2,*} and Carsten Peterson^{1,4,*}

¹Computational Biology and Biological Physics, Department of Astronomy and Theoretical Physics, Lund University, Lund, Sweden

²Division of Biology and Biological Engineering, 156-29, California Institute of Technology, Pasadena, CA 91125, USA

³These authors contributed equally

⁴Lead Contact

*Correspondence: evroth@its.caltech.edu (E.V.R.), carsten@thep.lu.se (C.P.)

<https://doi.org/10.1016/j.celrep.2020.108622>

SUMMARY

Intrathymic development of committed progenitor (pro)-T cells from multipotent hematopoietic precursors offers an opportunity to dissect the molecular circuitry establishing cell identity in response to environmental signals. This transition encompasses programmed shutoff of stem/progenitor genes, upregulation of T cell specification genes, proliferation, and ultimately commitment. To explain these features in light of reported *cis*-acting chromatin effects and experimental kinetic data, we develop a three-level dynamic model of commitment based upon regulation of the commitment-linked gene *Bcl11b*. The levels are (1) a core gene regulatory network (GRN) architecture from transcription factor (TF) perturbation data, (2) a stochastically controlled chromatin-state gate, and (3) a single-cell proliferation model validated by experimental clonal growth and commitment kinetic assays. Using RNA fluorescence *in situ* hybridization (FISH) measurements of genes encoding key TFs and measured bulk population dynamics, this single-cell model predicts state-switching kinetics validated by measured clonal proliferation and commitment times. The resulting multi-scale model provides a mechanistic framework for dissecting commitment dynamics.

INTRODUCTION

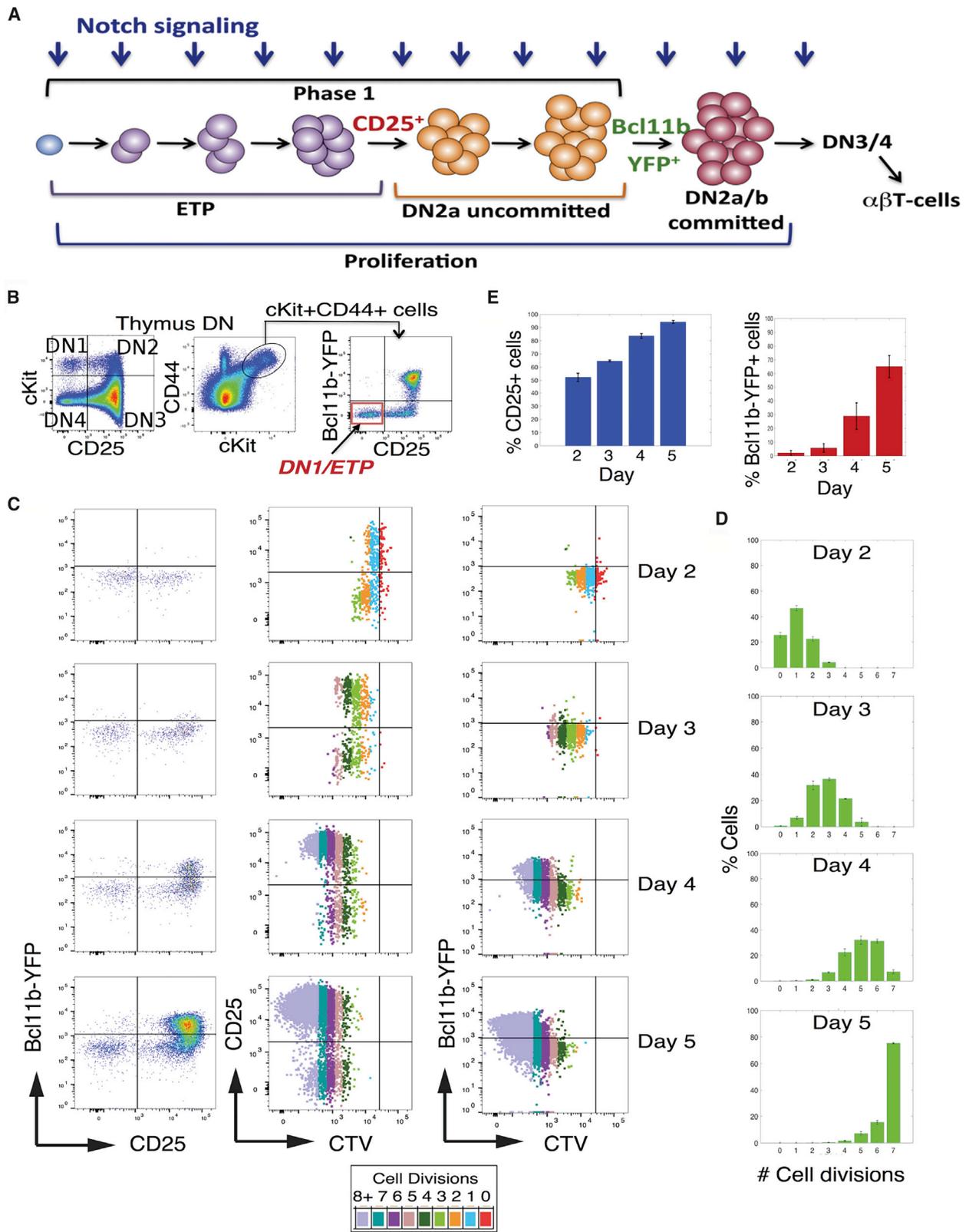
Hematopoietic progenitors continually replenish the body's supply of blood cells. While many of the molecular and cellular requirements of commitment to the various hematopoietic lineages have been studied in detail, how a single differentiating progenitor cell integrates multiple inputs into lineage decisions over time is less clear. Computational modeling approaches can provide new insights into the dynamics of this complex process.

T cell development provides an excellent model system for studying lineage commitment from a multipotent progenitor. Small numbers of multipotent hematopoietic precursors migrate continuously from bone marrow into the thymus (Ziętara et al., 2015; Zlotoff and Bhandoola, 2011) where they enter the T cell developmental pathway, driven by Notch signaling and cytokines in the thymic microenvironment. Many of the important regulatory factors controlling T-lineage commitment are known (for reviews, see De Obaldia and Bhandoola, 2015; Yui and Rothenberg, 2014) as well as their patterns of expression (David-Fung et al., 2009; Mingueneau et al., 2013; Zhang et al., 2012). Specific targets of several of these factors have been defined by perturbation studies, providing a strong basis for understanding the dynamic operation of an intrinsic gene regulatory network

(GRN) (Anderson et al., 2002; Champhekar et al., 2015; Chen et al., 1995; Del Real and Rothenberg, 2013; Dionne et al., 2005; Franco et al., 2006; Garcia-Ojeda et al., 2013; Germar et al., 2011; Hoogenkamp et al., 2007; Hosokawa et al., 2018; Kueh et al., 2016; Leddin et al., 2011; Li et al., 2010a, 2010b; Longabaugh et al., 2017; Scripture-Adams et al., 2014; Taghon et al., 2005, 2007; Weber et al., 2011; Xu et al., 2013; Zarnegar et al., 2010). However, the timing of observed transitions and the accompanying rapid proliferation leave open questions about how commitment decisions are made. In particular, recent experimental evidence has indicated a strong, rate-limiting role of *cis*-acting chromatin constraints in addition to the changes in *trans*-acting factors in controlling the timing of commitment (Ng et al., 2018). As a result, modeling the commitment process through a simple factor-target interaction approach alone is unrealistic. This work takes advantage of new and recently published experimental data to integrate these processes into a novel three-level model that explores the dynamics of T cell commitment at both population and single-cell levels.

Early stages of T cell development are well defined. Progenitor (pro)-T cells are CD4 and CD8 double negative (DN) and lack T cell receptor expression. The Kit^{high} early thymic progenitors (ETPs or DN1s) proliferate before transitioning to DN2a, a step marked by CD25 surface expression, and then they upregulate





(legend on next page)

Bcl11b and undergo Bcl11b-dependent T-lineage commitment (Figure 1A; Ikawa et al., 2010; Kueh et al., 2016; Li et al., 2010a, 2010b; Yui et al., 2010). Expression of the zinc-finger transcription factor (TF) *Bcl11b* not only alters many aspects of genomic activity and cell function (Avram and Califano, 2014; Hu et al., 2018; Kueh et al., 2016) but also correlates with the functional committed state of single DN2a cells (Kueh et al., 2016) and can thus serve as a proxy for commitment. After *Bcl11b* activation, Kit^{high} DN2a cells move through DN2b into the Kit^{low} DN3 stage, after which they can continue differentiation after rearranging a signaling-competent T cell receptor (Figure 1A). A broad shift in expression of multiple regulatory genes occurs during commitment, and the stem/progenitor TFs expressed in DN1 and DN2 stages are generally extinguished (for reviews, see Rothenberg et al., 2016; Yui and Rothenberg, 2014).

Computational modeling of GRNs represents an important starting point for studying mechanisms controlling cell-fate decisions (Olariu and Peterson, 2019). Our earlier dynamical models dealt, respectively, with the flux of pro-T cell populations through the DN1 to DN4 stages (Manesso et al., 2013) and with the transcriptional network for *Bcl11b* activation assuming deterministic, direct transcriptional regulation by positive regulators Notch, GATA3, and TCF1 (T cell factor 1, encoded by the *Tcf7* gene) (Manesso et al., 2016). In Ye et al. (2019), an exhaustive search of three-gene network topologies characterized by four attractors was pursued, recovering the one used in Manesso et al. (2016). Interestingly, such motifs can generate stepwise commitment as in T cell development. However, more recent empirical data have undermined basic assumptions of the deterministic direct model. First, we found that the roles of GATA3 and TCF1 are mostly required in a hit-and-run fashion, before *Bcl11b* is activated, in contrast to the roles of Runx1 and Notch signaling (Kueh et al., 2016). A simple combination of the earlier models, and evidence that all four known positive regulators are active by DN1 stage, would predict *Bcl11b* to be turned on multiple cell cycles earlier than it is, well beyond the range of measurement error. This suggests that the deterministic assumptions about *Bcl11b* control itself were incorrect or incomplete.

Second, recent evidence implies that the timing discrepancy also reflects an intervening local functionally important chromatin state change (epigenetic event). Extensive repression marks and methylation are removed from the *Bcl11b* locus during its activation (Ji et al., 2010; Li et al., 2013). This process can be rate limiting. Close analyses of cells with different fluorescent reporters in the two alleles of *Bcl11b* show that the two alleles

within the same DN2 cell nucleus can become activated at different times, with discordances of several days (and cell cycles) despite exposure to the same TFs (Ng et al., 2018). Consequently, in the model proposed here, we have introduced an explicit stochastic timing event for epigenetic activation in addition to the initial *trans*-acting regulatory activity.

Altogether, we propose a three-stage dynamic model of early T cell commitment from DN1 cells. The gene network aspects of the model take into account the unique linkage of commitment to *Bcl11b* activation and encompass *trans*-acting and epigenetic levels of *Bcl11b* regulation to account for its proximal activation kinetics. Step 1 is a single-cell GRN model based on previously published perturbation experiments with model parameters trained on new gene expression measurements from single-molecule RNA fluorescence *in situ* hybridization (smFISH). Step 2 is a collaborative epigenetic model to account for local chromatin constraints limiting *Bcl11b* activation, even after *trans*-acting factor requirements are met. At step 3, we embedded the single-cell two-step gene network model into a population growth model, trained with new cell culture dynamics data. Model predictions were confirmed by time-lapse imaging data from individual DN1 clones differentiating through commitment *in vitro*. The resulting multi-scale model closely represents measured early T cell commitment kinetics and helps to elucidate required mechanisms controlling this process, enabling future dissection of the controllers of commitment.

RESULTS

Experimental Results: Training Data Proliferation and Differentiation Kinetics from DN1 (ETP) to Commitment

To relate commitment rigorously to proliferation and to absolute time, we required new benchmark measurements linking cell proliferation and developmental kinetics of DN1 cell progeny through commitment. DN1 cells were purified from thymi of Bcl11b-YFP (yellow fluorescent protein) reporter mice (Kueh et al., 2016; Figure 1B) and then stained with CellTrace Violet (CTV) dye that tracks each cell's proliferation history through dilution. The cells were co-cultured with Notch-ligand-expressing OP9-DL1 stromal cells (Schmitt and Zúñiga-Pflücker, 2002) and growth-supporting cytokines interleukin-7 (IL-7) and Flt3 ligand (Flt3L) to promote T-lineage development, as previously described in detail (Yui et al., 2010). Parallel DN1 cultures were harvested daily, from day 2 to day 5, and assessed by flow

Figure 1. Proliferation and Developmental Kinetics from Multipotent DN1 Progenitors to T-Lineage-Committed Cells

(A) Schematic of T cell development showing initial Notch-dependent stages after multipotent precursors immigrate to thymus. The most immature Kit^{high} early T precursors (ETPs) are referred to here as DN1. At DN3, T cell receptor gene rearrangement occurs, and all subsequent development depends on T cell receptor expression and signaling.
(B) Flow cytometry plots showing early T cell developmental stages (DN1–DN4) and the gating strategies used for purifying progenitor DN1 cells, which are Kit^{high} and lack CD25 and Bcl11b-YFP, as inputs for cell culture.
(C–E) *In vitro* analysis of proliferation and development from DN1 cells. DN1 cells were purified from thymi of Bcl11b-YFP reporter mice, stained with the proliferation tracking dye CellTrace Violet (CTV), co-cultured with Notch ligand-expressing OP9-DL1 stromal cells to promote T cell development, and harvested for flow cytometric analysis on days 2–5. (C) Flow cytometry plots show the upregulation of CD25 followed by Bcl11b-YFP expression from days 2 to 5 (left), and the relationships between CD25 (middle) and Bcl11b-YFP (right) expression and the numbers of cell cycles each cell has experienced, as determined by stepwise decreases in CTV levels (color coded by cell division as shown; see Figure S1 for details). (D) Summary plots of the distributions of cell division numbers for progeny of cultured DN1 cells each day as measured by CTV. (E) Summary plots of the percentages of CD25-positive (blue) and of Bcl11b-YFP-positive (red) cells from each day. The histograms show mean and standard deviation of data from two experiments.

cytometry for total cell numbers, CTV fluorescence intensity, the developmentally regulated surface marker CD25 (marking DN2 transition), and Bcl11b-YFP (marking commitment) (Figure 1C). The number of divisions each cell had experienced was determined by flow cytometric measurement of residual CTV staining as calibrated in control cultures (Figure S1). Numbers of cell divisions experienced by individual cells are highlighted by different colors on the flow cytometry plots (Figure 1C) and as population distributions of cells with different numbers of cell divisions at each time (Figure 1D).

Approximately 50% of DN1 cells had differentiated to DN2, marked by CD25 expression, by day 2 of culture, with further percentage increases over time (Figures 1C and 1E). However, few if any committed Bcl11b-YFP⁺ cells appeared among the CD25⁺ cells until day 3, with percentages of YFP⁺CD25⁺ cells then increasing over time (Figures 1C and 1E). CTV analysis at day 2 showed that cell division was not required for onset of CD25 expression (red dots indicate cells with 0 divisions) and that the most rapidly dividing cells were neither most nor least likely to turn on CD25. By days 4 and 5, however, CD25⁺ DN2 cells on average had divided more rapidly than cells remaining CD25⁻, as seen by the shift toward a larger population with the lowest CTV levels (turquoise or lilac) (Figure 1C). Bcl11b-YFP was not expressed in any non-proliferating cells, and by days 4 and 5 the Bcl11b-YFP⁺ cells appeared to have proliferated more rapidly than cells remaining Bcl11b-YFP⁻, based on lower CTV levels for YFP⁺ cells (Figure 1C). Overall, these results confirm that CD25 is expressed before Bcl11b-YFP (Figures 1C and 1E) and that CD25 can be turned on in DN1s without cell division, while Bcl11b-YFP cannot be. Furthermore, more developmentally advanced cells in these cultures proliferated somewhat faster than those that were delayed.

