



T-cell commitment inheritance – an agent-based multi-scale model



Emil Andersson¹, Ellen V. Rothenberg², Carsten Peterson¹ & Victor Olariu¹ ✉

T-cell development provides an excellent model system for studying lineage commitment from a multipotent progenitor. The intrathymic development process has been thoroughly studied. The molecular circuitry controlling it has been dissected and the necessary steps like programmed shut off of progenitor genes and T-cell genes upregulation have been revealed. However, the exact timing between decision-making and commitment stage remains unexplored. To this end, we implemented an agent-based multi-scale model to investigate inheritance in early T-cell development. Treating each cell as an agent provides a powerful tool as it tracks each individual cell of a simulated T-cell colony, enabling the construction of lineage trees. Based on the lineage trees, we introduce the concept of the last common ancestors (LCA) of committed cells and analyse their relations, both at single-cell level and population level. In addition to simulating wild-type development, we also conduct knockdown analysis. Our simulations predicted that the commitment is a three-step process that occurs on average over several cell generations once a cell is first prepared by a transcriptional switch. This is followed by the loss of the Bcl11b-opposing function approximately two to three generations later. This is when our LCA analysis indicates that the decision to commit is taken even though in general another one to two generations elapse before the cell actually becomes committed by transitioning to the DN2b state. Our results showed that there is decision inheritance in the commitment mechanism.

Commitment of T-cell progenitors offers an opportunity to dissect the molecular circuitry establishing cell identity in response to environmental signals. This intrathymic development process encompasses programmed shutoff of progenitor genes¹, upregulation of T-cell specification genes, proliferation, and ultimately commitment². The core gene regulatory network controlling this process was identified from results from perturbation studies^{3,4}. It was also shown that Bcl11b expression correlates with the functional committed state cells and can thus serve as a proxy for commitment^{5,6}.

