Cell Reports, Volume *34*

Supplemental Information

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

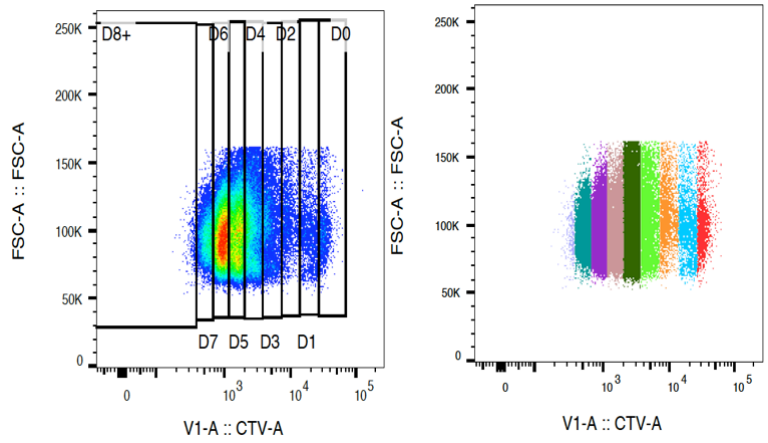
Victor Olariu, Mary A. Yui, Pawel Krupinski, Wen Zhou, Julia Deichmann, Emil Andersson, Ellen V. Rothenberg, and Carsten Peterson

SUPPLEMENTARY FIGURES AND FIGURE LEGENDS

Fig. S1-Fig. S7 and legends
Tables S1-S5
Legends to Tables S1-S5

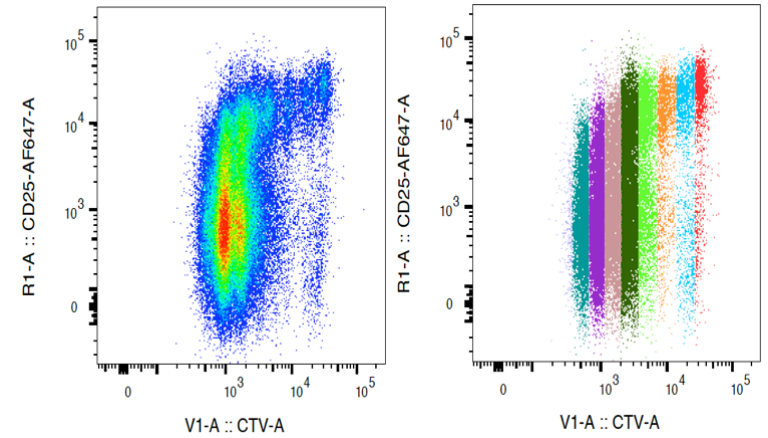
Supplementary Figure S1

A



	Sample Name	Subset Name	Count
	d3a Thy1 CTV	D8+	265
	d3a Thy1 CTV	D7	7141
	d3a Thy1 CTV	D6	37495
	d3a Thy1 CTV	D5	34097
	d3a Thy1 CTV	D4	15628
	d3a Thy1 CTV	D3	5227
	d3a Thy1 CTV	D2	2140
	d3a Thy1 CTV	D1	2170
	d3a Thy1 CTV	D0	1658