Single-Cell RNA-FISH Measurement of Key TF Genes

Potential TF drivers of commitment include TCF1 (*Tcf7*), Gata3, and Runx1, driver that are necessary positive inputs for *Bcl11b* expression (Kueh et al., 2016), while PU.1 (*Spi1*) antagonizes developmental progression (Champhekar et al., 2015; Ungerback et al., 2018). To determine the normal ranges of expression of *Spi1*, *Tcf7*, *Runx1*, and *Gata3* in individual T cell precursors in developmental stages leading up to *Bcl11b* activation, for use in the model, DN thymocytes were purified from 5-week-old mice and analyzed by multiplex smFISH (Figures 2A and 2B) using gene-specific readout probes (STAR Methods). Developmental stages of individual cells were scored based on measured Kit, CD25 protein, and mRNA levels. Absolute transcript counts showed that *Runx1*, *Gata3*, and *Tcf7* expression was variable in DN1s, increasing up to 2 times in DN2s (DN2 median transcript levels/cell: *Tcf7* ~ 40; *Runx1* ~ 12; *Gata3* ~ 8) (Figure 2C). By contrast, *Spi1* was highest in DN1 and DN2 and declined with further development, in accord with previous bulk population measurements (Mingueneau et al., 2013; Zhang et al., 2012). *Bcl11b* was not expressed in DN1s but turned on in CD25⁺ DN2s (median transcripts/cell ~ 15), remaining high into DN3 (Figure 2C).

These gene expression changes from DN1 to DN2 were used as input to the GRN model as an approximation for developmental changes in TF protein. The half-lives of these proteins can differ substantially; for example, PU.1 has a half-life at least

10 times that of TCF1, Runx1, and Gata3 (Kueh et al., 2013; Manesso et al., 2016); therefore, their relative expression patterns were used rather than their absolute protein levels.

Modeling Results

Overview

As described above, a GRN model for *Bcl11b* activation during commitment must incorporate three separate aspects of control. First, four positive regulatory inputs, Notch signaling, TCF1, Gata3, and Runx1, are needed in the DN1-DN2a-stage cells to make *Bcl11b* eligible for activation. Second, TCF1 and GATA3 work in a hit-and-run way and are dispensable for the final steps of *Bcl11b* induction, while Runx1 and Notch are still required (Kueh et al., 2016). Third, activation of each *Bcl11b* allele, in a cell fulfilling the eligibility requirements, still requires a slow stochastic epigenetic remodeling process (Ng et al., 2018). To accommodate these requirements, we separated an initial, transcriptionally driven specification process from a subsequent, epigenetic process.

Thus, we developed a three-level dynamic model. (1) In the specification level, network dynamics between Notch signaling, Runx1, *Tcf7* (for convenience in gene network model, using the same term for *Tcf7* gene and its product, also called TCF1), Gata3, and an antagonistic regulator (possibly PU.1) determine cell permissiveness for *Bcl11b* activation but act indirectly on *Bcl11b* via control of a critical, newly invoked function X, as described below. (2) In the stochastic level, the transition of *Bcl11b* from a closed to open state is influenced by the inhibitory function X, balanced against the continuing positive drivers Runx1 and Notch signaling. (3) Finally, in the population level, to predict how these mechanisms generate observed population phenotypes, the single-cell processes are combined with a population growth model that accounts for the expansion of cells that have and have not yet activated *Bcl11b*. Model development and validation procedures are shown in Figure 3.

Level 1: The Specification GRN Architecture

For the first level of the single-cell model, we proposed a GRN architecture based upon gene expression, perturbation experiments, and literature (Table 1), describing the interplay between the T cell specification regulatory genes *Tcf7*, *Gata3*, and *Runx1* and an opposing regulatory gene, which is here represented by *Spi1* (Figure 4A). However, these TFs alone could not account for the experimentally observed delay in *Bcl11b* upregulation (Kueh et al., 2016) because of the intervening *cis*-acting epigenetic process recently shown to be required (Ng et al., 2018). Therefore, rather than having these regulators control *Bcl11b* transcription directly, we proposed that their inputs collectively removed obstruction due to an antagonist of *Bcl11b* opening, X. X was defined as a composite of a slow initial chromatin opening mechanism and the actions of any additional DN1-specific antagonists of *Bcl11b* expression, still to be defined (Ng et al., 2018). Only X, Notch signaling, and Runx1 then propagate to the next stochastic, epigenetic level.

Specification Level: Deterministic Model, Parameter Setting Based on Pseudo-Time Series

To use the single-cell FISH data for setting TF values in the GRN (Figure 2), we created pseudo-time series from DN1 through DN2, which showed that in early DN1, expression of *Tcf7*

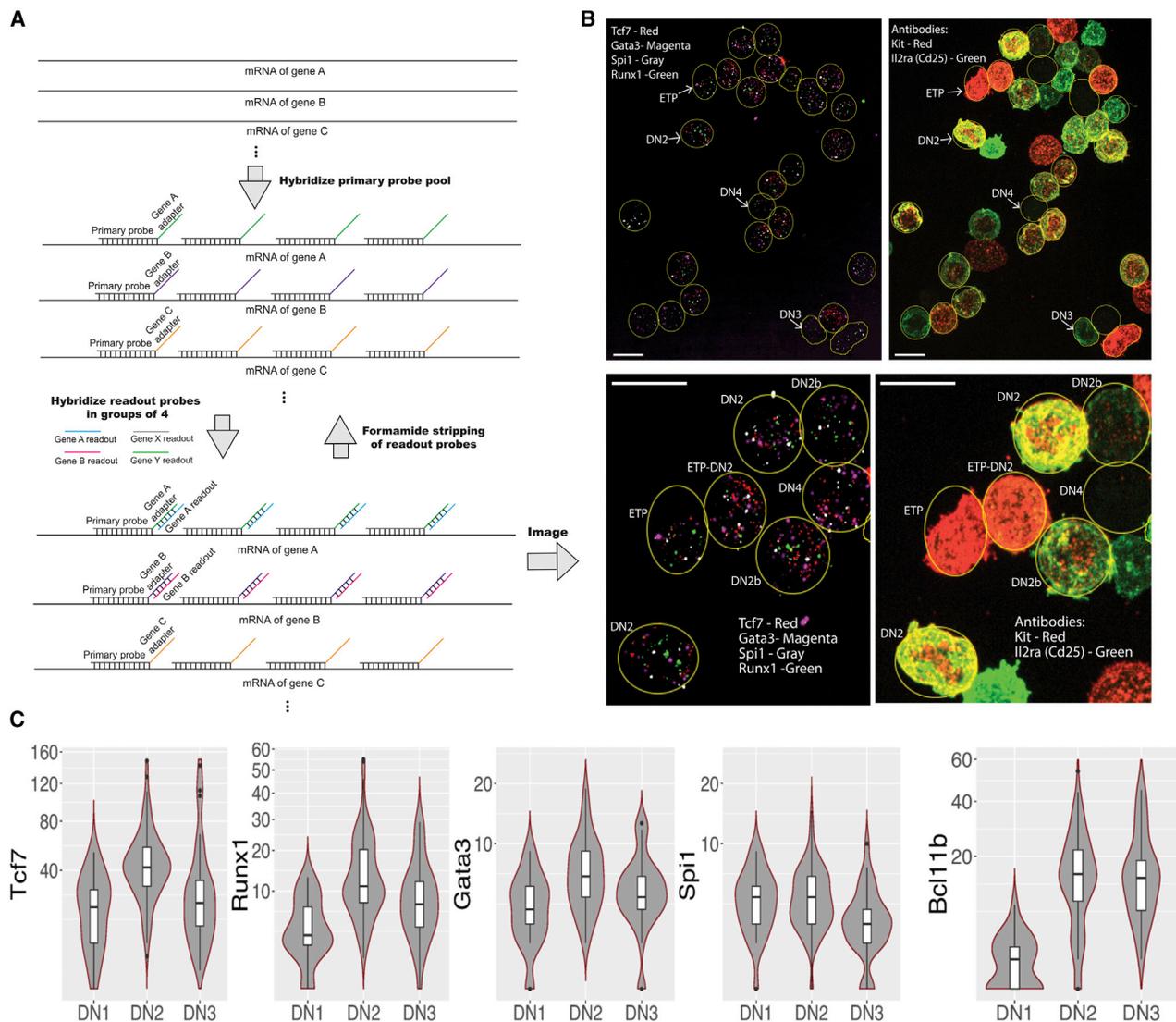


Figure 2. Regulatory Gene Expression Measurement by smFISH

(A) Schematics of smFISH hybridizations. Primary probes were hybridized against mRNAs in immobilized DN cells in pools, and genes were detected with readout probes (Table S6) in groups of four genes, imaged, and stripped with formamide before the second round of readout hybridizations.

(B) Representative microscopy images. Each mRNA transcript was represented by a single dot under fluorescent microscopy, and the transcripts of each gene were counted (Tcf7, red; Gata3, magenta; Spi1, gray; Runx1, green) and assigned to individual cells that were categorized into DN1-DN2-DN3-DN4 by antibody-stained surface markers Kit (red) and CD25 (encoded by Il2ra, green). Scale bars represent 10 μ m.

(C) Violin plots of transcript count distributions of *Runx1*, *Gata3*, *Spi1* (PU.1), *Tcf7*, and *Bcl11b* in single cells, showing median and quartiles, in developmental stages DN1–DN3. n = 169.

increases substantially and becomes predominant, while *Bcl11b* transcription begins only after a delay (200 arbitrary time units) (Figure 4B). *Gata3* expression slightly rises with increasing *Tcf7* expression, along with *Runx1*. By contrast, expression of the opposing gene, *Spi1* (PU.1), is fairly constant at a low level throughout these stages, turning off post-commitment (in DN3). For the following, note that PU.1 (*Spi1*) is the most prominent regulator in precommitment cells (Ungerback et al., 2018), but it is possible that another factor with a similar expression pattern to PU.1, or a complex of PU.1 with another factor, may be more directly involved than PU.1 alone in the initial restraint

on *Bcl11b* activation. The model simulations below predict the dynamics of the functionally relevant factor whether or not it is PU.1 itself.

We developed the computational model for the level 1 specification GRN (Figure 4A) using Shea-Ackers deterministic rate equations (Ackers et al., 1982; see STAR Methods). Initial conditions for *Tcf7*, *Gata3*, PU.1, and *Runx1* were set equal to the “earliest” pseudo-time-series expression values, and the X function was assumed to be highly active initially. The model results show that the system, when exposed to external Notch signaling, moves toward a steady state where the T cell factors

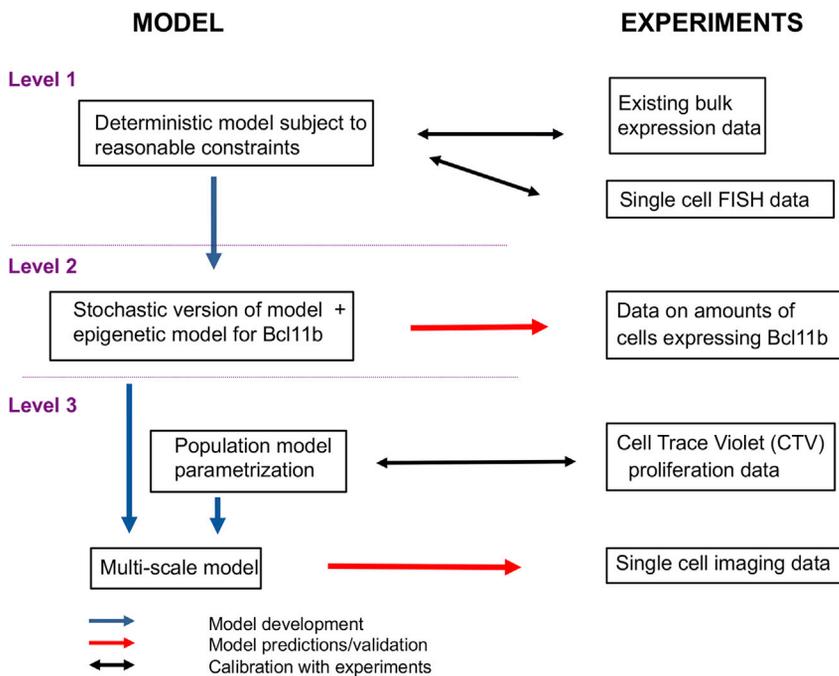


Figure 3. Modeling and Experimental Steps Emphasizing Predictions versus Calibrations

Level 1: with literature data and single-cell experiments, a GRN model is developed—both deterministic and stochastic versions. Level 2: the latter are augmented with an epigenetic model for *Bcl11b* expression. The resulting relative production rates of cells expressing *Bcl11b* are validated against bulk data. A population model is then developed using CTV experimental data. Level 3: this model is subsequently combined with the stochastic model from level 2 into a multi-level model that is validated against clonal imaging results.

accessibility (Yoshida et al., 2019) over 1 Mb of DNA flanking the locus, as summarized in Figure S3, along with published evidence for asynchronous activation of the two alleles in the same cells (Ng et al., 2018). Such state changes presumably are triggered by earlier TF actions but substantially affect later TF actions. While the roles for *Tcf7* and *Gata3* become dispensable by the stage immediately before *Bcl11b* activation, positive influences from Notch signaling and *Runx1* remain

are highly expressed, with a pronounced increase in *Tcf7*, while *PU.1* and *X* activity levels descend toward low values. The GRN model parameters (Table S1), optimized to the pseudo-time series, were picked and adjusted to accommodate the DN1-to-DN2 transition shown by single-cell RNA-FISH measurements (Figure S2) and were then frozen throughout the rest of this work. While the marker *CD25* plays an important role in interpreting the experimental results relative to the DN1-DN2 developmental transition, note that it is not a regulator and therefore was not explicitly included in these models.

Specification Level: Stochastic Model, Predicting Heterogeneity

Stochastic simulations (Gillespie, 1977) of the specification GRN model (Figure 4A) in individual cells with the same parameters (Table S1) showed that noise could be a source of heterogeneity in DN1 cells, as not all simulations exhibited dynamic switching toward a steady state with high T cell factors (Figure 4D). Of note, this heterogeneity is not a predetermined outcome as the system could have been in a dynamical region with high synchronization. However, such synchronization was not observed when using realistic parameter values determined from the FISH data. Moreover, the stochastic simulations showed that GRN intrinsic noise alone sometimes led to delays of the switch toward T cell commitment.

Level 2: Epigenetic Level Model

The second level of modeling was designed to account for the stochastic all-or-none activation timing observed for individual *Bcl11b* alleles, in terms of a simplified epigenetic model for the *Bcl11b* regulatory region with two possible states, open or closed (Figure 4C). Published data showing the extent of epigenetic changes at the *Bcl11b* locus during commitment reveal multiple increases in DNA looping (Hu et al., 2018) and chromatin

important. Notch signaling augments the likelihood of *Bcl11b* activation per unit time (Kueh et al., 2016), while *Runx1* positively modulates expression levels (Kueh et al., 2016; Ng et al., 2018). Therefore, our model places these regulators in opposition to an epigenetic resistance function *X*, which represents a composite of repressive chromatin state and any additional repressive TFs not yet defined (Figures 4A and 4C)

To account for the slow and potentially processive nature of this epigenetic transition, we used a model developed for calculating epigenetic activation as a function of the number of open (unmethylated) and closed (methylated) CpG dinucleotides at functional regulatory regions (Olariu et al., 2016; see STAR Methods). Here, it could apply to other probabilistic, repressive local chromatin states as well as CpG methylation. Notch signaling and *Runx1* levels obtained from stochastic GRN simulations serve as positive regulators for the probabilities that *Bcl11b*-regulating CpG sites (or equivalent chromatin features that are propagated by similar chromatin “reader-writer” mechanisms) become “open,” while *X* activity levels link to the probability of such CpGs remaining “closed.”