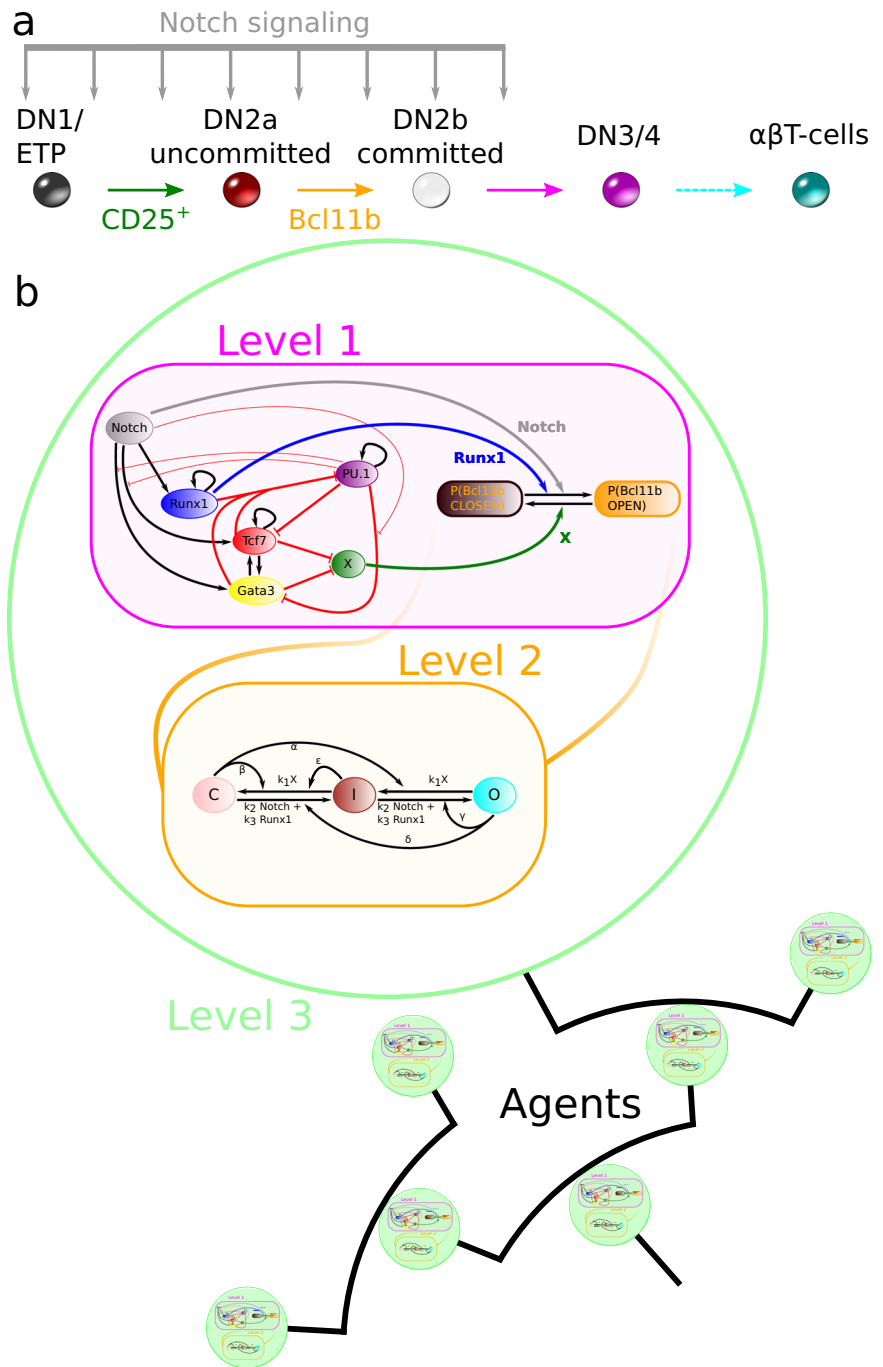
The stages of early T-cell development are well known experimentally. The T-cell progenitor cells are CD4 and CD8 double negative (DN) and do not express T-cell receptors. The Kit^{high} early thymic progenitors (ETPs or DN1s) transition to the DN2a state which is marked by CD25 surface expression^{7,8}. The DN2a cells upregulate the expression of Bcl11b, which correlates with the commitment to the T-cell lineage fate, as they transition to the DN2b state^{5,7–10}. The cells continue through the DN3 and DN4 stages and progress with the late T-cell development, however, these stages of development are not within the scope of this work. The cells proliferate

during the DN1 stage with an increasing proliferation rate as the cells progress to DN2. The steps of the early T-cell development are summarised in Fig. 1a and we refer to references^{1,2} for detailed reviews.

We have earlier developed and independently verified a model of the core gene regulatory network (GRN) governing the early stages of T-cell progenitor commitment¹¹. This GRN is built on experimentally shown interactions between T-cell specification genes Tcf7 (TCF1)^{5,12}, Gata3^{5,13} and Runx1^{4,5}, and the opposing Spi1 (PU.1) gene³. The dynamics governed by this network are influenced by an extrinsic T-cell positive input from Notch signalling present due to the thymic microenvironment. It has been shown that Runx1 levels control the timing of T-cell development¹⁴ and that Tcf7 is needed in order for Bcl11b to turn on^{5,15}. Connecting Bcl11b directly to the core GRN would predict an upregulation of Bcl11b synchronised with the expression of the T-cell-specific factors. However, it has been experimentally shown that Bcl11b is turned on a few cell cycles after CD25 expression and T-cell-specific factors upregulation^{5,16,17}. The expression of Bcl11b is controlled by a regulatory region consisting of many regulatory sites which each can be open, closed or in an intermediate state¹⁸. If enough sites become

¹Computational Science for Health and Environment, Centre for Environmental and Climate Science, Lund University, Lund, Sweden. ²Division of Biology and Biological Engineering, 156-29, California Institute of Technology, Pasadena, CA 91125, USA. ✉e-mail: victor.olariu@cec.lu.se