B

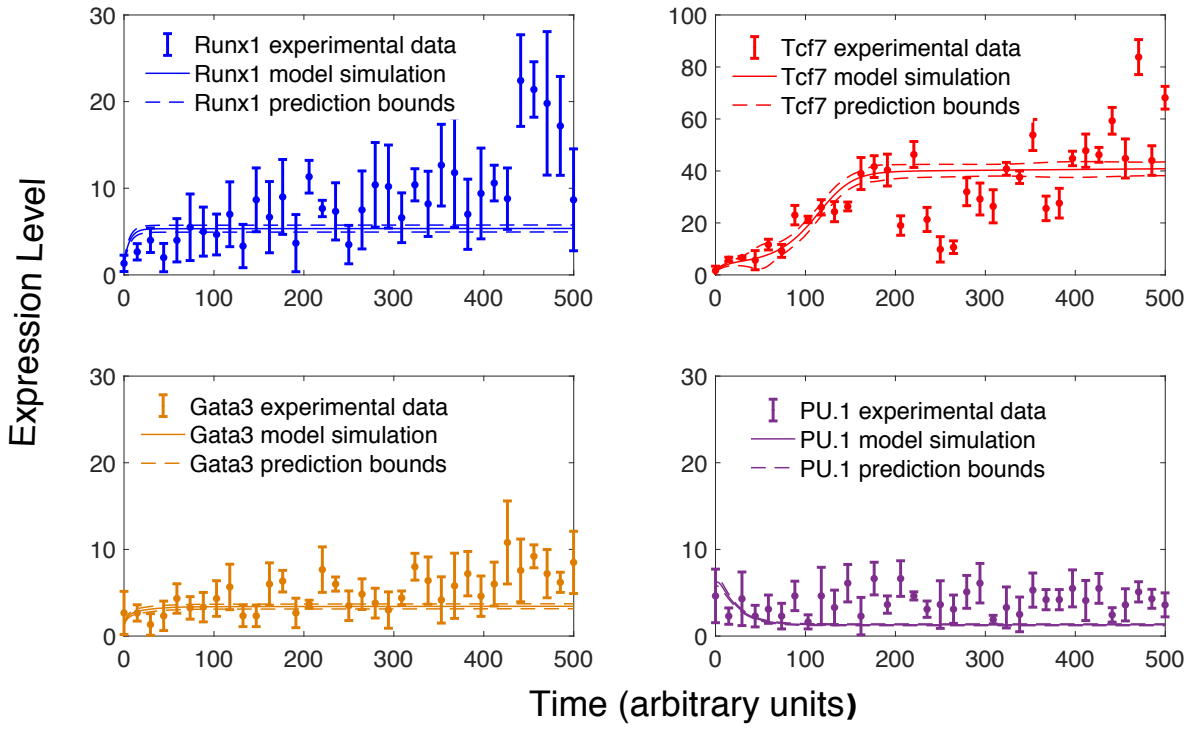


	Sample Name	Subset Name	Count
	d3a Thy1 CTV	D8+	265
	d3a Thy1 CTV	D7	7141
	d3a Thy1 CTV	D6	37495
	d3a Thy1 CTV	D5	34097
	d3a Thy1 CTV	D4	15628
	d3a Thy1 CTV	D3	5227
	d3a Thy1 CTV	D2	2140
	d3a Thy1 CTV	D1	2170
	d3a Thy1 CTV	D0	1658

Supplementary Figure S1, related to Fig. 1. Cell division determinations using Cell Trace Violet (CTV) levels in flow cytometric plots of control cells.

Unfractionated DN thymocytes were CTV stained and placed in OP9-DL1 culture for 4 days before harvesting, staining with surface differentiation marker antibodies, and FACs analysis (see Methods). CTV levels are diluted by dividing between daughter cells with every cell division. (A) Using Forward Scatter (FSC) vs. CTV plots, gates were drawn as shown, defining how many cell divisions each cell had experienced based on CTV ranges. Cells within each gate were assigned cell division numbers and color coded (right plot and legend): cells that had divided 0 times (D0, red) through 7 times (D7, teal) were distinguished and cells with CTV values below D7 were counted as at least 8 divisions (D8+, lilac). (B) FACs plots of the same cells but showing the relationship between CD25 and CTV in the developing DN cells. For total DN cells (mostly DN3 and DN4), CD25 generally declines/stays off with cell division.