Merging Transcriptional with Epigenetic Levels: Predicting Time Delay Distributions

We next conducted stochastic simulations of the single-cell multi-level model incorporating *Bcl11b* regulatory region state dynamics predicting that early pro-T cells are heterogeneous in their CpG methylation states (Li et al., 2013) or equivalent open/closed chromatin states in the *Bcl11b* regulatory regions. The simulations showed that the *Bcl11b* locus can remain closed or become opened depending on the status of these CpG (or equivalent) sites, as dictated by Notch signaling, *Runx1* expression, and opposing *X* activity (Figures 4D and 4E). The stochastic treatment of the epigenetic level predicted further delays in T cell commitment. In an

Table 1. Gene Network and Epigenetic Model Components and Sources of Evidence

From	To	Effect	Evidence Type	Reference(s)	Index
Gene Regulatory Network Components					
Notch	Runx1	activate	modest upregulation in pro-T cells	Del Real and Rothenberg, 2013	1
Notch	TCF1	activate	direct binding and activation	Germar et al., 2011 ; Weber et al., 2011	2
Notch	Gata3	activate	developmental activation requirements and protection from repression, but no direct binding evidence	Hosoya-Ohmura et al., 2011 ; Schmitt et al., 2004 ; Taghon et al., 2005 ; Tydell et al., 2007	3
Runx1	Runx1	activate	auto-regulation seen in HSC precursors in embryo (“Cbfa2”); many binding sites; not certain in pro-T	North et al., 1999	4
Runx1	PU.1	repress	direct molecular evidence of repression through interaction sites	Hoogenkamp et al., 2007 ; Hosokawa et al., 2018 ; Huang et al., 2008 ; Zarnegar et al., 2010	5
PU.1	PU.1	activate	functionally only in myeloid cells, promoter activation and auto-regulation via cell cycle	Chen et al., 1995 ; Kueh et al., 2013 ; Leddin et al., 2011	6
PU.1	TCF1	repress	functional perturbation; possibly indirect due to paucity of binding	Champhekar et al., 2015 ; Del Real and Rothenberg, 2013 ; Franco et al., 2006	7
PU.1	Gata3	repress	functional perturbation; dependent on absence of Notch signal	Del Real and Rothenberg, 2013	8
Notch	Repression of Gata3 by PU.1	inhibit	functional perturbation	Del Real and Rothenberg, 2013	9
PU.1	Ability of Notch to activate TCF1	inhibit	functional perturbation	Del Real and Rothenberg, 2013 ; Franco et al., 2006	10
PU.1	Ability of Notch to activate Gata3	inhibit	functional perturbation	Del Real and Rothenberg, 2013	11
TCF1	Gata3	activate	gain- and loss-of-function perturbation in pro-T cells	Weber et al., 2011	12
Gata3	TCF1	activate	genetic perturbation and retroviral interference	García-Ojeda et al., 2013 ; Scripture-Adams et al., 2014	13
TCF1	X	repress	conjectural	Interpretation [Kueh et al., 2016 ; Ng et al., 2018]	14
Gata3	X	repress	conjectural	Interpretation [Kueh et al., 2016 ; Ng et al., 2018 ; Scripture-Adams et al., 2014]	15
TCF1	TCF1	activate	gain and loss of function perturbation in pro-T cells	Weber et al., 2011	16
X	Bcl11b	repress	epigenetic constraint on Bcl11b activation, apparently not relieved until after DN1 to DN2 transition	Kueh et al., 2016 ; Ng et al., 2018	17
Epigenetic Model Components					
Runx1	Bcl11b	open	positively regulates <i>Bcl11b</i> expression amplitude at single-cell level based on gain and loss-of-function perturbation; direct DNA binding	Guo et al., 2008 ; Hosokawa et al., 2018 ; Kueh et al., 2016	–
Notch	Bcl11b	open	accelerates and enhances frequency upregulation from DN2; primes for activation in DN1	Tydell et al., 2007 ; Kueh et al., 2016	–
X	Bcl11b	close	conjectural based on stage-dependent prohibition and inter-allelic asynchrony	Kueh et al., 2016 ; Ng et al., 2018	–

Cbfa-2, core-binding factor α -2; HSC, hematopoietic stem cell.

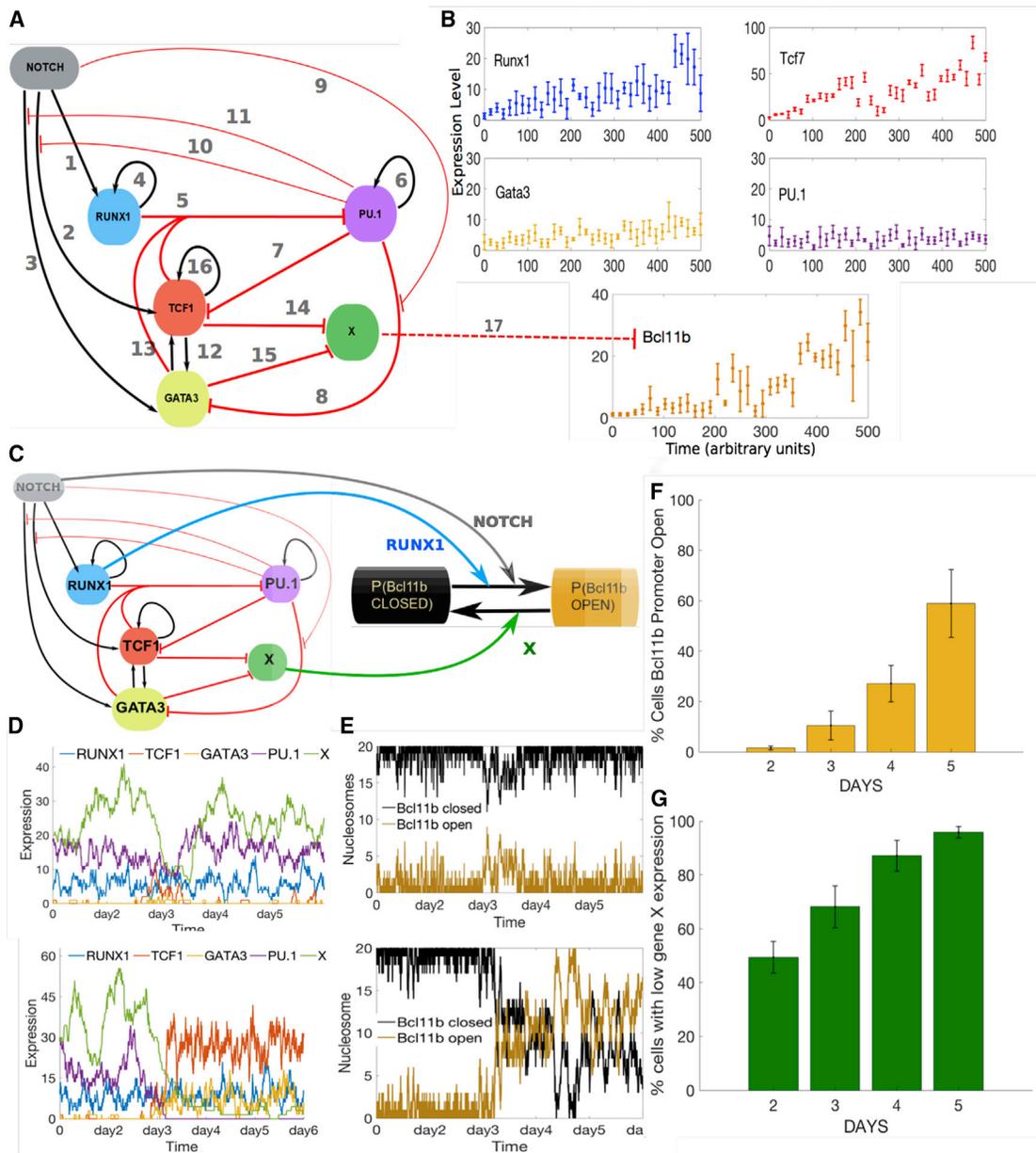


Figure 4. Core Circuit, Pseudo-Time-Series Data, Stochastic Simulations, and Multi-level Model Results

(A) GRN topology. The interaction indices correspond to the rightmost column in Table 1. The arrows and thick red lines are positive and negative DNA regulations, respectively. The thin red lines are PU.1 repression of Notch activation of TCF1 and Gata3 and Notch inhibition of PU.1 repression of Gata3. The dashed red line is the function X, causing epigenetic repression of Bcl11b.

(B) Pseudo-time series (mean and standard deviation) obtained from ordering the clusters given by the Gaussian mixture algorithm such that the Tcf7 (TCF1) and Bcl11b dynamics are similar to experimental behavior. Note the different y axis scales.

(C) Schematic representation of the epigenetic level model for the *Bcl11b* regulatory regions where Runx1 and Notch (blue and gray arrows) favor open DNA for transcription while X (green arrow) is keeping the *Bcl11b* regulatory regions closed.

(D) Stochastic simulation results obtained from transcriptional level model. In the first plot the system stays in a state controlled by PU.1 and X. In the second plot, the T cell factors are switched on, while PU.1 and X are downregulated.

(E) Stochastic simulations obtained from the model including an epigenetic level. In the first plot, the *Bcl11b* regulatory region remains closed, while in the second plot the regulatory region becomes open.

(F) Predicted distribution (yellow) of the percentage of single-cell multi-level model simulations with the outcome of *Bcl11b* regulatory region open.

(G) Predicted distribution (green) of the percentage of single-cell model simulations, where X activity levels have low values at time points corresponding to each day of the experiment (mean and standard deviations for three sets of 100 single-cell model simulations). (F and G) The model predicts a clear delay between the X loss of activity and *Bcl11b* regulatory region opening (e.g., at day 2 almost 50% of cells have lost X, while less of 10% are *Bcl11b* positive).

example when the switch was thrown by day 3 (i.e., Tcf7 and Gata3 becoming highly expressed, while X activity drops), the corresponding stochastic simulation of *Bcl11b* status showed that the regulatory system only opens after day 4 (Figure 4E, bottom).

Three sets of 100 stochastic simulations of this multi-level model predicted that on days 2 and 3, <20% of cells would express *Bcl11b*, ~40% on day 4, and ~50% on day 5 (Figure 4F). These predictions agree well with direct monitoring of Bcl11b-YFP during 5 days in DN1-derived populations going through commitment (Figure 1E). The distributions of cell percentages with low X (Figure 4G) are very similar to the distributions of CD25⁺ cells in experimental data (Figure 1E), suggesting that loss of X activity might be correlated with CD25 expression, and they are consistent with the observation that *Bcl11b* expression follows CD25 upregulation. Thus, the single-cell two-level model shows that stochastic mechanisms controlling eligibility for chromatin opening downstream of the core GRN can account for the observed delays in pro-T cell commitment.

Single-Cell Two-Level Model Predictions of Knockdown Effects

The two-level model predicts the effects of knocking down *Tcf7*, *Gata3*, and *Runx1*, which typically fall to 15%–25% of their original expression in RNA interference experiments. This corresponds to few molecular copies, thus requiring stochastic model simulations. When initialized in DN1 state, our model predicted that early knockdown of *Tcf7*, *Gata3*, or *Runx1* would halt progression toward a state of high T cell factors and low PU.1, resulting in a steady state with high PU.1 and low Tcf7, Gata3, and Runx1, keeping the *Bcl11b* locus closed (Figure S4A). By contrast, model simulations of the Tcf7 or Gata3 knockdown starting after X is inactivated predicted no continued requirement for Tcf7 or Gata3 (Figure S4B). These predictions were consistent with experimental data showing that for eventual *Bcl11b* expression, Tcf7 and Gata3 have stage-dependent roles and are much more important in the DN1 stage than in DN2 stage (Kueh et al., 2016). The concordance further supports the likelihood that X activity is lost after the DN1-to-DN2 transition.

Level 3: Population Model

We next developed population models based upon the CTV cell culture data (Figure 1). An initial hypothesis that cell-cycle length might be independent of generation was clearly incorrect (Table S3). We then relaxed this assumption, allowing each cell to have its division time stochastically determined from a normal distribution with mean and variance depending upon cell generation (Figure 5A). We chose the normal distribution for simplicity since the data would not disentangle this one from a more realistic long-tailed distribution. Normalized cell number distributions among different generations within simulated cell cultures up to day 5 were compared with the CTV data (Figure 1D), allowing us to determine the cell-cycle distribution parameters (Table S4). We monitored each cell generation and calculated the fraction of cells in each generation for each day from 172 simulations (Figure 5B), with resulting distributions very similar to experimental data (Figure 1D).

Full Multi-scale Model: Predicting Single Clone Developmental Dynamics

We next implemented the single-cell GRN model within the population mode to determine how cell proliferation affects popula-

tion distributions of fate-committed T cell versus non-committed cells. To take into account the impact of cell division on the *Bcl11b* locus epigenetic state, we used an advanced collaborative version of the epigenetic model with an intermediate state permitted between open and closed (STAR Methods). When the specification and epigenetic level models were embedded in this full three-level model, we found that a higher epigenetic barrier was required than in the two-level version, in order to match observed kinetics of *Bcl11b* activation. Even though the two-level model with 20 CpG sites gave very good results, when invoking the full multi-scale model, which takes into account the population growth, the observed kinetics was not matched (Figure 5C), and the number of CpG sites needed to be increased to 500 before results similar to observed data were achieved (Figure 5D; Figures S7G–S7I). In fact, this is consistent with the very extended potential regulatory regions associated with the *Bcl11b* gene (Hu et al., 2018; Li et al., 2013). Thus, at the population level, the most proliferative cells generate most of the Bcl11b⁺ cells, and this population bias toward proliferation of the most advanced cells initially underestimates the resistance to turning on *Bcl11b*, which is apparent at the single-cell level.

Full Multi-scale Model: Predicting Inter-clonal Kinetic Diversity

We computed the fraction of cells for which the *Bcl11b* regulatory region reached the open state in 172 simulations starting with one cell proliferating for 5 days. The model predicted that no cell has the *Bcl11b* regulatory region open at day 2, while at day 3 roughly 8% are open, increasing to 47% at day 4 and 60% at day 5 (Figure 5D). Our multi-scale model also predicts that X activity is lost in ~50% of the cells by day 2, and the proportion lacking X gradually increases, reaching >70% by day 5 (Figure 5D).

The multi-scale model predicted variability in proliferation kinetics among simulated single-cell clones (Figure 5E). The cell count increases over time for all clones. However, the cell numbers at each day of the experiment varied between clones (Figure 5E). Furthermore, the model predicted that the clones would not turn on Bcl11b and/or lose X activity in a fully synchronized way. Predictions indicated that 10/172 *in silico* “clones” produced 50% of *Bcl11b*⁺ cells as early as day 3, with others delaying until day 4, 5, or even later (Figure 5F). Similar results were obtained from monitoring the proportion of clones reaching 50% of cells without X activity (i.e., 95 of the simulated clones lost X activity at day 2, while 13 turned off X at day 5) (Figure 5G).

Experimental Results: Validating Data Clonal Kinetic Analysis of Differentiation from DN1

Finally, the multi-scale model predictions were tested by direct longitudinal imaging of DN1-derived T-lineage clones, through the DN1-to-DN2 transition and upregulation of *Bcl11b* (Figure 6). Data in Figure 1 had already shown population-level results, but because DN1s divide at different rates, output populations become dominated over time by progeny of the most rapidly proliferating input cells. By direct clonal monitoring, we could determine how well the three-level model predicted the distribution of behaviors corresponding to our model simulations,

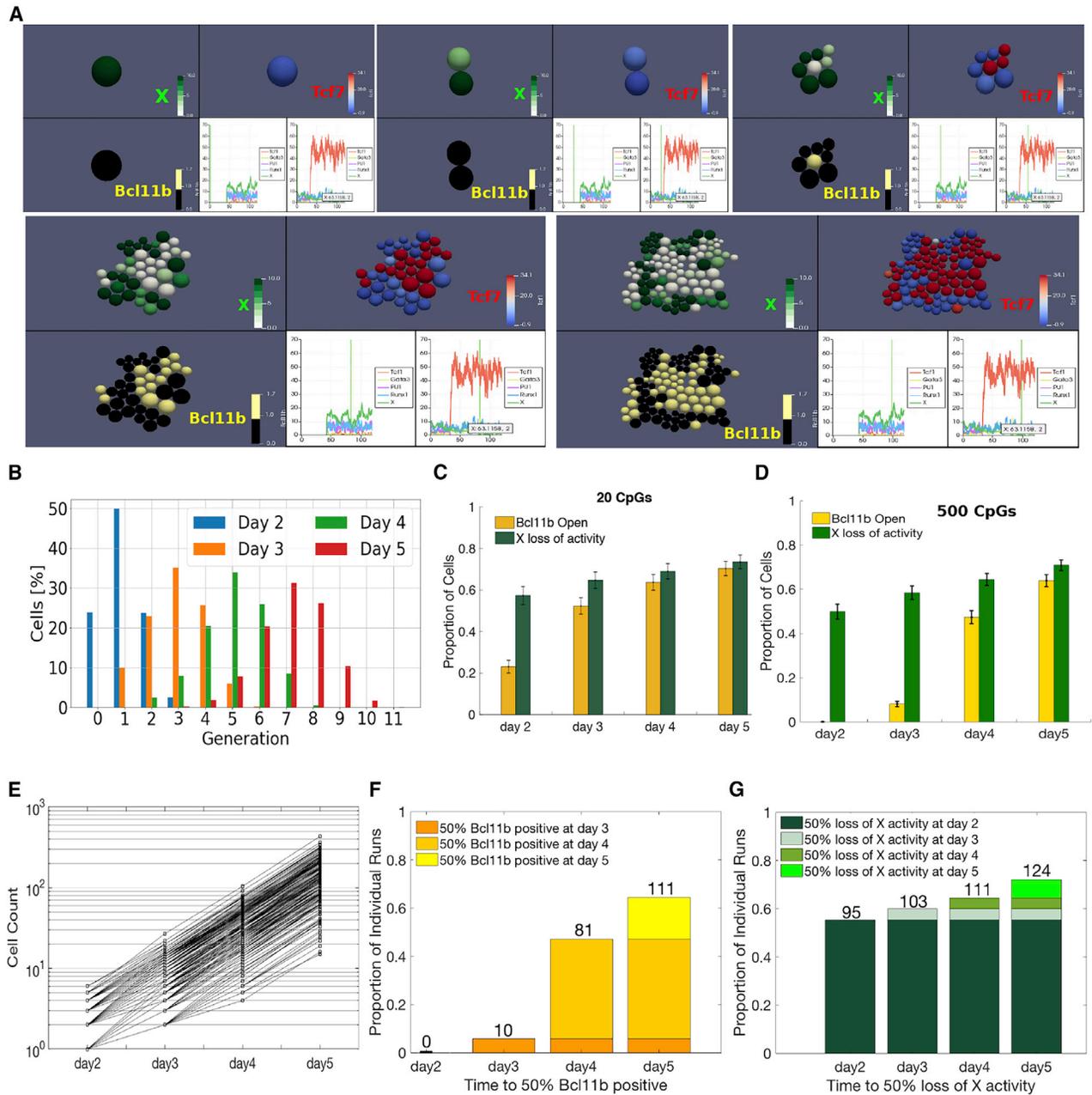


Figure 5. Full Multi-scale Model Predictions for Developmental Dynamics Starting from Individual DN1 Cells

(A) Snapshots of proliferation and differentiation kinetics of a single modeled DN1 cell at five time points. Each time point shows the following: top left: X function activity (green) impacting the epigenetic behavior of the *Bcl11b* regulatory regions; bottom left: cells with opened/closed regions are in gold/black, respectively; top right: *Tcf7* expression, which is predicted to be maximal during DN1-DN2a transition; and bottom right: expression timeline of all genes for two selected cells, one non-switching and one switching to DN2a, respectively.