Fig. 1 | Agent-based multi-scale model for early T-cell development commitment. a Schematics over the in vivo early T-cell development. Early thymic progenitors (ETPs or DN1) transition to the DN2a state which is marked by CD25 surface expression. Commitment to the T-cell fate is observed by Bcl11b upregulation as the cells progress to the DN2b state. The T-cell lineage development continues through DN3 and DN4 stages and eventually becomes mature T-cells. The early T-cell development takes place under the influence of Notch signalling inside the thymus. **b** Depiction of the multi-scale agent-based model for the T-cell development stages from DN1/ETP to DN2b. As CD25 is a surface marker it is not included in the model. The magenta-coloured box illustrates level 1 and contains the GRN topology. The black arrows and thick red blunted arrows represent positive and negative direct regulation respectively. The thin red blunted arrows represent inhibition of regulation. The blue and grey arrows represent that Runx1 and Notch promote the opening of Bcl11b regulatory sites, while the green arrow shows that X keeps the sites closed. The orange box illustrates the epigenetic mechanism of level 2. The regulatory sites can change between three different states (closed, intermediate and open) and are affected by input signals from level 1. Each cell of level 3 (green circles) contains a copy of levels 1 and 2. The agent-based model implementation tracks the relation between the proliferating cells in lineage trees.



open, the whole region is considered to be open and Bcl11b is expressed. Furthermore, the ultimate activation of Bcl11b has been found to depend on a slow cis-acting epigenetic mechanism, because direct observation of live cells showed that the two equivalent alleles within a single cell can be activated days apart¹⁹, and activation timing is strongly modulated by repressive histone marks across the regulatory region¹⁷. Therefore, we proposed a model for a transcription level of regulation which propagates into an epigenetic level of Bcl11b regulation¹⁸. This single-cell model was trained with data from RNA fluorescence in situ hybridisation (FISH). Moreover, this two-level regulation model was augmented with a proliferation level, resulting in a multi-scale model which allowed us to investigate the T-cell commitment mechanism at bulk level. The predictions from this model were validated with clonal kinetic data. Our multi-scale model predicted that DN1 population developmental heterogeneity can arise solely from GRN

noise. It also showed that the observed heterogeneous delay in lineage commitment marked by Bcl11b arises both from GRN and epigenetic stochasticity. Furthermore, we observed that the delay between the loss of the Bcl11b opposing function X and expression of Bcl11b was similar to the delay between the CD25 surface expression and the expression of Bcl11b. However, we could not specify the exact timing of the decision to commit with respect to the observed Bcl11b upregulation or whether there is an inheritance of progression towards the T-cell fate.

To this end, we have further developed our model by putting the full multi-scale model into an agent-based setting. By this augmentation, we unlock the ability to study in silico single-cell properties of T-cell development. With this framework, we can investigate if inheritance of decisions plays a role in the T-cell commitment mechanism. We do so by studying lineage trees of simulated T-cell colonies and introducing the concept of last

common ancestors (LCAs) of Bcl11b-positive cells. With inheritance of decision we mean that a cell is influenced by events that occurred in its ancestors, e.g. event *B* could only occur if event *A* already happened in an ancestor cell. If decisions are not inherited, then events may occur in any order without affecting cells downstream, i.e. pure stochastic decisions. We find that inheritance of decisions does indeed play an important role and that the commitment mechanism takes place over several cell generations. Importantly, we find that a cell's decision to commit can actually be made one to two generations before Bcl11b is upregulated.

Results

The model

We investigate decision inheritance in T-cell commitment by performing in silico experiments. We propose an agent-based model with a core consisting of our previously published stochastic multi-scale model¹¹. By treating single cells as separate agents, we can construct lineage trees corresponding to an entire cell colony, enabling us to access details like inheritance and relations between cells which was previously not readily available. The model consists of three levels: level 1 is a transcriptional level inside a cell, containing a GRN governing the core differentiation process; level 2 is an epigenetic level, also inside a cell, where the Bcl11b regulatory region is treated by opening and closing of regulatory sites depending on signals from the transcriptional level; and finally, level 3, a proliferation level, where the cells as separate agents divide and pass down properties to their descendants (see Fig. 1b). The mathematical details of all three levels in the agent-based model are described in Methods. Differentiation is simulated starting from the equivalent of the ETP stage. Note that although Bcl11b can be turned on in many DN2a precursors without cell division⁵, ETPs normally undergo several cell cycles before turning Bcl11b on¹¹. Hence, our model encompasses multiple cycles of cell division.