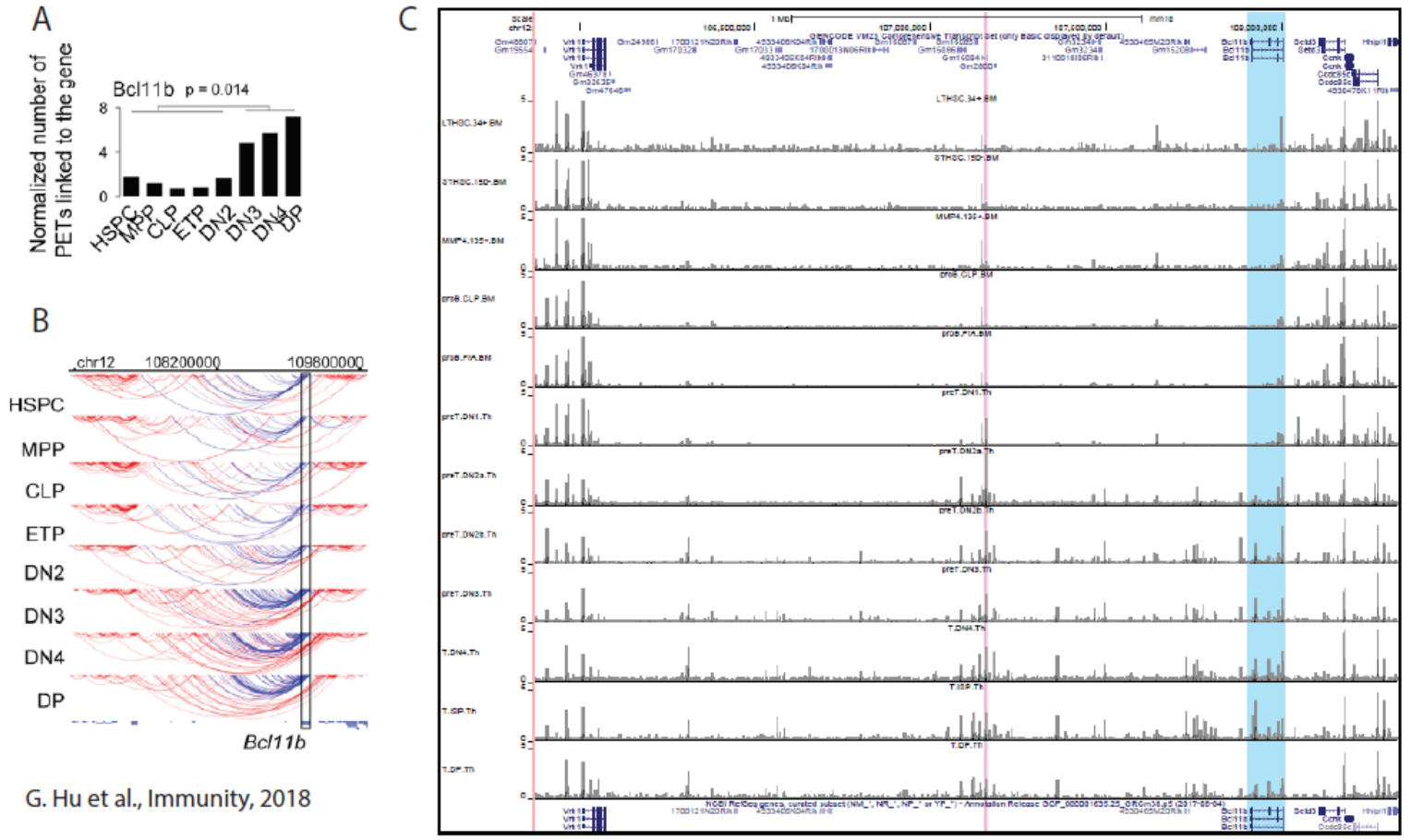
Supplementary Figure S2



Supplementary Figure S2, related to Fig. 4B. Preferred model fits to pseudo-time-series data.

Simulation dynamics from the gene regulatory network model with preferred parameters versus the pseudo-time-series data. The continuous lines depict the model deterministic simulation results while the dots show the mean mRNA count in each cluster obtained from FISH data and the bars show standard deviations. The intermittent lines depict the calculated prediction bounds. To avoid negative values on the y axis due to standard deviations, both model and data were shifted up by 1.3 in the fourth subpanel.

Supplementary Figure S3



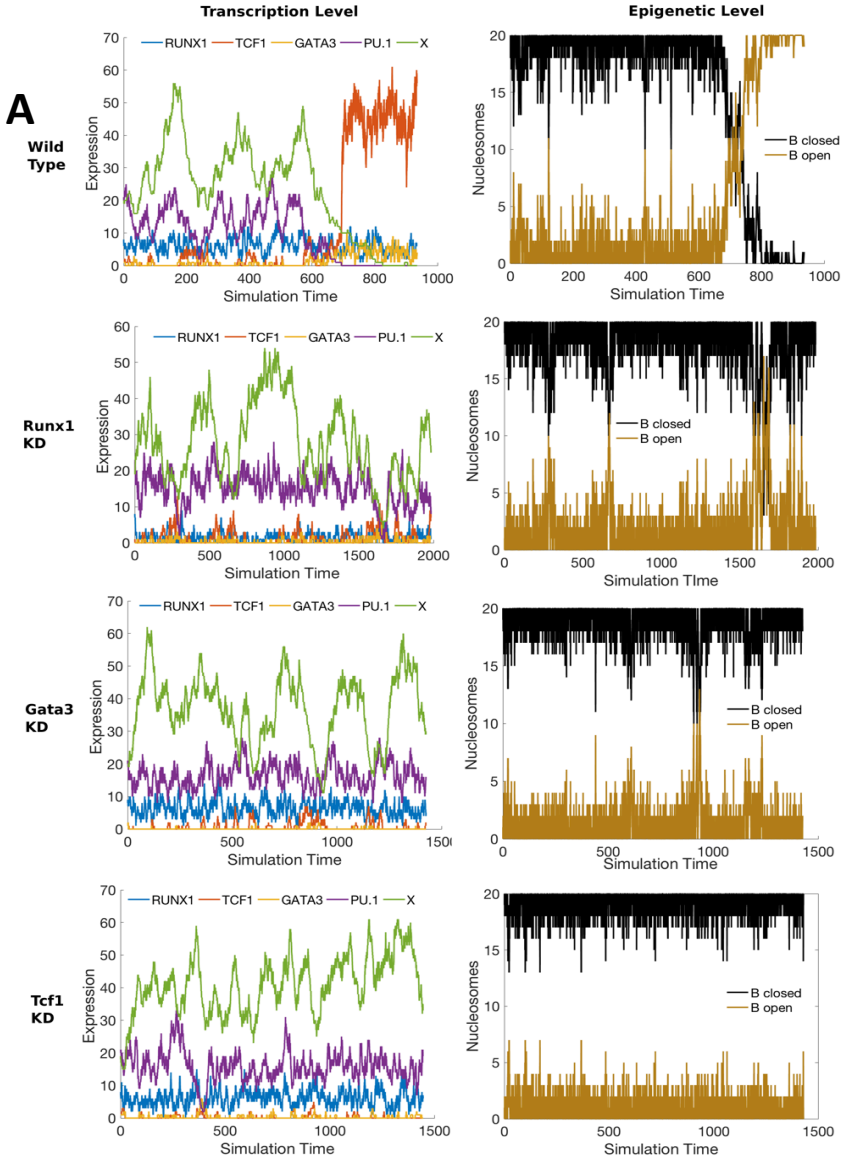
Supplementary Figure S3, related to Fig. 4C. Published evidence for role of epigenetic change in the mechanism of *Bcl11b* gene activation between DN2a and DN2b stages of T-cell development.