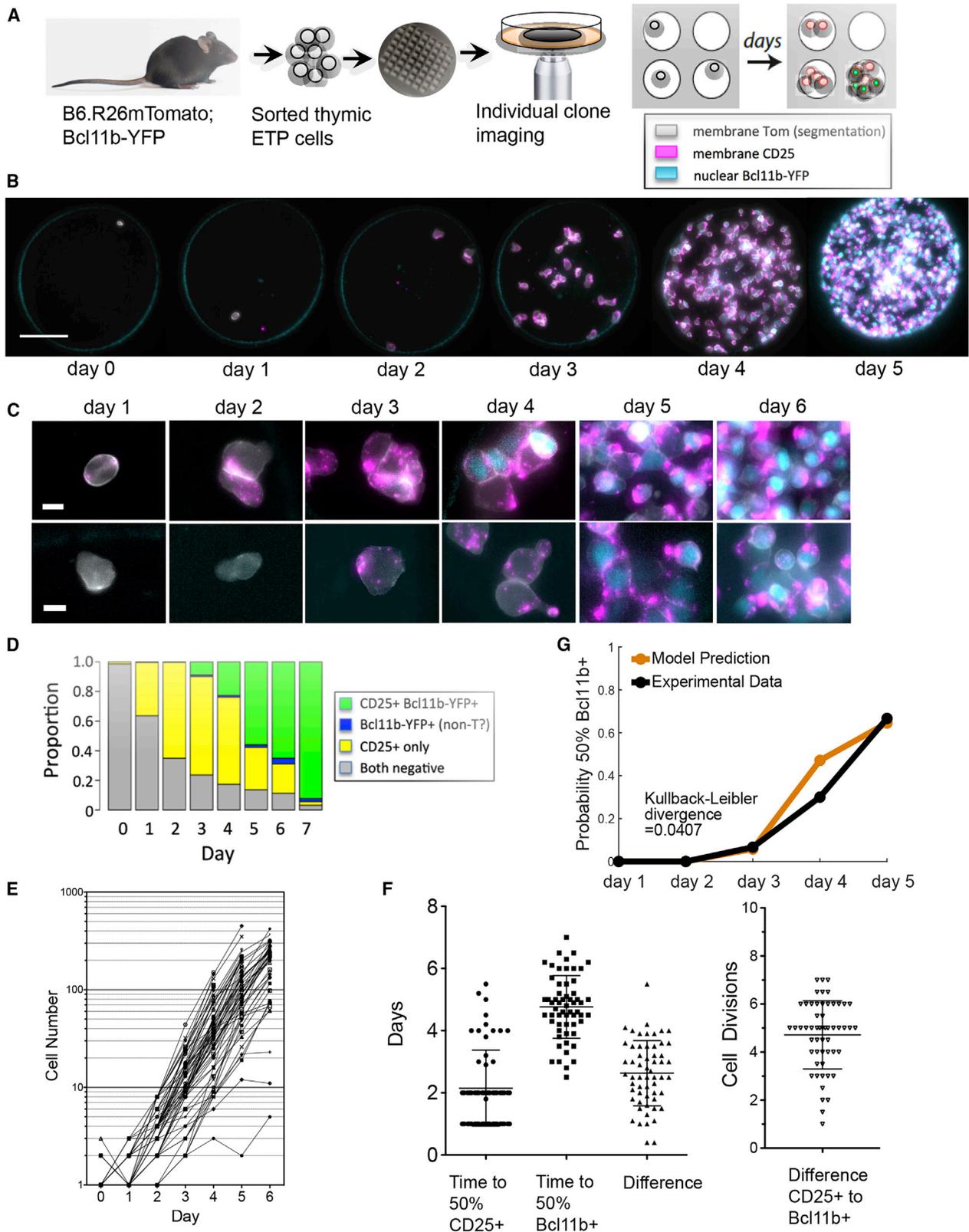
(B) Multi-scale model distributions of cell generations for each day.

(C and D) Multi-scale model predictions of the average fraction of cells with low X (green) and cells with the *Bcl11b* regulatory region open (yellow) at simulation times corresponding to days 2–5 of CTV and single-cell imaging experiments for epigenetic level model with 20 and 500 CpGs, respectively.

(E) Model predictions on variability in proliferation kinetics between clones.

(F) Multi-scale model predictions for the fraction of simulations versus required time to reach 50% *Bcl11b* regulatory region open in each run. The number of simulations reaching 50% at each day are shown on top.

(G) Multi-scale model predictions for the fraction of simulations versus time to reach 50% X activity loss. Also shown are the number of simulations reaching 50% at each day. The statistics were calculated from 172 model simulations.



(legend on next page)

starting with single clone founders and including cell divisions, GRN states, and *Bcl11b* epigenetic dynamics.

To determine the heterogeneity of DN1 cells and to resolve at a clonal level how consistently *Bcl11b* activation relates to absolute time or cell cycles, individual thymic DN1s were co-cultured with OP9-DL1 stroma in microwells and microscopically imaged over time (Kueh et al., 2016; Figure 6A). The DN1s were purified from thymi of mice with constitutively expressed membrane (m) Tomato, for cell segmentation and counts, as well as the nuclear *Bcl11b*-YFP reporter gene, and plated into microwells pre-seeded with OP9-DL1 cells. Microwells verified to start with single live DN1 cells were imaged daily for 6–7 days to determine clonal expansion and onsets of surface CD25 and nuclear *Bcl11b*-YFP expression for cells within individual clones. Merged false-color images of microwells in mTomato, CD25-Alexa Fluor 647, and *Bcl11b*-YFP fluorescence channels over time are shown in Figure 6B from one representative clone of 64 clones imaged. Two of 64 DN1 clones generated only non-T-lineage cells, possibly granulocytes, but the remainder turned on CD25 and most turned on *Bcl11b*-YFP within 7 days, indicating that they were indeed in the T lineage (Figure S5). Fluorescence thresholds for determining CD25 and *Bcl11b*-YFP positivity were calculated using background fluorescence estimates (Figure S6A). Higher magnification images for cells from two representative clones from days 1 to 6 (Figure 6C) show that surface CD25 (magenta) was turned on by days 1–3, with nuclear *Bcl11b*-YFP⁺ cells (cyan) appearing in some cells by day 4 or day 5.

Differentiation dynamics of CD25 and *Bcl11b* expression for all imaged wells combined are shown in Figure 6D, demonstrating comparable overall dynamics to bulk DN1 cultures (Figure 1E) and confirming that CD25 is turned on before *Bcl11b*. Based on cell counts, individual clones achieved similar proliferation rates in most cases, but with different lag times (Figure 6E), reflecting heterogeneity within the starting DN1 population, similar to model predictions (Figure 5E). At early time points, cells that remained DN1 had proliferated less than the CD25⁺ DN2s (Figure S6B), in agreement with CTV data (Figure 1C).

Wide variability in the timing of CD25 and *Bcl11b*-YFP expression between clones was observed (Figure 6F; Figure S5), and CD25 upregulation was not synchronous for progeny within

some clones (Figure S5), hinting at a stochastic element even in the early transition to DN2. Because of this asynchrony in expression of both CD25 and *Bcl11b* within clones, CD25 and *Bcl11b* positivity for each clone was scored as the time at which at least 50% of cells within a clone were CD25⁺ or *Bcl11b*⁺ (Figure 6F). These results show that most clones became $\geq 50\%$ CD25⁺ at days 1–2, although 22% of clones waited until days 3–5, while clones became $\geq 50\%$ *Bcl11b*-YFP⁺ between days 3 and 7, with most doing so on days 4–5 (Figure 6F; Figure S5).

These clonal data allow us to test the determinism of commitment, whether cells are programmed to turn on *Bcl11b* after a fixed period of time or fixed number of cell cycles after making the DN1-DN2 transition. Figure 6F shows the differences observed for individual clones between the times when $\geq 50\%$ of cells were scored as CD25⁺ and when $\geq 50\%$ were scored as *Bcl11b*-YFP⁺, in elapsed time and in estimated cell cycles. The results showed that the intervals between entering the DN2 state (CD25⁺) and commitment (*Bcl11b*-YFP⁺) were highly variable both in absolute times and in cell cycles, taking from <1 to >5 days and from one to seven cell divisions. Furthermore, *Bcl11b*-YFP was not turned on synchronously within a clone, generally requiring 2–3 days from expression in the first cells to expression in 100% of the cells (e.g., Figure 6C, top row, day 4; Figure S5). Thus, daughter cells from individual DN1 cells are not only heterogeneous in their clonal founders' differentiation states but also show a large stochastic element in timing of *Bcl11b* upregulation even after entering the DN2 state (Ng et al., 2018).

The *Bcl11b* expression dynamics in these clonal cultures appear to be in close accord with our multi-scale model predictions (Figures 5D and 6F), so to compare the model predictions directly with the experimental data, in Figure 6G we show the probability distributions of reaching 50% *Bcl11b*-YFP⁺ predicted by model simulations with the experimental results (Figure 5F). These distributions appear to be very similar, and the Kullback-Leibler divergence value calculated between them has a very low value of 0.0407 (a value of 0 indicates that the two distributions are identical). Furthermore, rates of CD25 expression gave comparable dynamics to the loss of X activity predicted by the model (Figures 5D and 6D), suggesting a potential relationship between these parameters.

Figure 6. Kinetic Analysis of Proliferation and Differentiation for Individual DN1 Clones

- (A) DN1 clonal imaging. Thymic DN1 cells were purified from mice homozygous for mTomato and the nuclear *Bcl11b*-YFP reporter, added to microwells pre-seeded with OP9-DL1 stroma, and cultured with cytokines IL-7 and Flt3L and CD25-Alexa Fluor 647 antibodies (marking progression to DN2). Individual microwells with one DN1 cell were imaged daily from day 0 to day 6 or 7.
- (B) Merged false-color images of one DN1 clone for days 0–5: mTomato (gray), surface CD25 (magenta), and nuclear *Bcl11b*-YFP (cyan). Exposures were adjusted to allow visualization of cells in all channels. Scale bar represents 100 μm .
- (C) Higher magnification merged false-color images of selected cells from two DN1 clones showing the kinetics of CD25 (magenta) and *Bcl11b*-YFP (cyan) expression. Scale bars represent 10 μm .
- (D) Combined fractions of cells expressing CD25 and *Bcl11b*-YFP, alone or together over time. $n = 67$ (including 62 T-lineage clones, 2 non-T clones, and 3 non-clonal wells).
- (E) Cell numbers counted from daily images of 62 clonal T-lineage wells.
- (F) (Left) Timing of CD25 and *Bcl11b*-YFP expression scored as days until $>50\%$ of cells in individual clones were CD25⁺ ($n = 62$ clones) or *Bcl11b*-YFP⁺ ($n = 58$) and the difference in time between the two events for the clones that turned on both markers ($n = 58$). (Right) The difference in cell cycles for individual clones between 50% of cells turning on CD25 and 50% turning on *Bcl11b*-YFP ($n = 58$). Each symbol represents a single clone. Means and standard deviations are indicated with lines.
- (G) The probability of a clone reaching 50% *Bcl11b*-YFP⁺ on each day predicted by the multi-scale model simulations compared directly with the results obtained from the DN1 clone experimental data. The Kullback-Leibler divergence calculated between them has a very low value, confirming the similarity.

Thus, DN1 clonal heterogeneity in differentiation potential and stochasticity in the execution of commitment dynamics, as predicted by the three-level model, was validated by the kinetics and the intra- and inter-clonal variation revealed by experimental clonal monitoring.

DISCUSSION

We have developed, explored, and validated an integrated computational multi-scale model for early T cell development and commitment kinetics. Commitment of multipotent precursors to the T cell fate is particularly accessible for modeling because, at the single-cell level, commitment corresponds to *Bcl11b* activation. The model explains commitment kinetics in terms of the gene network circuitry controlling *Bcl11b* expression by incorporating results from both single-cell- and population-level experiments elucidating the process on different scales. Although the process has a core of classic transcriptional circuitry, the output of the process was recently shown to be substantially modulated by epigenetic resistance to transition (Ng et al., 2018), and, as shown here, by nonuniform cell proliferation as differentiation proceeds. A transcriptional gene network model optimized to ignore these additional factors would necessarily be inaccurate. Thus, after a transcriptional first level, our model has built in a second level of stochastic epigenetic control to predict chromatin opening. Finally, the two-step gene control model is integrated into a cell growth population model to predict how the internal molecular events generate measured population behavior.

The model delivers several validations and predictions using transcriptional and population model parameters set by fitting gene network parameters to experimental RNA expression data and population growth parameters to experimental proliferation data.

DN1 population developmental heterogeneity can arise solely from GRN noise, although we cannot rule out an additional role for external noise.

The heterogeneous delay in lineage commitment marked by *Bcl11b* expression arises from both GRN and epigenetic variability.

Effects of knockdown of *Tcf7* and *Gata3* are substantially different before and after *X* downregulation, consistent with experimentally observed differences between effects of disrupting these inputs at DN1 and DN2 stages.

DN1 single clone commitment and proliferation kinetics resulting from GRN, epigenetic, and population dynamics are validated by clonal cell cultures.

The high *Bcl11b* epigenetic barrier is partly masked by bulk population growth.

Computational models for epigenetics and GRNs have been proposed in the context of pluripotency acquisition through cell reprogramming (Artyomov et al., 2010; Olariu et al., 2016). Multi-scale modeling approaches, unifying observations from the intra-cellular to cell population scale, have also been used to model heterogeneity and function of mature peripheral T cells (for review, see Carbo et al., 2014). Whereas other he-

matopoietic stem cell commitment processes have been subject to mathematical model investigations (for review, see Olariu and Peterson, 2019), few attempts have been made for early lymphocyte development (Collombet et al., 2017; Manesso et al., 2016). The choice to use continuous-valued rather than Boolean transcriptional models here has the advantage of avoiding discretizing the data in instances when experimental real-valued gene expression levels are relevant for the problem studied, as in this case. Whereas a Boolean approach could probe larger networks than that modeled here (for review, see Olariu and Peterson, 2019), previous investigations of important regulators of *Bcl11b* enable us to focus on a few key genes only.