Level 1: transcriptional level. The gene regulatory network in the transcriptional level consists of 6 interacting genes: Runx1, Gata3, PU.1 (Spi1), Tcf7 (TCF1), Notch signalling, and a Bcl11b inhibitory function *X* (see Fig. 1b). *X* includes a slow initial chromatin opening mechanism and other possible Bcl11b-antagonists, inhibiting the opening of the Bcl11b regulatory sites. Runx1 and Notch act positively for opening the Bcl11b regulatory sites and Notch is promoting Runx1, Tcf7 and Gata3. Both Tcf7 and Gata3 operate towards opening the Bcl11b regulatory sites by inhibiting *X* and PU.1. In turn, PU.1 inhibits Tcf7 and Gata3 as well as their Notch activation. Experimental data supporting these linkages were from refs. 3,5,13,15,20–24; these were largely confirmed in a comprehensive update using recent genome-wide perturbation analysis data¹⁴. Experimentally, CD25 marks the transition from ETP to DN2a, but since CD25 is just a surface marker, we have not included this gene in the model. Intrinsic transcriptional noise is present in experimental data¹¹, thus we simulate the GRN stochastically with the Gillespie algorithm²⁵ (see Methods). Bcl11b is not treated directly as a node in the GRN, but is instead regulated in level 2. The communications between the transcriptional level and epigenetic level are conducted through the signals from *X*, Runx1 and Notch.

Level 2: epigenetic level. In vivo and in vitro, Bcl11b becomes expressed in the transition from the T-cell development stage DN2a to DN2b. In our in silico model, Bcl11b is regulated by an epigenetic mechanism consisting of the opening and closing of regulatory sites instead of treating the gene expression directly. The 500 Bcl11b regulatory sites can either be closed (C), intermediate (I) or open (O). Cells at generation 0 have the Bcl11b regulating region completely closed. Cells with more than 75 % of the regulatory sites open are considered to have an open Bcl11b regulatory region, corresponding to Bcl11b being expressed. When the Bcl11b regulatory region has opened up, the cell is considered to have committed to the T-cell lineage fate⁵. The epigenetic level is regulated by the transcriptional level. Runx1 and Notch increase the probability of opening regulatory sites while *X* increases the probability of

closing them. More details about the stochastic implementation can be found in Methods.

Level 3: proliferation level. The proliferation level of the model simulates an entire cell population. The simulations always start with one individual cell at generation 0. The dynamics of the multi-scale model containing the transcriptional and epigenetic levels inside the cell are simulated over time. Cell division is also implemented with the daughter cells containing the same multi-scale model as the mother, continuing the transcriptional and epigenetic simulations. The colony evolves through multiple divisions and its simulation corresponds to 120 h in experimental time. We sample the division times from fitted distributions that vary with cell generations, to account for cell cycle length variability experimentally observed¹¹. This creates heterogeneity of the age of cells from different generations, i.e. one simulated colony may undergo 6 cell cycles while another may undergo 11.

During cell division, all the gene expression levels are copied to the daughter cells, keeping them at the same levels as inside the mother cell. However, the divisions follow a set of conditional rules imposed on the regulatory sites¹⁸. If *X* is expressed, the regulatory sites are copied from the mother cell, while the complete loss of *X* promotes the opening of the regulatory sites (see Methods). Thus, a strong driving force to opening up the Bcl11b regulatory region is the repression of *X*.