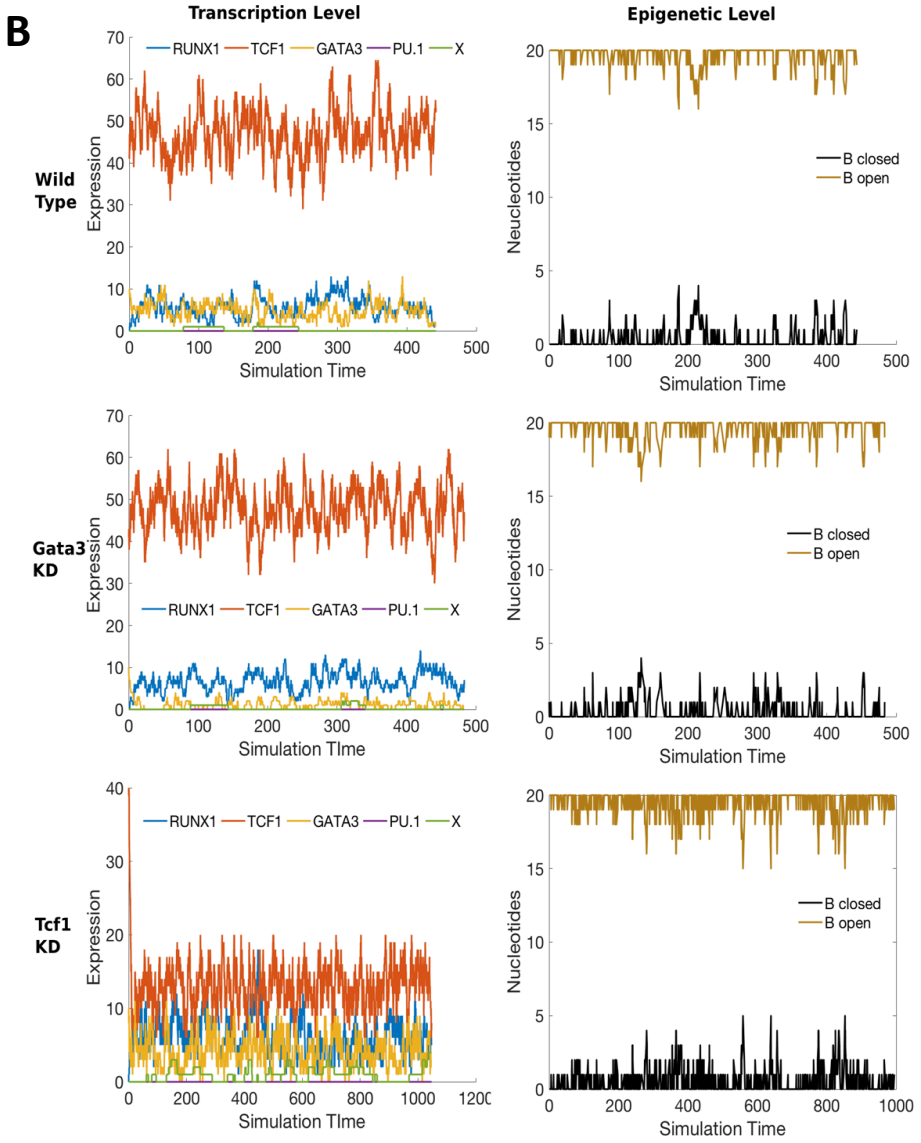
(A, B) Stage-specific increases in chromatin looping between *Bcl11b* gene and distal superenhancer complex, during differentiation from DN1 (ETP) to DN3 stage, as measured by Hi-C (panels reproduced from Hu et al., 2018, Fig. 3). (A) quantitation of loops to the *Bcl11b* gene at indicated stages: HSPC= hematopoietic stem and progenitor cells (prethymic); MPP=multipotent progenitor (prethymic); CLP=Common Lymphoid Progenitor (prethymic); ETP=Early T-cell precursor/ DN1 (intrathymic, T lineage); DN2, DN3, DN4=pre- and post-commitment early T cell developmental stages (intrathymic); DP=$CD4^+$ $CD8^+$ "double positive" T cell precursors (later intrathymic). (B) Map positions of DNA involved in significant looping interactions, as measured in indicated populations. Blue curves in B show loops that involve the *Bcl11b* gene itself.

(C) Patterns of chromatin accessibility from *Bcl11b* through the gene desert containing known regulatory elements, in stages from LongTerm Hematopoietic Stem Cell through contrasting early B cell stages, and in T cell developmental stages from DN1 through DP. ATAC-seq data are from the Immunological Genome Project, www.immgen.org (data published in Yoshida et al., 2019). Accessible sites are mapped to mm10 coordinates in browser provided at http://rstats.immgen.org/Chromatin/chromatin.html. A region of approximately 2.5 Mb is shown.