Importantly, our model incorporates the complex molecular mechanisms shown recently to be involved in *Bcl11b* activation including the four known positive regulators and the additional, slow epigenetic step that creates a stochastic delay in *Bcl11b* activation, long after the positive regulators are present. This level of control is probably not unique to *Bcl11b* regulation but has been demonstrated in particular depth for developmental activation of this gene (Ng et al., 2018). The design of the stochastic epigenetic level of the network is novel, incorporating some features not usually included in GRNs, in order to encompass the known mechanisms involved. Published experimental measurements indicate that a default repressive chromatin state is probably a major contributor to the stochastic timing of activation, with a delay from the activation of one allele to the other—a measure of noise in time to relieve an epigenetic barrier—on the order of days (Ng et al., 2018). We designed this negative function (*X*) to oppose the stepwise chromatin opening that is a prerequisite for *Bcl11b* activation. This strongly affects the predicted developmental timing control in the model as it does in experimental observation. For parsimony, we have proposed an architecture in which the positive regulators in the first transcriptional layer of the model work toward downregulating *X*, leaving the second stochastic layer to determine when *Bcl11b* will respond by activation. In the complete three-layer implementation of the model, the predicted kinetics of *X* downregulation in individual clones match well with the observed kinetics of DN1-to-DN2 developmental progression before *Bcl11b* activation in individual clones, as detected by the cell surface marker CD25. Similar timing of *X* downregulation and DN1-to-DN2 progression would also fit the distinct effects of *Tcf7* or *Gata3* knockdown before and after *X* downregulation (*in silico*) and before and after DN1 to DN2 transition (*in vitro*), thus explaining the experimental observation that *Bcl11b* is activated only after DN1-to-DN2 progression, despite the presence of all its positive regulators in DN1 stage.

In this model *X* is a composite function, not a single undefined regulatory gene. As *Bcl11b* starts out in an epigenetically silent state inherited from hematopoietic stem cells, *X* measures the resistance to the opening of relevant *cis*-regulatory elements at each allele of the locus. In the stochastic level of the model, we have taken the balance of Notch signals and Runx1 against decreasing *X* to drive a stepwise epigenetic opening process that we have modeled analogously to the removal of methylated CpGs. While demethylation may be part of this process, the use of this formalism is more abstract here and indicates the stochastic but neighbor-biased and progressive nature that

is shared by demethylation with other epigenetic opening processes.

In different DN1 clones progressing toward T cell commitment, the timing of CD25 expression (DN1-to-DN2 transition, approximating X downregulation) is quite heterogeneous, confirming DN1 cell heterogeneity (Manesso et al., 2013; Yui and Rothenberg, 2014). If this were reflecting a pipeline of distinct developmental sub-stages within DN1, the cells might actually be starting with different initial regulatory states. However, stochastic simulations of the GRN controlling the first steps in T cell commitment show that the switch toward a state with low X levels does not occur in a synchronized manner between model simulations with identical initial conditions. This suggests that GRN intrinsic noise could be another source of heterogeneity in timing of the loss of X. The explicitly stochastic basis of progressive locus opening that follows X downregulation, in the epigenetic level of the model, is further supported by the high variability of *Bcl11b* activation timing in different clones even after they transition to DN2.

The very similar results obtained from the single-cell model and clonal cell cultures suggest a high level of heterogeneity in critical GRN components during the DN1-to-DN2 stages. The experimental and model population kinetics show an additional level of proliferative heterogeneity between these stages that is apparent at the single-cell level, but much less apparent in bulk populations. The general increase in proliferation rate as cells transition to DN2 not only biases population phenotypes to favor the fastest differentiating clones but also might feed back on the network by dilution of inherited proteins affecting the potential speed of epigenetic changes (Haerter et al., 2014; Kueh et al., 2013). The need to invoke a *Bcl11b*-regulating function, X, in the GRN and epigenetic models, to achieve proper model behavior, calls for identifying critical *cis*- as well as *trans*-acting GRN elements. Work to identify these factors is currently underway.

In summary, this new model, that integrates three distinct types of mechanistic control, provides a quantitative understanding of commitment kinetic requirements of early T cell precursors and a valuable *in silico* tool to help investigate new regulatory factors and processes and their roles in lineage commitment.

STAR★METHODS

Detailed methods are provided in the online version of this paper and include the following:

- KEY RESOURCES TABLE
- RESOURCE AVAILABILITY
 - Lead Contact
 - Materials Availability
 - Data and Code Availability
- EXPERIMENTAL MODEL AND SUBJECT DETAILS
 - Animals
 - Cell lines
- METHOD DETAILS
 - DN thymocyte purification
 - Single Molecule Multiplex Fluorescent *in situ* Hybridization (sm-FISH)

- Bulk cell cultures
- Clonal Image Analysis
- Modeling: Pseudo-time series
- Single Cell Model: Transcription factor Level
- Single Cell Model: Epigenetic Level
- Multi-scale model of proliferation and gene expression
- A simplified model
- Multi-scale Model: Parameters Collaborative Epigenetic Model
- QUANTIFICATION AND STATISTICAL ANALYSIS

SUPPLEMENTAL INFORMATION

Supplemental Information can be found online at <https://doi.org/10.1016/j.celrep.2020.108622>.

ACKNOWLEDGMENTS

The authors thank Dr. Long Cai for support for the smFISH analysis; Dr. Jeffrey Longmate for data analysis; Dr. Hao Yuan Kueh for helpful discussions and advice on imaging and analysis; Kenneth Ng for technical help; Diana Perez, Jamie Tijerina, and Rochelle Diamond of the Caltech Flow Cytometry Facility for fluorescence-activated cell sorting (FACS); and Dr. Andreas Collazo and the Caltech Biological Imaging Facility for microscopy assistance. The authors gratefully acknowledge the support of the US National Institutes of Health (USPHS grant R01HL119102 to E.V.R. and C.P.) and the Albert Billings Ruddock Professorship (to E.V.R.).

AUTHOR CONTRIBUTIONS

V.O., M.A.Y., E.V.R., and C.P. designed the study. V.O., M.A.Y., E.V.R., and C.P. wrote most of the manuscript. M.A.Y. performed the CTV and kinetics experiments. W.Z. performed the FISH experiments and wrote part of the manuscript. V.O. developed the transcriptional and epigenetic models and analyzed the data. P.K. developed the population model. E.A. conducted parameter optimization and confidence bounds calculations. J.D. implemented the pseudo-time-series and multi-scale models.

DECLARATION OF INTERESTS

The authors declare no competing interests.

Received: October 15, 2019

Revised: April 24, 2020

Accepted: December 18, 2020

Published: January 12, 2021

REFERENCES

- Ackers, G.K., Johnson, A.D., and Shea, M.A. (1982). Quantitative model for gene regulation by λ phage repressor. *Proc. Natl. Acad. Sci. USA* 79, 1129–1133.
- Anderson, M.K., Weiss, A.H., Hernandez-Hoyos, G., Dionne, C.J., and Rothenberg, E.V. (2002). Constitutive expression of PU.1 in fetal hematopoietic progenitors blocks T cell development at the pro-T cell stage. *Immunity* 16, 285–296.
- Artyomov, M.N., Meissner, A., and Chakraborty, A.K. (2010). A model for genetic and epigenetic regulatory networks identifies rare pathways for transcription factor induced pluripotency. *PLoS Comput. Biol.* 6, e1000785.
- 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.

- Byrd, R.H., Lu, P.H., Nocedal, J., and Zhu, C.Y. (1995). A Limited Memory Algorithm for Bound Constrained Optimization. *SIAM J. Sci. Comput.* *16*, 1190–1208.
- Carbo, A., Hontecillas, R., Andrew, T., Eden, K., Mei, Y., Hoops, S., and Basaganya-Riera, J. (2014). Computational modeling of heterogeneity and function of CD4+ T cells. *Front. Cell Dev. Biol.* *2*, 31.
- Champhekar, A., Damle, S.S., Freedman, G., Carotta, S., Nutt, S.L., and Rothenberg, E.V. (2015). Regulation of early T-lineage gene expression and developmental progression by the progenitor cell transcription factor PU.1. *Genes Dev.* *29*, 832–848.
- Chen, H., Ray-Gallet, D., Zhang, P., Hetherington, C.J., Gonzalez, D.A., Zhang, D.E., Moreau-Gachelin, F., and Tenen, D.G. (1995). PU.1 (Spi-1) autoregulates its expression in myeloid cells. *Oncogene* *11*, 1549–1560.
- Collombet, S., van Oevelen, C., Sardina Ortega, J.L., Abou-Jaoudé, W., Di Stefano, B., Thomas-Chollier, M., Graf, T., and Thieffry, D. (2017). Logical modeling of lymphoid and myeloid cell specification and transdifferentiation. *Proc. Natl. Acad. Sci. USA* *114*, 5792–5799.
- David-Fung, E.S., Butler, R., Buzi, G., Yui, M.A., Diamond, R.A., Anderson, M.K., Rowen, L., and Rothenberg, E.V. (2009). Transcription factor expression dynamics of early T-lymphocyte specification and commitment. *Dev. Biol.* *325*, 444–467.
- 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.
- De Obaldia, M.E., and Bhandoola, A. (2015). Transcriptional regulation of innate and adaptive lymphocyte lineages. *Annu. Rev. Immunol.* *33*, 607–642.
- Dionne, C.J., Tse, K.Y., Weiss, A.H., Franco, C.B., Wiest, D.L., Anderson, M.K., and Rothenberg, E.V. (2005). Subversion of T lineage commitment by PU.1 in a clonal cell line system. *Dev. Biol.* *280*, 448–466.
- Franco, C.B., Scripture-Adams, D.D., Proekt, I., Taghon, T., Weiss, A.H., Yui, M.A., Adams, S.L., Diamond, R.A., and Rothenberg, E.V. (2006). Notch/Delta signaling constrains reengineering of pro-T cells by PU.1. *Proc. Natl. Acad. Sci. USA* *103*, 11993–11998.
- García-Ojeda, M.E., Klein Wolterink, R.G., Lemaître, F., Richard-Le Goff, O., Hasan, M., Hendriks, R.W., Cumano, A., and Di Santo, J.P. (2013). GATA-3 promotes T-cell specification by repressing B-cell potential in pro-T cells in mice. *Blood* *121*, 1749–1759.
- Germar, K., Dose, M., Konstantinou, T., Zhang, J., Wang, H., Lobry, C., Arnett, K.L., Blacklow, S.C., Aifantis, I., Aster, J.C., and Gounari, F. (2011). T-cell factor 1 is a gatekeeper for T-cell specification in response to Notch signaling. *Proc. Natl. Acad. Sci. USA* *108*, 20060–20065.
- Gillespie, D.T. (1977). Exact Stochastic Simulation of Coupled Chemical-Reactions. *J Phys Chem.* *81*, 2340–2361.
- Guo, Y., Maillard, I., Chakraborti, S., Rothenberg, E.V., and Speck, N.A. (2008). Core binding factors are necessary for natural killer cell development and cooperate with Notch signaling during T-cell specification. *Blood* *112*, 480–492.
- Haerter, J.O., Lövkvist, C., Dodd, I.B., and Sneppen, K. (2014). Collaboration between CpG sites is needed for stable somatic inheritance of DNA methylation states. *Nucleic Acids Res.* *42*, 2235–2244.
- Hoogenkamp, M., Kryszinska, H., Ingram, R., Huang, G., Barlow, R., Clarke, D., Ebralidze, A., Zhang, P., Tagoh, H., Cockerill, P.N., et al. (2007). The *Pu.1* locus is differentially regulated at the level of chromatin structure and noncoding transcription by alternate mechanisms at distinct developmental stages of hematopoiesis. *Mol. Cell. Biol.* *27*, 7425–7438.
- Hosokawa, H., Ungerback, J., Wang, X., Matsumoto, M., Nakayama, K.I., Cohen, S.M., Tanaka, T., and Rothenberg, E.V. (2018). Transcription factor PU.1 represses and activates gene expression in early T cells by redirecting partner transcription factor binding. *Immunity* *48*, 1119–1134.e7.
- Hosokawa, H., Romero-Wolf, M., Yang, Q., Motomura, Y., Levanon, D., Groner, Y., Moro, K., Tanaka, T., and Rothenberg, E.V. (2020). Cell type-specific actions of Bcl11b in early T-lineage and group 2 innate lymphoid cells. *J. Exp. Med.* *217*, e20190972.
- Hosoya-Ohmura, S., Lin, Y.H., Herrmann, M., Kuroha, T., Rao, A., Moriguchi, T., Lim, K.C., Hosoya, T., and Engel, J.D. (2011). An NK and T cell enhancer lies 280 kilobase pairs 3' to the *gata3* structural gene. *Mol. Cell. Biol.* *31*, 1894–1904.
- Hu, G., Cui, K., Fang, D., Hirose, S., Wang, X., Wangsa, D., Jin, W., Ried, T., Liu, P., Zhu, J., et al. (2018). Transformation of accessible chromatin and 3D nucleosome underlies lineage commitment of early T cells. *Immunity* *48*, 227–242.e8.
- Huang, G., Zhang, P., Hirai, H., Elf, S., Yan, X., Chen, Z., Koschmieder, S., Okuno, Y., Dayaram, T., Grownney, J.D., et al. (2008). PU.1 is a major downstream target of AML1 (RUNX1) in adult mouse hematopoiesis. *Nat. Genet.* *40*, 51–60.
- Ikawa, T., Hirose, S., Masuda, K., Kakugawa, K., Satoh, R., Shibano-Satoh, A., Kominami, R., Katsura, Y., and Kawamoto, H. (2010). An essential developmental checkpoint for production of the T cell lineage. *Science* *329*, 93–96.
- Isoda, T., Moore, A.J., He, Z., Chandra, V., Aida, M., Denholtz, M., Piet van Hamburg, J., Fisch, K.M., Chang, A.N., Fahl, S.P., et al. (2017). Non-coding Transcription Instructs Chromatin Folding and Compartmentalization to Dictate Enhancer-Promoter Communication and T Cell Fate. *Cell* *171*, 103–119.e18.
- Ji, H., Ehrlich, L.I., Seita, J., Murakami, P., Doi, A., Lindau, P., Lee, H., Aryee, M.J., Irizarry, R.A., Kim, K., et al. (2010). Comprehensive methylome map of lineage commitment from haematopoietic progenitors. *Nature* *467*, 338–342.
- Krupinski, P., Chickarmane, V., and Peterson, C. (2012). Computational multi-scale modeling of embryo development. *Curr. Opin. Genet. Dev.* *22*, 613–618.
- Kueh, H.Y., Champhekar, A., Nutt, S.L., Elowitz, M.B., and Rothenberg, E.V. (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., Pease, S.S., Zhang, J.A., Damle, S.S., Freedman, G., Siu, S., Bernstein, I.D., Elowitz, M.B., and Rothenberg, E.V. (2016). Asynchronous combinatorial action of four regulatory factors activates *Bcl11b* for T cell commitment. *Nat. Immunol.* *17*, 956–965.
- Leddin, M., Perrod, C., Hoogenkamp, M., Ghani, S., Assi, S., Heinz, S., Wilson, N.K., Follows, G., Schönheit, J., Vockentanz, L., et al. (2011). Two distinct auto-regulatory loops operate at the PU.1 locus in B cells and myeloid cells. *Blood* *117*, 2827–2838.
- Li, L., Leid, M., and Rothenberg, E.V. (2010a). An early T cell lineage commitment checkpoint dependent on the transcription factor Bcl11b. *Science* *329*, 89–93.
- Li, P., Burke, S., Wang, J., Chen, X., Ortiz, M., Lee, S.C., Lu, D., Campos, L., Goulding, D., Ng, B.L., et al. (2010b). Reprogramming of T cells to natural killer-like cells upon Bcl11b deletion. *Science* *329*, 85–89.
- Li, L., Zhang, J.A., Dose, M., Kueh, H.Y., Mosadeghi, R., Gounari, F., and Rothenberg, E.V. (2013). A far downstream enhancer for murine Bcl11b controls its T-cell specific expression. *Blood* *122*, 902–911.
- Longabaugh, W.J.R., Zeng, W., Zhang, J.A., Hosokawa, H., Jansen, C.S., Li, L., Romero-Wolf, M., Liu, P., Kueh, H.Y., Mortazavi, A., and Rothenberg, E.V. (2017). Bcl11b and combinatorial resolution of cell fate in the T-cell gene regulatory network. *Proc. Natl. Acad. Sci. USA* *114*, 5800–5807.
- Manesso, E., Chickarmane, V., Kueh, H.Y., Rothenberg, E.V., and Peterson, C. (2013). Computational modelling of T-cell formation kinetics: output regulated by initial proliferation-linked deferral of developmental competence. *J. R. Soc. Interface* *10*, 20120774.
- Manesso, E., Kueh, H.Y., Freedman, G., Rothenberg, E.V., and Peterson, C. (2016). Irreversibility of T-cell specification: Insights from computational modelling of a minimal network architecture. *PLoS ONE* *11*, e0161260.
- Mingueneau, M., Kreslavsky, T., Gray, D., Heng, T., Cruse, R., Ericson, J., Bendall, S., Spitzer, M.H., Nolan, G.P., Kobayashi, K., et al.; Immunological Genome Consortium (2013). The transcriptional landscape of $\alpha\beta$ T cell differentiation. *Nat. Immunol.* *14*, 619–632.
- Ng, K.K.H., Yui, M.A., Mehta, A., Siu, S., Irwin, B., Pease, S., Hirose, S., Elowitz, M.B., Rothenberg, E.V., and Kueh, H.Y. (2018). A stochastic epigenetic switch controls the dynamics of T-cell lineage commitment. *eLife* *7*, e37851.