Agent-based model. Our agent-based implementation enables us to record the relations between the cells within a lineage tree. This way, we can investigate the existence of inheritance in T-cell commitment. This type of modelling makes it possible to know the transcription levels, epigenetic status and age of each cell at any simulation time point. Having access to this type of cell information is very informative for future experimental efforts.

Investigating commitment inheritance

As illustrated in Fig. 1a, surface expression of CD25 marks the transition from ETP to DN2a and Bcl11b expression marks the advent of T-cell commitment in experiments⁵. Bcl11b heterogeneity was shown in colonies of differentiating T-cell progenitors both in vitro and in silico in Olariu et al. (2021)¹¹. Furthermore, the model predicted a delay between when the colonies reached 50 % of the cells with loss of *X* activity and when 50 % of cells in these colonies became Bcl11b positive. Similarly, for in vitro experiments, a large heterogeneity was observed for the number of cell divisions and the time when the colonies reached 50 % CD25 positive cells and 50 % Bcl11b positive cells.

These findings raise a few questions which we attempt to answer using our new agent-based implementation of the multi-scale model:

- What are the mechanisms leading to the observed heterogeneity within a single colony?
- When is the decision taken to open or keep closed a cell's Bcl11b regulatory region?
- How are system perturbations affecting the decision to commit?

Lineage trees. To answer the questions above, we use a tool that represents the simulated colonies as phylogenetic trees²⁶ or lineage tree diagrams. Figure 2a illustrates a simple version of a lineage tree, with key features highlighted. The initial cell at generation 0 is the root of the tree and every cell division leads to two new branches. Each cell is a node in the tree and the radial length of a connection in the tree is proportional to the cell's lifetime, i.e. the time-axis points radially outwards. The colour of the node represents the status of the cell where black depicts a cell with the Bcl11b regulatory region closed and *X* expression greater than 0, a red node represents a cell where the Bcl11b is closed but *X* is not expressed, and white represents a cell where the Bcl11b region is open. With this colouring scheme, one can observe heterogeneity of open Bcl11b within a colony along with the variable delay between the loss of *X* and the opening of Bcl11b.