(D, E) Evidence for asynchronous activation of two alleles of *Bcl11b* within the same cell during DN2a to DN2b transition. (Panels reproduced from Ng et al., 2018, Fig. 2.) Mice were generated with a yellow fluorescent protein (YFP) reporter non-disruptively knocked into one allele of *Bcl11b* and a red fluorescent protein (mCherry) reporter non-disruptively knocked into the other allele of *Bcl11b*. Panels depict the differentiation of populations of DN2a cells from these animals, initially purified before the onset of any *Bcl11b* expression, as they differentiated in real time through the activation of one or both copies of the *Bcl11b* gene. (D) Time course of activation of YFP and mCherry-tagged *Bcl11b* alleles in populations of DN2 cells from the two-color mice. Note the large fractions of monoallelically expressing cells detected at 35-65 hr, before the majority turn on both alleles. (E) Panels depict patterns of activation of the indicated *Bcl11b* alleles in representative clones differentiating in vitro in real time from these animals. Note that cells produced in the first clone turn on only the red allele by the time of harvest at 110 hr, while cells in the second clone turn on the yellow allele in one clonal daughter by 42 hr but only activate the red allele between 73 and 96 hr. Monoallelic expression is indicative of epigenetic, cis-acting mechanisms, since both copies of the gene are exposed to the same TF environment in the same cell nucleus.

Supplementary Figure S4



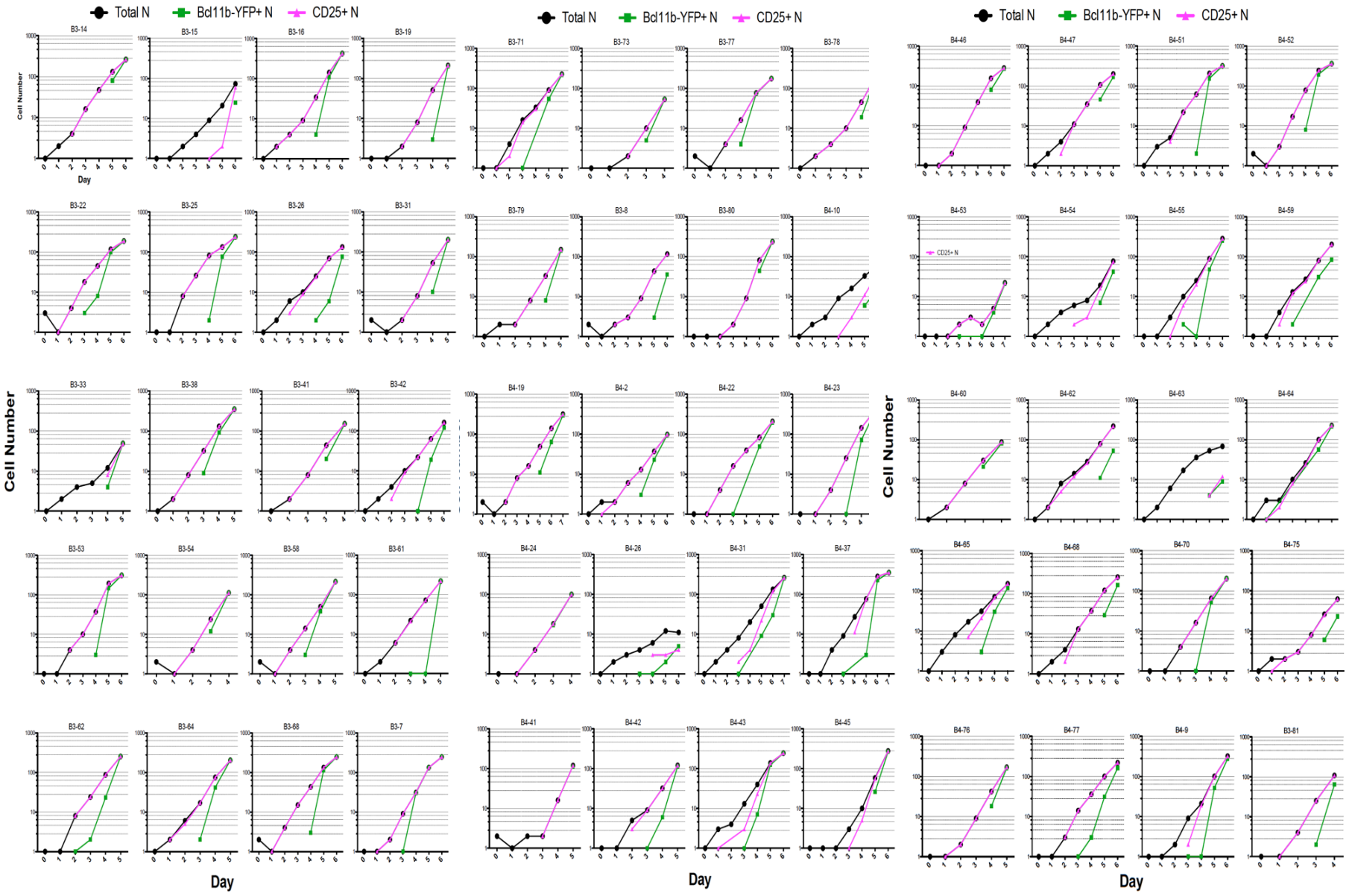


Supplementary Figure S4, related to Fig. 4D. Model validation and predictive tests