- North, T., Gu, T.L., Stacy, T., Wang, Q., Howard, L., Binder, M., Marín-Padilla, M., and Speck, N.A. (1999). *Cbfa2* is required for the formation of intra-aortic hematopoietic clusters. *Development* *126*, 2563–2575.
- Olariu, V., and Peterson, C. (2019). Kinetic models of hematopoietic differentiation. *Wiley Interdiscip. Rev. Syst. Biol. Med.* *11*, e1424.
- Olariu, V., Lökvist, C., and Sneppen, K. (2016). *Nanog*, *Oct4* and *Tet1* interplay in establishing pluripotency. *Sci. Rep.* *6*, 25438.
- Pedregosa, F., Varoquaux, G., Gramfort, A., Michel, V., Thirion, B., Grisel, O., Blondel, M., Prettenhofer, P., Weiss, R., Dubourg, V., et al. (2011). Scikit-learn: Machine Learning in Python. *J. Mach. Learn. Res.* *12*, 2825–2830.
- Raue, A., Schilling, M., Bachmann, J., Matteson, A., Schelker, M., Kaschek, D., Hug, S., Kreutz, C., Harms, B.D., Theis, F.J., et al. (2013). Lessons learned from quantitative dynamical modeling in systems biology. *PLoS ONE* *8*, e74335.
- Rothenberg, E.V., Ungerback, J., and Champhekar, A. (2016). Forging T-Lymphocyte Identity: Intersecting Networks of Transcriptional Control. *Adv. Immunol.* *129*, 109–174.
- Schmitt, T.M., and Zúñiga-Pflücker, J.C. (2002). Induction of T cell development from hematopoietic progenitor cells by delta-like-1 in vitro. *Immunity* *17*, 749–756.
- Schindelin, J., Arganda-Carreras, I., Frise, E., Kaynig, V., Longair, M., Pietzsch, T., Preibisch, S., Rueden, C., Saalfeld, S., Schmid, B., et al. (2012). Fiji: an open-source platform for biological-image analysis. *Nature Methods* *9* (7), 676–682. <https://doi.org/10.1038/nmeth.2019>.
- Schmitt, T.M., Ciofani, M., Petrie, H.T., and Zúñiga-Pflücker, J.C. (2004). Maintenance of T cell specification and differentiation requires recurrent notch receptor-ligand interactions. *J. Exp. Med.* *200*, 469–479.
- Scripture-Adams, D.D., Damle, S.S., Li, L., Elihu, K.J., Qin, S., Arias, A.M., Butler, R.R., 3rd, Champhekar, A., Zhang, J.A., and Rothenberg, E.V. (2014). GATA-3 dose-dependent checkpoints in early T cell commitment. *J. Immunol.* *193*, 3470–3491.
- Shah, S., Lubeck, E., Zhou, W., and Cai, L. (2016). In situ transcription profiling of single cells reveals spatial organization of cells in the mouse hippocampus. *Neuron* *92*, 342–357.
- Taghon, T.N., David, E.S., Zúñiga-Pflücker, J.C., and Rothenberg, E.V. (2005). Delayed, asynchronous, and reversible T-lineage specification induced by Notch/Delta signaling. *Genes Dev.* *19*, 965–978.
- 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.
- Tydell, C.C., David-Fung, E.S., Moore, J.E., Rowen, L., Taghon, T., and Rothenberg, E.V. (2007). Molecular dissection of prethymic progenitor entry into the T lymphocyte developmental pathway. *J. Immunol.* *179*, 421–438.
- Ungerback, J., Hosokawa, H., Wang, X., Strid, T., Williams, B.A., Sigvardsson, M., and Rothenberg, E.V. (2018). Pioneering, chromatin remodeling, and epigenetic constraint in early T-cell gene regulation by SPI1 (PU.1). *Genome Res.* *28*, 1508–1519.
- Weber, B.N., Chi, A.W., Chavez, A., Yashiro-Ohtani, Y., Yang, Q., Shestova, O., and Bhandoola, A. (2011). A critical role for TCF-1 in T-lineage specification and differentiation. *Nature* *476*, 63–68.
- Xu, W., Carr, T., Ramirez, K., McGregor, S., Sigvardsson, M., and Kee, B.L. (2013). E2A transcription factors limit expression of Gata3 to facilitate T lymphocyte lineage commitment. *Blood* *121*, 1534–1542.
- Ye, Y., Kang, X., Bailey, J., Li, C., and Hong, T. (2019). An enriched network motif family regulates multistep cell fate transitions with restricted reversibility. *PLoS Comput. Biol.* *15*, e1006855.
- Yoshida, H., Lareau, C.A., Ramirez, R.N., Rose, S.A., Maier, B., Wroblewska, A., Desland, F., Chudnovskiy, A., Mortha, A., Dominguez, C., et al.; Immunological Genome Project (2019). The cis-Regulatory Atlas of the Mouse Immune System. *Cell* *176*, 897–912.e20.
- Yui, M.A., and Rothenberg, E.V. (2014). Developmental gene networks: a triathlon on the course to T cell identity. *Nat. Rev. Immunol.* *14*, 529–545.
- Yui, M.A., Feng, N., and Rothenberg, E.V. (2010). Fine-scale staging of T cell lineage commitment in adult mouse thymus. *J. Immunol.* *185*, 284–293.
- Zarnegar, M.A., Chen, J., and Rothenberg, E.V. (2010). Cell-type-specific activation and repression of PU.1 by a complex of discrete, functionally specialized cis-regulatory elements. *Mol. Cell. Biol.* *30*, 4922–4939.
- Zhang, J.A., Mortazavi, A., Williams, B.A., Wold, B.J., and Rothenberg, E.V. (2012). Dynamic transformations of genome-wide epigenetic marking and transcriptional control establish T cell identity. *Cell* *149*, 467–482.
- Zhou, X., Yu, S., Zhao, D.M., Harly, J.T., Badovinac, V.P., and Xue, H.H. (2010). Differentiation and persistence of memory CD8(+) T cells depend on T cell factor 1. *Immunity* *33*, 229–240.
- Ziętara, N., Łyszkiewicz, M., Puchalka, J., Witzlau, K., Reinhardt, A., Förster, R., Pabst, O., Prinz, I., and Krueger, A. (2015). Multicongenetic fate mapping quantification of dynamics of thymus colonization. *J. Exp. Med.* *212*, 1589–1601.
- Zlotoff, D.A., and Bhandoola, A. (2011). Hematopoietic progenitor migration to the adult thymus. *Ann. N Y Acad. Sci.* *1217*, 122–138.

STAR★METHODS

KEY RESOURCES TABLE

REAGENT or RESOURCE	SOURCE	IDENTIFIER
Antibodies		
Anti-human/mouse CD44 PE	eBioscience	Cat#12-0441-83; RRID: AB_465665
Anti-mouse CD117 (cKit) APC	eBioscience	Cat#17-1171-82; RRID: AB_469430
Anti-mouse CD117 (cKit) PE	eBioscience	Cat# 12-1171-82; RRID: AB_465813
Anti-mouse CD25 eFluor-450	eBioscience	Cat#48-0251-82; RRID: AB_10671550
Anti-mouse CD25 APCe780	eBioscience	Cat#47-0251-82; RRID: AB_1272179
Anti-mouse CD25-Alexa Fluor 647	Biolegend	Cat# 102019; RRID: AB_493459
Anti-mouse CD45 PECy7	eBioscience	Cat#25-0451-82; RRID: AB_2734986
Anti-mouse CD44 APC-e780	eBioscience	Cat# 47-0441-82; RRID: AB_1272244
Anti-mouse CD44 eFluor 450	eBioscience	Cat #48-0441-82; RRID: AB_1272246
Anti-mouse NK1.1 Biotin	eBioscience	Cat#13-5941-85; RRID: AB_466805
Anti-mouse CD19 Biotin	eBioscience	Cat#13-0193-85; RRID: AB_657658
Anti-mouse Ter119 Biotin	eBioscience	Cat#13-5921-85; RRID: AB_466798
Anti-mouse CD11b Biotin	eBioscience	Cat#13-0112-86; RRID: AB_466361
Anti-mouse CD11c Biotin	eBioscience	Cat#13-0114-85; RRID: AB_466364
Anti-mouse CD8 α Biotin	eBioscience	Cat#13-0081-86; RRID: AB_466348
Anti-mouse TCR $\gamma\delta$ Biotin	eBioscience	Cat#13-5711-85; RRID: AB_466669
Anti-mouse TCR β Biotin	eBioscience	Cat#13-5961-85; RRID: AB_466820
Anti-mouse Ly6G/C (Gr1) Biotin	Biolegend	Cat# 108404; RRID: AB_313369
Streptavidin PerCP-Cy5.5	eBioscience	Cat#45-4317-82; RRID: AB_10311495
Biological Samples		
Primary murine thymocytes	This work	N/A
Chemicals, Peptides, and Recombinant Proteins		
MEM Alpha	GIBCO	Cat#12561-056
Fetal Bovine Serum	SigmaAldrich	Cat#F7305
Human IL-7	PeptoTech Inc	Cat#200-07
Human FLT-3-Ligand	PeptoTech Inc	Cat#300-19
Stem Cell Factor	PeptoTech Inc	Cat#250-03
Murine M-CSF	PeptoTech Inc	Cat#315-02
Cell Trace Violet	Molecular Probes	Cat#34571
HBSS	GIBCO	Cat#14175-095
HEPES	GIBCO	Cat#15630-080
Pen Strep Glutamine	GIBCO	Cat#10378-016
Puromycin	SigmaAldrich	P8833-10MG
MACS Streptavidin Microbeads	Miltenyi Biotec	Cat#130-048-101
MACS LS columns	Miltenyi Biotec	Cat#130-042-401
37% formaldehyde	ThermoFisher Scientific	Cat#28908
7AAD	eBioscience	Cat#00-6993-50
β -mercaptoethanol	SigmaAldrich	Cat#M6250
NaBH ₄	SigmaAldrich	Cat#452882
DNaseI recombinant, RNase-free	Roche	Cat#4716728001
20 \times SSC	Invitrogen	Cat#15557-036
Formamide	Ambion	Cat#AM9344
HCR amplification hairpins	Molecular Instruments	Custom order

(Continued on next page)

Continued

REAGENT or RESOURCE	SOURCE	IDENTIFIER
Dextran Sulfate	SigmaAldrich	Cat#D8906
Trolox	Calbiochem	Cat#648471
Pyranose oxidase	SigmaAldrich	Cat#P4234
Catalase	SigmaAldrich	Cat#C3155
NHS ester labeled Alexa 488	Invitrogen	Cat#A20000
NHS ester labeled Alexa 594	Invitrogen	Cat#A20004
NHS ester labeled Alexa 647	Invitrogen	Cat#A20006
NHS ester labeled Cy3B	GE Healthcare	Cat#PA63101
Experimental Models: Cell Lines		
OP9-DL1	Schmitt and Zúñiga-Pflücker, 2002	N/A
OP9-DL1 dGFP	This work	N/A
OP9-control	Schmitt and Zúñiga-Pflücker, 2002	N/A
Experimental Models: Organisms/Strains		
Mouse: C57BL/6	Jackson laboratories	Stock NO: 664
Mouse: B6.Bcl11b ^{yfp/yfp} (Bcl11btm1.1Evr)	Kueh et al., 2016	N/A
Mouse: B6.ROSA26-mTom;Bcl11b-YFP	This work	N/A
Mouse: B6.129(Cg)-Gt(ROSA)26Sor ^{tm4(ACTB-tdTomato,-EGFP)}	Jackson laboratories	Stock NO: 007676
Oligonucleotides		
Listed in Table S6	This work	N/A
Recombinant DNA		
CRISPR puromycin plasmid (pSpCas9(BB)-2A-Puro PX459) V2.0)	Addgene	# 62988
Software and Algorithms		
FlowJo (v10.0.8)	N/A	https://www.flowjo.com/
Fiji (ImageJ) Imaging software	(Schindelin et al., 2012)	https://imagej.net/Fiji
Other		
BD FACS Aria II Cell Sorter	BD Bioscience	N/A
Lonza Nucleofector Kit	Lonza	N/A
iCyt Mission Technology Reflection Cell Sorter	Sony	N/A
BD FACSAria FUSION Cell Sorter	BD Bioscience	N/A
Miltenyi Biotec MACSQuant 10 Flow Cytometer	Miltenyi Biotec	N/A
hyb-cells	Grace Bio-Labs	RD478685-M
Microscope	Olympus	IX81
Confocal Scanner Unit	Yokogawa	CSU-W1
CCD camera	Andor	iKon-M 934
100x Oil Objective Lens NA1.40	Olympus	N/A
Motorized stage MS2000	ASI	N/A
Leica wide-field fluorescence inverted microscope w/ 40X objective; 504/542,560/607, 650-680 filters; Metamorph software	Leica	6000
24-well glass-bottom plate (P24G-1.0-13-F)	MatTek	P24G-1.0-13-F
Black PDMS micromesh inserts	Microsurfaces	MMA-0250-100-08-01
MATLAB version 9..3.0.713579 R (2017b)	N/A	https://www.mathworks.com
Python Language Reference, version 2.7	N/A	https://www.python.org/
SciPy: Open source scientific tools for Python	N/A	https://www.scipy.org
Scikit-learn: Machine Learning in Python	N/A	https://scikit-learn.org/stable/

RESOURCE AVAILABILITY

Lead Contact

Further information and requests for algorithms and code should be directed to and will be fulfilled by the Lead Contact Carsten Peterson (carsten@thep.lu.se).

Materials Availability

Requests for animals, experimental resources and reagents should be directed to and will be fulfilled by the Corresponding Author for the experimental work, Ellen V. Rothenberg (evroth@its.caltech.edu). All genotypes of mice used in this study were crossed from strains available from Jackson Laboratories, or from strains reported previously by the Rothenberg lab (Kueh et al., 2016), which are available upon reasonable request from Dr. Rothenberg with a minimal standard MTA. Cell lines generated in this study were derived from OP9-DL1 stock originally provided by J. C. Zúñiga-Pflücker, Sunnybrook Research Institute, University of Toronto. They are available on request with an MTA from Dr. Rothenberg and an MTA from Dr. Zúñiga-Pflücker.

Data and Code Availability

All mathematical models were coded in MATLAB version 9.3.0.713579 R (2017b), The Mathworks, Inc. Available at <https://www.mathworks.com> and in Python – Python Software Foundation. Python Language Reference, version 2.7. Available at <https://www.python.org/>. The ordinary differential equations were solved using Runge-Kutta methods encoded both in Python (SciPy: Open source scientific tools for Python. Available at <https://www.scipy.org>) and MATLAB (ode45 function). The stochastic simulations were done using the Gillespie algorithm implemented from scratch. All model implementations was done from scratch and the code is available upon request from the corresponding author.

The clustering algorithm we used is from Scikit-learn: Machine Learning in Python available at <https://scikit-learn.org/stable/>

The limited memory algorithm for bound constrained optimization L-BFGS-B is available at <https://docs.scipy.org/doc/scipy/reference/optimize.minimize-lbfgsb.html>

EXPERIMENTAL MODEL AND SUBJECT DETAILS

Animals

Mice of a variety of genotypes were used exclusively as sources of primary cells to be analyzed *ex vivo* in these studies. C57BL/6J, B6.*Bcl11b*^{YFP/YFP} (*Bcl11b*^{tm1.1Evr}) reporter mice (Kueh et al., 2016), and B6.129(Cg)-*Gt(ROSA)26Sor*^{tm4(ACTB-tdTomato,-EGFP)/J} mice expressing ubiquitous membrane Tomato (Jackson Laboratory) were bred, crossed, and maintained in the Caltech Lab Animal Resources Facility under specific pathogen-free conditions in accordance with protocols reviewed and approved by the Animal Care and Use Committee at California Institute of Technology.