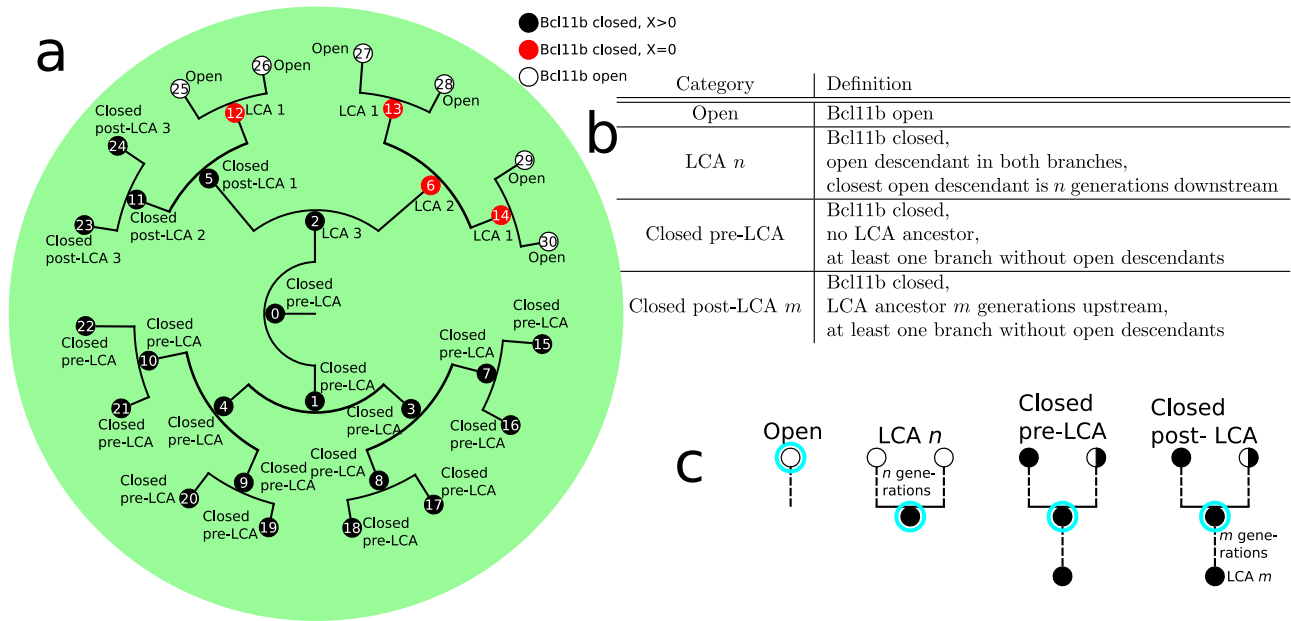


Fig. 2 | Small lineage tree and last common ancestor (LCA) definitions. **a** A lineage tree example. Every node is a cell which is uniquely identifiable by an index. Each cell branches into two daughter cells. The radial connecting lines are proportional to a cell’s lifetime. The colour of the node indicates the status of the cell’s Bcl11b regulatory region and X expression, as described by the legend. Each cell is labelled with

its LCA label. **b** Definitions of the LCA categories. **c** The graphical definition of each LCA category for the cells encircled with cyan. Note the subtle difference between ‘closed post-LCA m ’ and ‘closed pre-LCA’, i.e. the only difference is whether they have an LCA-ancestor or not.

Last Common Ancestors. Since the delay between CD25 activation and Bcl11b opening varies between cells in experiments, both in time and number of divisions, we have to track time relative to interesting events rather than experimental time or generation counting when analysing our model simulation outcomes. The cells’ main characteristic considered here is whether the Bcl11b regulatory region is open or closed. This is due to the fact that a cell with Bcl11b open has normally committed to the T-cell fate and has become DN2b. Therefore, we introduce a system of categorising the in silico cells uniquely according to their relations with cells with open Bcl11b within a simulated lineage tree (see Fig. 2b and c).

We define the first category ‘open’ which includes all cells with open Bcl11b regions. Two arbitrarily chosen cells from the ‘open’-category which have mother cells with Bcl11b closed have a last common ancestor (LCA) which is the last Bcl11b-closed cell they both originate from. Therefore, we define any Bcl11b-closed cell that has at least one ‘open’ descendant in each of its two daughter branches as an ‘LCA’-cell. The ‘LCA’-cells are further classified depending on how many generations down the lineage the closest ‘open’ cell is located. For instance, if a closed cell has both of its daughter cells ‘open’, it is of order 1, i.e. categorised as ‘LCA 1’. If a closed cell instead has an ‘open’ cell five generations downstream in one branch, but six generations downstream in the other branch, it is of order 5, i.e. ‘LCA 5’.

Bcl11b-closed cells upstream from the earliest LCA in a lineage are defined as ‘closed pre-LCA’. The so-far unclassified cells are Bcl11b-closed but located downstream from LCAs. These are defined as ‘closed post-LCA’. These cells are also given an order depending on how many generations downstream from the latest LCA the cell is. The definitions of the four categories are summarised in Fig. 2b and c.

All the cells in the example tree in Fig. 2a have been categorised according to this system. From this tree, it is clear that a colony can be heterogeneous in Bcl11b. The root cell, cell 0, is ‘Closed pre-LCA’ since every cell in its lower branch, rooted in cell 1, is closed. The top half of the tree, rooted in cell 2, has both open and closed sub-branches. Cell number 2 is an LCA 3’ cell since both its daughters have descendants which are ‘open’, where the closest is three generations away. Cell 2 together with cell 5 (which is a ‘closed post-LCA 1’-cell) are interesting cells since their branches have