A Knock-down simulations for DN1 cells. First row shows the wild-type where no gene was knocked-down (K-D). The switch is thrown at the transcriptional level and *Bcl11b* becomes open. Second, third and fourth rows show Runx1, Gata3 and TCF1 K-D simulations. The T-cell factors do not get expressed and *Bcl11b* remains closed.

B Knock-down simulations for DN2 cells. First row shows the wild-type, where no gene was knocked-down. The T-cell factors are high and X is low, thus *Bcl11b* is open. Second and third rows show Gata3 and TCF1 K-D simulations. *Bcl11b* remains open, even though Gata3 and TCF1 were K-D.

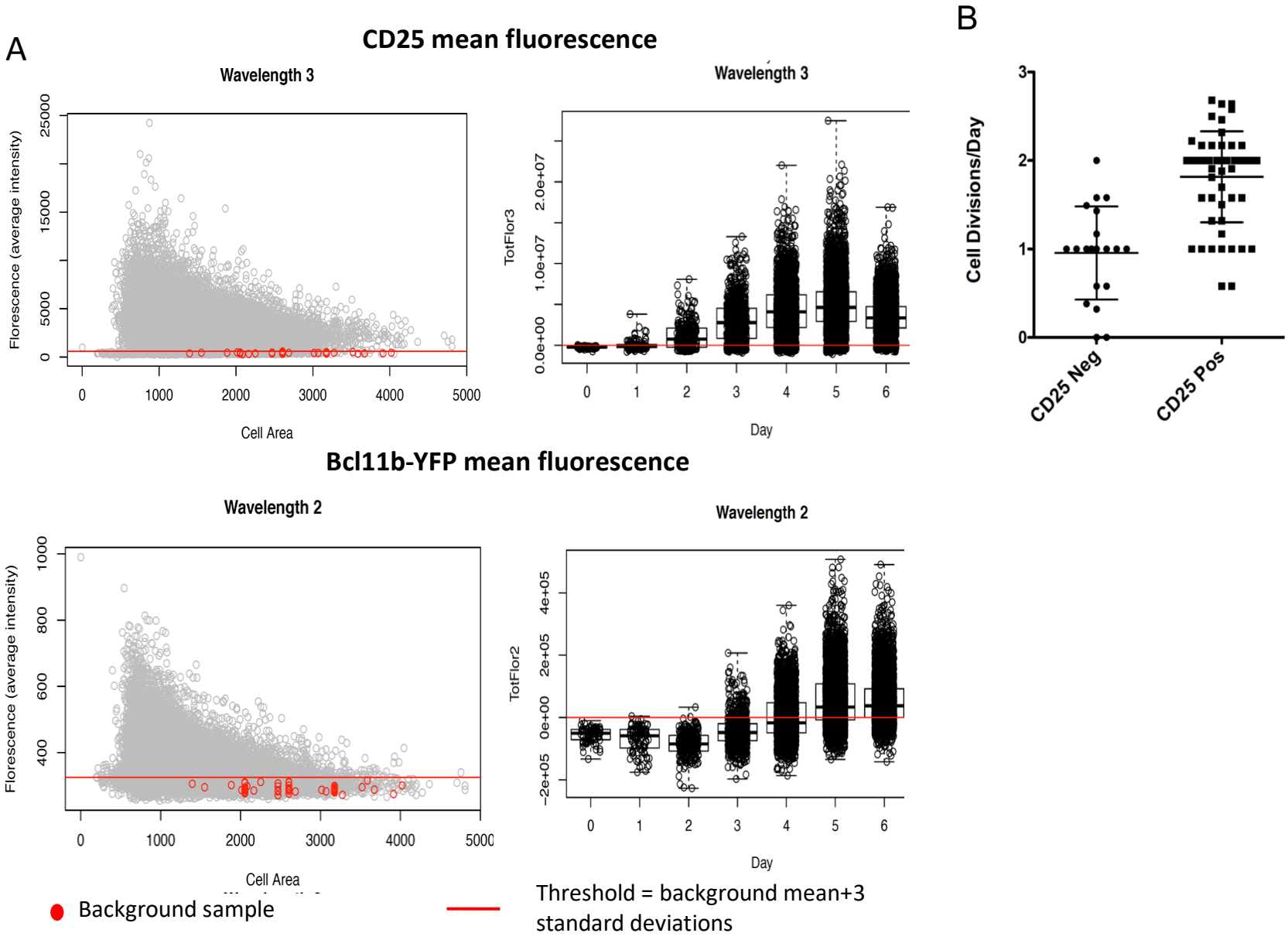
Supplementary Figure S5



Supplementary Figure S5, related to Fig. 6. Clonal variations in proliferation and differentiation rates of differentiating clones seeded from single DN1 cells.