C57BL/6J (B6) mice (bred from stock originally from Jackson Laboratories) were used for multiplex single-molecule FISH. B6.*Bcl11b*^{YFP/YFP} reporter (Kueh et al., 2016) mice were used to track the kinetics of *Bcl11b* upregulation *in vitro* in bulk studies shown in Figure 1. B6.*ROSA26-mTom;Bcl11b-YFP* mice for clonal imaging analysis were generated by crossing and backcrossing B6.129(Cg)-*Gt(ROSA)26Sor*^{tm4(ACTB-tdTomato,-EGFP)Luo/J} mice, which express ubiquitous membrane Tomato (Jackson Laboratories), with the B6.*Bcl11b*^{YFP/YFP} reporter mice until both loci were homozygous. These animals were then used for clonal imaging analyses. Note that the *Bcl11b*^{YFP} allele is a nondisruptive insertion of an internal ribosome entry site (IRES)-mCitrine (YFP) reporter into the 3' untranslated region of the last *Bcl11b* exon, which allows fully normal expression of the *Bcl11b* protein from the same allele.

All adult animals used were mice between 4 and 8 weeks of age, and all experiments were performed with samples consisting of cells pooled from multiple age- and sex-matched animals. Animals used for these experiments were bred and maintained at the Animal Facilities at California Institute of Technology under conventional Specific Pathogen-Free conditions, and animal protocols were reviewed and approved by the Institute Animal Care and Use Committee of the California Institute of Technology (Protocol #1445-18G). To maximize both thymus population sizes and fertility of the mice in the colony, care was taken to protect these animals from stress throughout their lifetimes to the greatest extent possible.

Cell lines

To provide a microenvironment that supports T-lineage differentiation *in vitro*, we co-cultivated primary cells with the OP9-DL1 stromal cell line (Schmitt and Zúñiga-Pflücker, 2002), which were obtained from Dr. Zúñiga-Pflücker (Sunnybrook Research Institute, University of Toronto) and maintained in our laboratory as described in the original reference. Details of the differentiation cultures vary with individual experimental designs, and are given below under Method Details.

METHOD DETAILS

DN thymocyte purification

Single cell suspensions were made from thymuses from 4-6-week-old mice. Thymuses were removed, passed through sterile metal meshes and collected in 1X Hanks Balanced Salt Solution supplemented with 0.5% fraction V bovine serum albumin,

10mM HEPES buffer (GIBCO), 5mM MgCl₂, and DNaseI. After pelleting, the cells were resuspended in a cocktail of biotinylated antibodies to deplete unwanted cells: CD8 α (53-6.7), TCR $\gamma\delta$ (GL3), TCR β (572597), Gr1 (R86.8C5), Ter119 (Ter119), NK1.1 (PK136), CD11b, and CD11c (N418), after which the cells were incubated with streptavidin-coated magnetic beads and then passed through a magnetic column (Miltenyi Biotec). The eluted DN cells were either used directly for RNA-FISH analysis or further purified by staining with fluorescent antibodies to CD44, Kit, CD25 and then sorting on a BD Biosciences FACSria or FACSriaFusion or a Sony SY3200 Cell Sorter in the Caltech Flow Cytometry Facility. For bulk cultures or clonal imaging, DN1 cells were sorted as CD44^{high} Kit^{high} CD25-negative Bcl11b-YFP-negative cells. Note that for simplicity, throughout this paper, we use “DN1” to refer to the Kit^{high} subset of CD44⁺ CD25⁻ cells that contains the T cell precursor function, also known as ETP.

Single Molecule Multiplex Fluorescent *in situ* Hybridization (sm-FISH)

Probe design and synthesis. The gene-specific primary probes for each gene tested were designed as previously described (Shah et al., 2016), with some modifications: each probe comprised a mRNA-complementary 35-mer sequence plus a shorter “gene handle” sequence that was shared by all probes against the same gene. All probes were blasted against the mouse transcriptome and expected copy numbers of off-target probe hits were calculated using predicted RNA counts in the ENCODE database for murine thymocytes. BLAST hits on any sequences other than the target gene with a 15-nt match were considered off-target hits. Any probe that hit an expected total off-target copy number exceeding 500 in count table was dropped. Probes were sequentially dropped from genes until any off-target gene was hit by no more than 6 probes from the entire pool. At this stage, all of the “viable” candidate probes for each gene had been identified. For the final probe set (Table S6), the best possible subset from the viable probes for each gene was selected such that none of the final probes used were within 2-nt bases of each other on the target mRNA sequence, with no-overlapping hybridization regions and GC-content close to 55%. Primary probes were synthesized and amplified from array-synthesized oligo-pool as previously described (Shah et al., 2016). The readout oligos against specific gene handles on primary probes were ordered from IDT (Integrated DNA Technologies, Coralville, Iowa) with 5'-amino modification and were coupled with NHS (N-hydroxysuccinimide)-ester labeled fluorescent dyes (Alexa 488, 594, 647 (Thermo Fisher Scientific) and Cy3B (GE Healthcare)) and purified through HPLC.

smFISH experiment and image acquisition. First, the isolated cells were spun onto an aminosilane modified coverslip, crosslinked with 4% Formaldehyde (ThermoScientific 28908) in 1X PBS for 10min, and permeabilized in 70% EtOH overnight at 4C. Samples were imaged first to record the surface antibody signals, followed by briefly bleaching away antibody signals through incubation in 0.1% NaBH₄ (Sigma 452882) in 1XPBS for 10 min. Then, the samples were **1)** hybridized overnight at 37°C with primary mRNA probes at 1 nM each oligo concentration in 50% Hybridization Buffer (50% HB: 2X SSC (saline sodium citrate, Invitrogen 15557-036), 50% (v/v) Formamide (Ambion AM9344), 10% Dextran Sulfate (Sigma D8906) in Ultrapure water (Invitrogen 10977-015)); then **2)** washed in 50% Wash Buffer [2X SSC, 50% (v/v) Formamide, 0.1% Triton X-100 (Sigma X-100)] for 20 minutes, followed by incubation in 2X SSC for 10 minutes. The samples were then **3)** incubated with fluorophore-coupled readout oligos, in 30% Hybridization buffer (30% HB: 2X SSC, 30% (v/v) Formaldehyde, 10% Dextran Sulfate in Ultrapure water) at concentrations of 10 nM each oligonucleotide for 30 minutes; this was followed by **4)** 5 minutes wash in 30% Wash Buffer (2X SSC, 30% Formamide (v/v), 0.1% Triton X-100 (Sigma X-100)), 3 minute wash in 2X SSC and DAPI staining. We then **5)** proceeded to imaging of this round of hybridization as described below. After image acquisition, **6)** the samples were incubated with 70% formamide with 1x PBS at room temperature for 30 minutes, followed by 3 rounds of washing in 1x PBS for 5 minutes each round. The procedures **3)-6)** were then repeated with a set of gene-specific readout oligos until the completion of measurements of all genes of interest, as illustrated in Figure 2A.

Samples were imaged in an anti-bleaching buffer (20 mM Tris-HCl, 50 mM NaCl, 0.8% glucose, saturated Trolox (Acros Organics 218940050), pyranose oxidase (OD405 = 0.05) (Sigma P4234), and catalase at a dilution of 1/1000 (Sigma C3155)) with the microscope (Olympus IX81) equipped with a confocal scanner unit (Yokogawa CSU-W1), a CCD camera (Andor iKon-M 934), 100x oil objective lens (Olympus NA 1.4), and a motorised stage (ASI MS2000). Lasers from CNI and filter sets from Semrock were used. Snapshots were acquired with 0.5 μ m z steps for more than 10 positions per sample.

The cells were segmented and categorised according to surface antibodies, and dots representing individual mRNA molecules were assigned to individual cells as shown in Figure 2B.

Bulk cell cultures

For experiments on developmental kinetics DN1 cells were cultured on OP9-DL1 stromal cells as previously described (Kueh et al., 2016; Yui et al., 2010) in α MEM medium supplemented with L-glutamine, penicillin, streptomycin (OP9 culture medium) and IL-7 (5 ng/ml) and Flt3L (10 ng/ml). For tracking proliferation in bulk cultures, FACS purified DN1 cells were stained in 5 μ M Cell Trace Violet (Molecular Probes) for 7 minutes at 37°C in HBSS and washed 2X with whole medium before being plated in 96-well OP9-DL1 co-cultures. Cells were incubated at 37°C 7% CO₂, harvested on days 2, 3, 4, and 5 by vigorous pipetting, and evaluated for developmental status by staining with antibodies, CD25 and Kit for development and CD45 to separate developing cells from stroma, plus a viability dye, 7AAD. The cells were then analyzed for expression of these markers plus Bcl11b-YFP using a Miltenyi MACSQUANT flow cytometer. Analysis was carried out using FlowJo software (Treestar). CTV measurements of cell division histories were calibrated using total DN (DN1–4) cultures on OP9-DL1 stroma (Figure S1).

Clonal Image Analysis

Specialty biological inputs. To follow the kinetics of development and proliferation in individual DN1-derived clones by microscopic imaging, several technical problems had to be solved. First, to allow for accurate identification and segmentation of developing cells of varying shapes and sizes on a background layer of OP9-DL1 stroma, we needed to use cells purified from B6.R26^{mTom/mTom} Bcl11b^{yfp/yfp} J mice expressing ubiquitous membrane Tomato (Jackson Laboratory) as well as an mCitrine yellow fluorescent reporter for *Bcl11b* expression, as described above. For imaging, we used animals in which both fluorescent reporter loci were homozygous (B6.ROSA26-mTom;Bcl11b-YFP mice).

To enable the use of the Bcl11b-YFP reporter for imaging the commitment status of individual T cell precursors developing on an OP9-DL1 stromal layer, GFP had to be removed from the original OP9-DL1 stromal cells to reduce spectral interference (Schmitt and Zúñiga-Pflücker, 2002). To do this, we designed guide RNAs targeting GFP and electroporated them along with a CRISPR puromycin plasmid v2.0 (Plasmid # 62988, Addgene) into OP9-DL1-GFP cells using the Lonza Nucleofector Kit. After puromycin selection, different clones were analyzed for GFP expression and one clone found to be completely negative for GFP (OP9-DL1-delGFP1) which was selected for use in imaging.

Imaging culture conditions. Developing T cells are extremely active in motility, so individual DN1 cells had to be confined in wells for imaging. We used a protocol similar to our previous experiments with DN2 cells (Kueh et al., 2016), but we found that the background of the clear wells was too high in the fluorescence channels used in these experiments, so black microwells were substituted. Black 8mm circular poly (dimethyl siloxane) PDMS micromeshes with over 150 microwells punched/micromesh, each approximately 250 μ M wide \times 100 μ M deep, were custom fabricated by Microsurfaces (Australia). One day before the start of co-culture and imaging these micro meshes were adhered to each of 2-4 (macro)wells of a 24-well glass-bottom plate (P24G-1.0-13-F) (MatTek, Ashland, CA), sterilized with ethanol, and washed in accordance with manufacturer's instructions, and then seeded with 5,000 OP9-DL1-delGFP stromal cells. Cultures were carried out in OP9 culture medium prepared as previously described except for the omission of pH indicator, phenol red, from the medium, which gives background fluorescence, and with the addition of 10mM HEPES buffer to stabilize the pH of the wells during imaging. On the first day of culture the following were added to the medium: 10 ng/ml Flt3L, 5 ng/ml IL-7, and 0.05 μ g/ml anti-CD25-AlexaFluor647 antibody (BioLegend) for detection of CD25 surface expression. DN cells were isolated from B6.ROSA26-mTom/Bcl11b-YFP mice and DN1 cells were FACS-purified as described above. 1,500 sorted DN1 cells were added to each well of the 24-well plate which had the microwells pre-seeded with OP9-DL1-delGFP. The cells were allowed to recover for at least one-hour at 37C, 7% CO₂, before transfer to the fluorescence microscope.

Imaging data collection. Imaging was carried out with a Leica 6000 wide-field fluorescence microscope with Metamorph software and an incubation chamber pre-set to 37C, 7% CO₂. Each microwell was briefly checked using the 40X objective, which allowed imaging of one entire microwell, for mTomato-positive cells. Microwells with 1 or 2 cells had their X-Y stage positions marked. All marked microwells were imaged daily using differential interference contrast (DIC) and three fluorescence channels: 504 excitation-542 filter for YFP, 560-607 for mTomato, and 650-684 for CD25-AlexaFluor647. Wells found to have one mTomato positive cell on either day 1 or 2 and cells present throughout the culture period were selected to further analysis. mTomato-positive cells were segmented and analyzed using Fiji (ImageJ) and by hand to obtain cell area and CD25 and Bcl11b-YFP fluorescence data. For each cell, total fluorescence was calculated by multiplying mean fluorescence by area. Because microwells became quite crowded by late time points (days 6-7), and cell clumps were very difficult to segment, all cells were counted but only a subset of the distinguishable cells were segmented and sampled to obtain fluorescence data for the well. Fifty background fluorescence sample cells were taken and the mean total fluorescence for CD25 and Bcl11b-YFP plus 3 standard deviations was used to set the thresholds for determination of positive versus negative cells (Figure S6).

Modeling: Pseudo-time series

From the single cell FISH data, pseudo-time-series were created by clustering the data and ordering the clusters, assuming that the cells measured are in states distributed over the whole-time intervals of both stages DN1 and DN2. The data points were first classified into clusters with a Gaussian-Mixture algorithm (Pedregosa et al., 2011), assuming that they follow several Gaussian distributions with individual parameters. Moreover, the number of data points in each cluster should be similar in order to regard all data with the same significance. The mean expression and standard deviation of all data points in a cluster were then computed for each gene and used to arrange the clusters in time. The cluster order was set by the fact that both *Tcf7* and *Bcl11b* expression levels are known to increase during cell commitment (Tydell et al., 2007; Zhou et al., 2010). We considered the relative values of gene expression levels with respect to the maximum level at each stage to prohibit *Tcf7* from dictating the order of the clusters alone, since its absolute expression is higher than *Bcl11b* expression. For each cluster, expression values of *Tcf7* and *Bcl11b* were added. Finally, the clusters were ordered such that this sum is increasing and placed equidistantly on an arbitrary time axis.

Single Cell Model: Transcription factor Level

For the circuit in Figure 4A we obtained the following set of rate equations from a thermodynamic approach (Ackers et al., 1982). The equations describe the dynamics of T cell specific genes *Tcf7*, *Gata3* and *Runx1* along with gene opposing the T cell fate *PU.1* and X function, with concentration levels denoted as: [T], [G], [R], [P] and [X]. The Notch signaling activity is denoted as N.

$$\frac{\partial[R]}{\partial t} = \frac{p_1[R] + p_2N}{1 + p_1[R] + p_2N} - \gamma_R[R]$$

$$\frac{\partial[T]}{\partial t} = \frac{p_3[T] + p_4[G] + \frac{p_5N}{p_6 + [P]}}{1 + p_3[T] + p_4[G] + p_5N + p_7[P]} - \gamma_T[T]$$

$$\frac{\partial[G]}{\partial t} = \frac{p_8[T] + \frac{p_9N}{p_{10} + [P]}}{1 + p_8[T] + p_9N + p_{11}[P]} - \gamma_G[G]$$

$$\frac{\partial[P]}{\partial t} = \frac{p_{12}[P]}{1 + p_{12}[P] + p_{13}[R][G][T]} - \gamma_P[P]$$

$$\frac{\partial[X]}{\partial t} = \frac{1}{1 + p_{14}[T] + p_{15}[G]} - \gamma_X[X]$$

$$\frac{\partial N}{\partial t} = \frac{p_{16}}{1 + N}$$

We considered that repression of *PU.1* by the T cell specific factors follows AND logic i.e., Runx1, Tcf7 and Gata3 all need to be present for active repression of *PU.1*. This accounts for the experimentally observed increases in *PU.1* expression upon deletion or downregulation of any one of these regulators (see Table 1 for references). We modeled the Notch activity such that the N levels increase throughout the simulations, however, parameter p_{16} values were chosen to be very low to assure only a slight increase of the Notch activity, avoiding a jump to infinity. *PU.1* limits the upregulation of Tcf7 and Gata3 by Notch signaling. This was modeled by dividing the Notch activation term, in the nominator of the rate equations of Tcf7 and Gata3, by a linear term including *PU.1* expression level. The rest of interactions in the model follow a standard Shea-Ackers formalism implementation. The parameters p_i with $i = 1:15$ except p_6, p_{10} , which are part of implementation of *PU.1* impeding the Notch positive regulation of Tcf7 and Gata3, correspond to binding affinities, while γ parameters model the decay rates linked to the half-lives of the respective molecules. The parameter values (Table S1) were chosen such as the resulting expression levels dynamics for all the genes in the network are within the range levels of genes expression observed in the single cell FISH experiments (Figure S2). It should be noted that parameters p_{14} and p_{15} were estimated from FISH expression data of candidate genes for X function. The model parameters were optimized (see details below) from the pseudo-time-series data using simulated annealing, genetic algorithms and a bound constrained optimization algorithm (L-BFGS-B) (Byrd et al., 1995). From optimization solutions (Raue et al., 2013) we picked one that with minor adjustment including rounding accommodated the DN1 to DN2 transition shown by RNA-FISH measurements (Figure 2C). The determined parameters exhibit robustness, see below.