Plots are of cell numbers (log scale) and numbers of CD25 and Bcl11b-YFP positive cells by day, within 60 individual clones that clearly entered the T-cell pathway in the course of the differentiation cultures, as determined by turning on CD25 by day 6 (see Methods). Note x axis scales: not all clones were followed for a full 6 days. Black circles=total cell numbers; pink triangles=number of CD25+ cells; green squares=number of Bcl11b-YFP+ cells; 0 values are not plotted.

Supplementary Figure S6

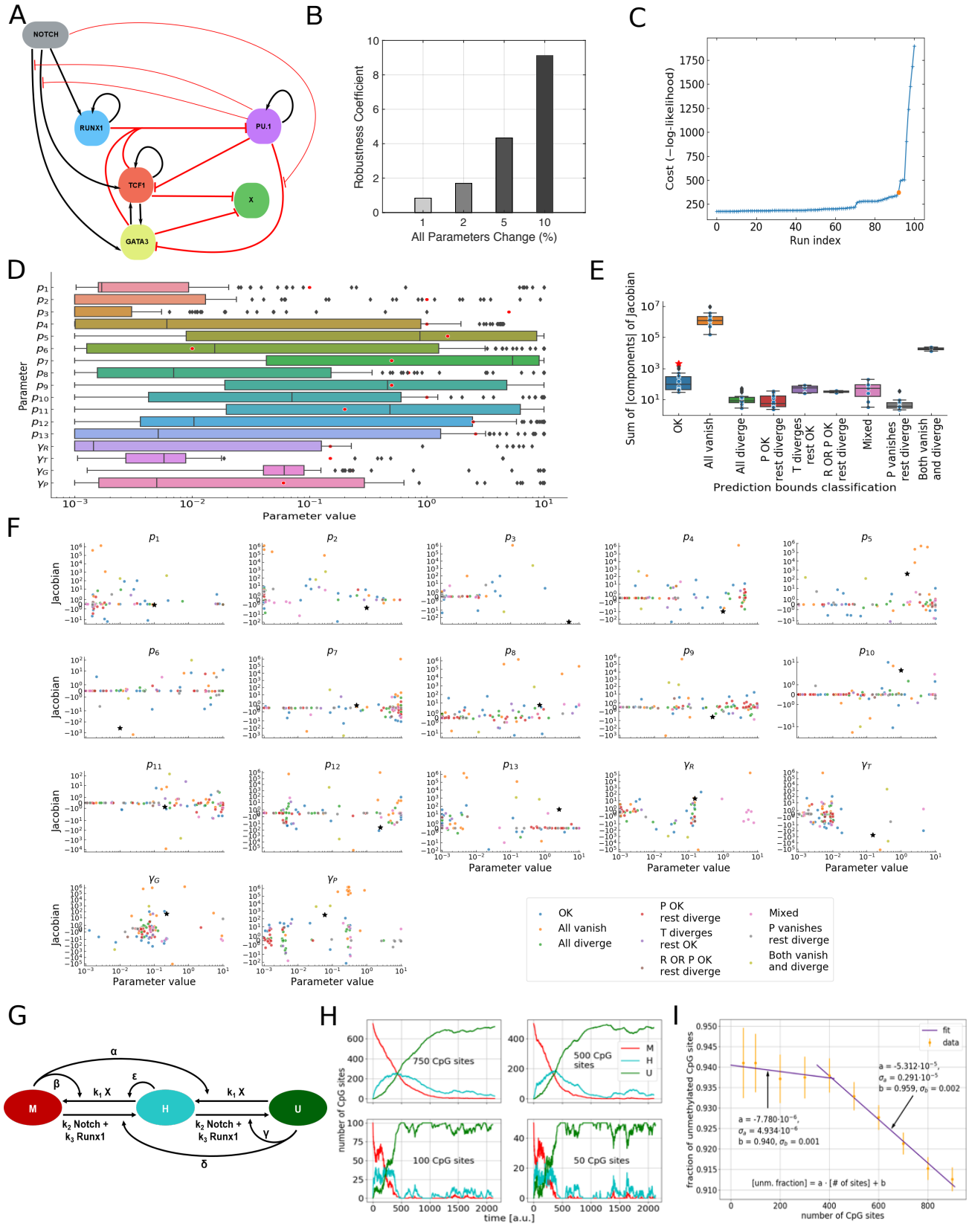


Supplementary Figure S6, related to Fig. 6. Fluorescence thresholds and cell division rates from in vitro clonal cultures monitored by live imaging.

A Fluorescence threshold level determinations for CD25 and Bcl11b-YFP expression. Plots of fluorescence values for ~13K segmented cells from all ETP clones and all times (left) and fluorescence values for all cells by day (right; median +/- 25th percentile). Background values were taken from 50 samples from 13 wells (red circles) and the average + 3 standard deviations (red line) was selected as the threshold level of expression for scoring a cell as being positive for CD25 or Bcl11b-YFP.

B The estimated number of divisions/day for clones in which all daughter cells were scored as either CD25 negative (left) or positive (right) for two consecutive days in early cultures, from days 1-3. Cells from clones that had turned on CD25 (DN2) proliferated more rapidly than those that had not turned on CD25 (still DN1) (DN1, n=20; DN2, n=55; $p<0.0001$ paired t-test).

Supplementary Figure S7



Supplementary Figure S7, related to Supplementary Text, Fig. 4 and Fig. 5. Characterization of the gene regulatory network and epigenetic de-repression layers of the model

A Gene regulatory network topology for the transcription model.

B Robustness analysis shows that the model is robust when all parameters are changed by 1, 2, 5 and 10 % of their values.

C Cost for the 100 optimization runs. The picked solution is marked with the orange dot.

D Distributions of the parameter values for the 100 optimization runs represented as box plots. Parameter values for the picked solutions are showed as red dots.

E The sum of the absolute values of the components of the Jacobian for each optimization run, classified according to the behavior of its prediction bounds. The red star marks the picked solution and ends up in the 'OK' prediction bounds cluster.

F Component of the Jacobian corresponding to each parameter against the parameter value. The picked parameter set is represented with black stars. They clearly lie close to zero in most cases, and where they do not, they are not extreme.

G Collaborative model (left panel) used to simulate the demethylation of the Bcl11b regulatory region. CpG sites can be methylated (M) -- closed state, hemi-methylated (H) -- intermediate state or unmethylated (U) -- open state.

H Number of methylated (M), hemimethylated (H) and unmethylated (U) CpG sites vs. time for methylated regions having sizes of 750, 500, 100 and 50 CpG sites. For low numbers of CpG sites noise is important. The size of the methylated region influences the fraction of unmethylated CpG sites after steady state.

I The fraction of unmethylated CpG sites decreases linearly with the number of CpG sites and the decrease is stronger for regions greater than 400 CpG sites.

SUPPLEMENTARY TABLES
Tables S1-S5

Table S1: Parameter Values for the Transcription/Specification Level Model, related to Star Methods, Single Cell Model: Transcription Factor Level and to Figures 3 and 4. Best-fit parameter values are shown for rate constants p_1-p_{16} governing predicted levels of Runx1, TCF1, Gata3, PU.1, Notch signaling and X, based on the pseudo-time-ordered smFISH measurements, as described in STAR Methods and optimized as shown in Fig. S7.

p_1	p_2	p_3	p_4	p_5	p_6	p_7	p_8	p_9	p_{10}	p_{11}
0.10	1.00	5.00	1.00	1.50	0.01	0.50	0.70	0.50	1.00	0.20
p_{12}	p_{13}	p_{14}	p_{15}	p_{16}		γ_R	γ_T	γ_G	γ_P	γ_X
2.50	2.60	2.00	1.00	0.01		0.15 h^{-1}	0.15 h^{-1}	0.23 h^{-1}	0.06 h^{-1}	0.02 h^{-1}

Table S2: Parameter Values for the Epigenetic Level Model, related to Star Methods, Single Cell Model: Epigenetic Level and to Figure 3 and 4. Best-fit parameter values p_{17}-p_{21} governing the opening and transcriptional activation of Bcl11b, based on input levels of Runx1, Notch, and X, as described in STAR methods and optimized as shown in Figure S7.

p_{17}	p_{18}	p_{19}	p_{20}	p_{21}
0.05	0.05	0.01	0.20	0.10

Table S3: Relative error between the simplified population model outcomes and the population developmental kinetics data, related to Star Methods, Multiscale-Model: Population Level and Figure 5. Table shows deviation from measured data in the simplified population model in which cell cycle lengths are assumed to be constant across all cell generations. Note high error in the early timepoints.

	DN1 day 2	**DN1 day 3**	**DN2a day3**	**DN2a day 3**
Relative error of distribution fit	16.9%	5.3%	17.4%	4.2%

Table S4: Average and standard deviation parameters for cell cycle length distributions for each generation fitted to experimental measurements, related to Star Methods, Multiscale-Model: Population Level and Figure 5 and 6. Table shows fitted cell cycle time parameters, with cell cycle times decreasing by generation, as used in the final population model.

generation	μ	σ
0	34	13
1	15	5
2	13	5
3	12	4
≤ 4	12	3

Units: (times in hr).

Table S5: Parameter values for the collaborative Epigenetic Level Model, related to Star Methods, Multiscale-Model: Epigenetic Level. Final parameter values are shown for the Collaborative Epigenetic Level Model, governing the number of closed epigenetic sites (modeled as methylated CpG residues) that would need to be opened with the indicated rate constants in order to fit the observed population kinetics in the integrated multilevel model.

k_1	k_2	k_3	α	β	γ	δ	ε
0.28	0.20	0.20	0.002	0.002	0.0005	0.0005	0.002