Single Cell Model: Parameter Optimization Details. The cost function of the model parameter set \mathbf{p} measuring the squared difference between the pseudo-time-series data D and model simulated concentration values of the genes in the network, all grouped under variable M has the following expression:

$$S(\mathbf{p}) = \sum_k \sum_{i=1}^n (D_{kj} - M_{kj}(\mathbf{p}))^2 \text{ where } k = [R], [T], [G], [P]$$

n is the number of data points for each gene in the pseudo-time-series data, genes *Tcf7*, *Gata3* and *Runx1* along with gene opposing the T cell fate *PU.1*, concentration levels were denoted as: $[T]$, $[G]$, $[R]$, $[P]$.

$S(\mathbf{p})$, or rather its -log likelihood counterpart, was minimized with respect to the model parameters when we used the bound constrained optimization algorithm (L-BFGS-B) (Byrd et al., 1995). This was done 100 times with different initializations from the parameter space by Latin Hypercube sampling (Raue et al., 2013) with most solutions having similar levels for the cost function value. From the optimization outcomes we picked a solution, denoted preferred, that respected the experimentally observed transition from DN1 to DN2 (Figure 2C). The picked solution was used as input further in the work, even though it had somewhat higher cost function value compared to other solutions (Figure S7C). The rounded values of optimized parameters are shown in Table S1. The set of parameters that we picked (agreeing both with pseudo-time-series data and experimentally observed DN1 to DN2 transition in Figure 2C) has most of the values within the distribution of the values of parameters obtained from 100 optimization runs (Figure S7D). We also estimated the model prediction bounds using methods inspired from the MATLAB Curve Fitting Toolbox functions. We found that, for the

set of parameters that we used throughout this work, the model prediction bounds are within the experimental data levels, see [Figure S2](#). The prediction bounds for many of the other parameter sets either diverged or vanished. In order to compare properties among parameter sets that yielded different types of prediction bounds, we manually categorized the prediction bounds depending on their behavior. [Figure S7E](#) shows the distribution of the sums of the absolute values of the components of the Jacobian, from which the prediction bounds are calculated, for each category. The preferred parameter set ends up in the ‘OK’ prediction bounds cluster. There is also a clear separation between the three main cases. [Figure S7F](#) shows the component of the Jacobian that corresponds to each parameter, plotted against the parameter value. The preferred parameter set clearly lies close to zero in most cases. One should keep in mind that the fitting is not performed on original data but rather on pseudo time series. These originate from clustering procedures and do not yield unique results—“horizontal” errors from the time ordering are hard to estimate.

Robustness Analysis. We also performed robustness analysis by calculating the robustness indexes given by

$$R_p^M = \frac{\delta M(x)/M(x)}{\delta p/p}$$

where p are the varied parameters, δp is the change in p , i.e., increased/decreased by 1%, 2%, 5% and 10% of its value, $M(x)$ is the model output with unperturbed parameters with x representing the gene expression levels and $\delta M(x)$ is the change of $M(x)$ due to change δp in p . In [Figure S7B](#) robustness coefficients are shown when all model parameters simultaneously are varied with 1%, 2%, 5% and 10% respectively. The robustness coefficients are lower than the perturbation level, increasing with the level of perturbation, leading to the conclusion that the model is robust.

The deterministic and stochastic simulation results presented in [Figures 4, 5, S2, and S4](#) were obtained using the parameter values shown in [Tables S1, S2, S4, and S5](#). Among the parameters, the decay rates are given in real units, whereas the other parameters cannot be tied with experimental numbers as there are overall implicit constants multiplying the ratios appearing on the right-hand side of the rate equations.

It should be noted that we consider the starting state of the system to correspond to an uncommitted early progenitor cell with high expression levels of *PU.1* and *X* activity along with *Runx1* being expressed and low expression levels of *Tcf7*, *Gata3*. In actual DN1-DN2a stage cells, note that several other progenitor-specific transcription factors in addition to *PU.1* are still expressed until commitment, but the network connections of these factors to *Tcf7*, *Gata3*, *Runx1* and Notch signaling have not yet been studied in depth ([Yui and Rothenberg, 2014](#)). Thus, it is possible that *PU.1* is not the only factor involved in antagonizing *GATA3*, *Tcf7* and Notch, and further work should define better the roles of other genes in this initial period.

Single Cell Model: Epigenetic Level

We implemented a simplified computational model for *Bcl11b* regulatory system transition from chromatin closed to open state. The epigenetic level model is linked to the transcription level model through the fact that the expression of *Runx1* and *X* and level of activity of Notch signaling (outputs of transcription level model) serve as inputs to the *Bcl11b* regulatory system model. We consider that *Runx1* along with Notch signaling activity push toward an open *Bcl11b* regulatory system state thus the probability of achieving this state is direct proportional to *Runx1* and Notch levels. *X* activity is considered to help maintain a closed state of the *Bcl11b* regulatory system, dictating the probability of the regulatory region to be closed. We initially put forward a very simplified model in which we considered the size of the *Bcl11b* regulatory system to have a finite low number of CpG sites i.e., 20. This was chosen to make sure that the simulations were very fast, but that we considered enough CpG sites so that the transitions between the opened and closed states are not instant and not completely controlled by noise. The amount of existing CpG sites in one of the states influences negatively the amount of CpG sites in the other state, because of the finite total number of CpG sites. We conducted a stochastic implementation of the simplified model for epigenetic state evolution of *Bcl11b* regulatory region governed by the following master equation:

$$\frac{\partial B_{open}}{\partial t} = p_{17}R + p_{18}N - p_{19}B_{closed}$$

$$\frac{\partial B_{closed}}{\partial t} = p_{20}X - p_{21}B_{open}$$

where N denotes Notch signaling activity, R represents the expression level of *Runx1* and X is the *X* activity level. The model parameter values are shown in [Table S2](#).

The initial conditions for stochastic simulations of the epigenetic level model consider the *Bcl11b* regulatory system to be in closed state i.e., the number of closed CpG sites is greater than the number of sites considered to be open. In some simulations the *Bcl11b* regulatory system becomes open i.e., the number of opened CpG sites becomes and remains greater than the number of closed CpG sites during simulation time.

Multi-scale model of proliferation and gene expression

Multi-scale Model: Population Level. To analyze the effects of cell proliferation on the population scale gene expression we devised a model in which we can track the gene expression and division history of each individual cell. The model is an extension of in house

developed framework, written in C++ and used previously in the context of cell-based simulations (Krupinski et al., 2012). We have implemented mechanical interactions between cells for the sole purpose of displaying the evolution of cells in a simple way. Each cell in the model contains a copy of a gene network presented in Single Cell Model (both transcription and epigenetic levels) and evolves its gene expression levels independently from other cells by a stochastic Gillespie simulation (Gillespie, 1977). Cell cycle length for each cell is also a stochastic variable and chosen from a normal distribution. The cell divisions are assumed to be symmetric such that the daughter cells inherit the mother cell content, but then evolve independently. Since we assume that the presented network describes transition from DN1 to DN2a cell state in the multi-cell simulations we turn on the network in its initial state between day 1 and day 2 while the start of the simulation at day 0 corresponds to introduction of a single DN1 clone to the well.

We constructed a unified model of cell proliferation consistent with the data by assuming that the cell cycle length is a function of the cell generation. The parameters of the cell cycle length normal distributions were chosen independently for each cell based on its generation and fitted globally to the CTV data in Table S4.

Population Model – Details. In order to take population dynamics into account in a multi-level model, we developed population models with parameters extracted from the CTV data. As CTV staining intensity provides information about the number of divisions that an individual cell has gone through, it allows us to build generation profiles for each of the assessed cell type groups (DN1 and DN2a) and at different time points of measurements. These profiles exhibit dispersion of the generation distributions in time.

A simplified model

To assess if this dispersion can be explained by proliferation, which is uniform for all generations, we first developed a population proliferation model in which the division rate of a cell does not depend on its generation.

A three-parameter model was employed that gave the best fit of the predicted to measured cell numbers with the substantial different average relative error at the points of measurement. The prediction of DN1 cell numbers differed on average from experimental values by 16.9% at day 2 and by 5.3% at day 3 (Table S3). At day 2 in both groups of cells we observed a large proportion (above 50%) of the cell population not dividing between measurements. At day 3 this proportion was lowered to about 20%, with most of the cells (around 50%) having divided twice. This demonstrates that the cells after sorting and seeding into the cell plate culture experienced an initial slowdown in proliferation rate to a value smaller than 1/24h but recovering to the rate of about 1/12h at later times.

The results of this simple population model of cell proliferation suggest that the dispersion of the cell proportions among different generations is not uniform either between different cell types (DN1 or DN2a) or for different measurement time points (day 2, day 3) within the same cell group.

In order to estimate the cell cycle lengths in a population of immature thymocytes as they progress through their development, we first devised a simple population model of cell division fitted to experimental CTV data. This “null model” assumes that a cell can divide between measurement time points from 0 to 3 times. This assumption matches our observations from confocal imaging of the cell cultures. The proportions of the cells in the population undergoing 1 to 3 divisions are denoted a , b and c respectively. These parameters have to satisfy the relation $a+b+c < 1$. The proportion of cells not dividing between measurements points is given by $1-a-b-c$, which exhausts all the cases considered in the model. We also hypothesize in this “null model” that these proportions are uniform through all the cell generations in the population. Correctness of this working hypothesis will be assessed from the results of the model. This means that number of cells in generation G_i^{n+1} at time t_{n+1} is given in terms of generations at previous time t_n by

$$G_i^{n+1} = (1 - a - b - c)G_i^n + 2aG_{i-1}^n + 4bG_{i-2}^n + 8cG_{i-3}^n$$

Measuring the goodness of the fit to CTV data is given by the relative error (Table S3).

This demonstrates that the cells after seeding into the cell plate culture experience initial slowdown in proliferation rate to a value smaller than 1/24h recovering the rate of about 1/12h at later times.

The results and the outputs of this population “null model” of cell proliferation suggest that the dispersion of the cell proportions among different generations is not uniform either between different cell types (DN1 or DN2a) or for different measurement time points (day 2, day 3) within the same cell group.

The CTV data provides complete untruncated distributions of generations 0 to 6 for two groups of cells (DN1 and DN2a cells) for days 2, 3 and 4. This allows us to predict with the model distributions at day 3 given data at day 2 and to predict distributions at day 4 given data at day 3. Comparison of these predictions to the actual data gives us best fit parameters of population level cell proliferation at given time point (Table S4). In this way the dispersion of cell proportions among different generations can be included. These parameters were found by global minimization of the normalized mean error measure between predicted and measured cell numbers in generations.

Multiscale Model: Epigenetic Level. In order to take into account the effect of cell division on the amounts of epigenetic factors inside a cell, we implemented a Bcl11b regulation region epigenetic model where the regulatory system can be open, close or at an intermediate state corresponding, in a CpG site methylation model, to the methylated, unmethylated and hemi-methylated states (Figure S7G). This type of collaborative model was proposed in Haerter et al. (2014) and Olariu et al. (2016) and used here to simulate the opening and closing of the *Bcl11b* regulation region. In the model the CpG sites can be methylated (M)–closed state, hemi-methylated (H)–intermediate state or unmethylated (U)–open state. Transition rate constants depend on the level of X activity, expression

level of Runx1 as well as Notch signaling activity. The arrows labeled with α to ε indicate reactions that require a mediator nearby (e.g., the α arrow defines a transition from U to H in the presence of a mediator in state M).

The collaborative model is described by the following set of equations:

$$\frac{\partial U}{\partial t} = (k_2 N + k_3 [R])H - k_1 [X]U + HU - UM$$

$$\frac{\partial H}{\partial t} = k_1 [X]U + (k_2 N + k_3 [R])M - (k_1 [X] + k_2 N + k_3 [R])H + UM + MU - HU - HM - HH$$

$$\frac{\partial M}{\partial t} = k_1 [X]H - (k_2 N + k_3 [R])M + HM + HH - MU$$

The parameter values are shown in [Table S5](#).

When we simulate division in the multi-scale model, two new Gillespie simulations are started. In the absence of X, and in response to Runx1 (R) and Notch signals (N), the states of the mother cell *Bcl11b* regulatory region CpG sites are transferred to the daughters following the rule: M -> H, H -> 50% H + 50% U, U -> U ([Haerter et al., 2014](#)). It should be noted that when the collaborative model was used, the *Bcl11b* regulatory system had to be of a size of minimum 500 CpG sites, [Figures S7H](#) and [S7I](#). The results shown in [Figure 5D](#) were obtained only if the regulatory region of *Bcl11b* went through epigenetic events corresponding to demethylation of at least 500 CpG sites. Note that the real *Bcl11b* gene includes multiple methylated CpG islands in the gene body and also possesses extended enhancers spread over more than 850 million bp of DNA ([Figure S3](#)) ([Hu et al., 2018](#); [Li et al., 2013](#)), so this constraint is quite likely to be biologically meaningful.

Multi-scale Model: Parameters Collaborative Epigenetic Model

Since the previously presented “simple” population based model assumed cell cycle lengths independent from cell generations and required different parameter sets for each data time point, we wanted to see if now we can construct unified model of cell proliferation consistent with the data by assuming that the cell cycle length is a function of the cell generation. As such, the parameters of the cell cycle length normal distributions were chosen independently for each cell based on its generation and fitted globally to the CTV data in [Table S4](#). In order to take into account, the effect of cell division on the amounts of epigenetic factors inside a cell, we implemented a *Bcl11b* regulation region collaborative epigenetic model [Figure S7G](#) with parameters shown in [Table S5](#).

We considered the number of CpG sites of *Bcl11b* regulatory region to be also a model parameter and identified that at least 500 CpG sites have to be considered for the collaborative model. This result originates from analyzing the impact of noise when the number of CpG sites were varied ([Figure S7H](#)). For low number of CpG sites the noise has a larger impact. Therefore, we conclude that a minimal number of 500 CpG sites is required for observing a stable switch. We observed larger fluctuations at the steady states for a low number of CpG sites (see error bars [Figure S7I](#)) and also that the fraction of unmethylated sites in steady state decreases with higher CpG number. These results are not strongly affected by our choice of parameter values for the DNA methylation collaborative model (see [Figure S7G](#)).

This result is in good agreement with experimental findings in [Figure S3](#) showing that the *Bcl11b* regulatory region includes > 30 stage-specifically opening elements spread over > 1 Mb of gene desert, many of which loop to the promoter once *Bcl11b* is activated. Not shown, over 30 of them bind Runx1 and multiple elements bind GATA3, Notch/RBPJ, and/or TCF1, although this can be seen from genome browser analysis of published datasets. Only the elements in the vicinity of the “Major Peak” enhancer (magenta highlight, [Figure S3](#)) have been functionally analyzed so far ([Ng et al., 2018](#); [Isoda et al., 2017](#); [Li et al., 2013](#)), and those results plus newer work by [Hosokawa et al. \(2020\)](#) indicate that more than this one regulatory element is involved in the full extent of *Bcl11b* regulation. It would be realistic to say that this large genomic region contains at least 500 functionally relevant CpG dinucleotides.

QUANTIFICATION AND STATISTICAL ANALYSIS

Experimental data: Cell trace violet experimental data shown in [Figure 1](#) are from two independent experiments as described in the text and figure legend. FISH transcript count distributions in [Figure 2](#) show median and quartiles from 169 cells. Clonal imaging data shown in [Figure 6](#) are from images of 62 clonal T-lineage wells, which were individually tracked as shown in [Figure S5](#). Means and standard deviations are shown, as described in the text and figure legend. Fluorescence thresholds for the imaging data were determined by using the average + 3 standard deviations of background values taken from 50 samples, as described in the legend for [Figure S6](#).

The statistical tests used are indicated in the individual figure legends.

The modeling results shown in [Figures 1, 4, and 5](#) along with the ones in the [Figures S2, S4, and S7](#) were obtained using MATLAB version 9.3.0.713579 R (2017b), The Mathworks, Inc. Available at <https://www.mathworks.com>.