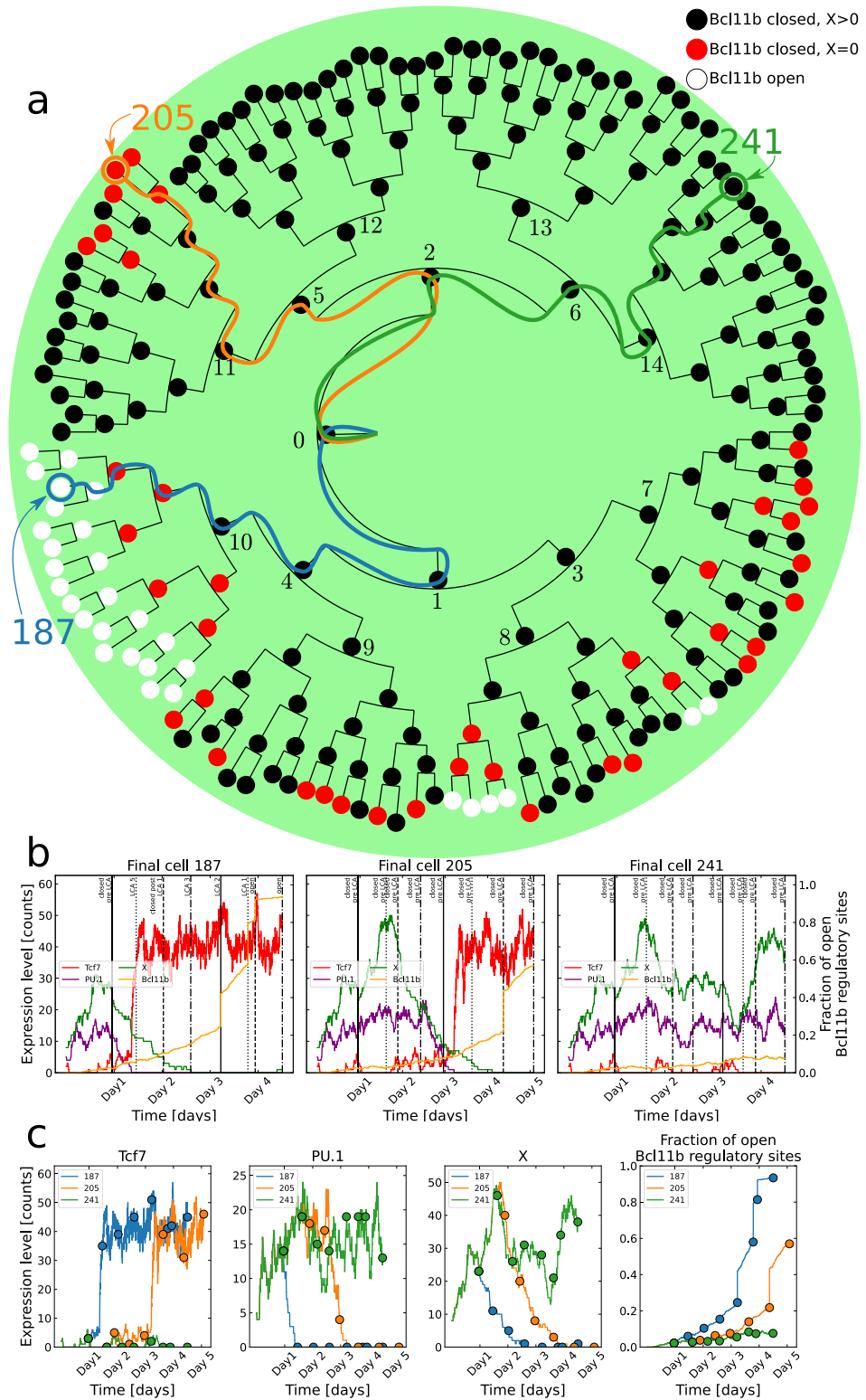
different fates; some sub-branches open up while others stay closed. Branching points like these are key points to investigate in order to elucidate the inheritance mechanism in T-cell commitment.

By defining these categories, we can examine the internal relations between the cells in a tree in a generation-number-independent way. This is important since the cells in some colonies in our model may only go through 5 divisions while the cells in other colonies could go through 10 divisions over the time span of the simulation, consistent with the wide variation in cell cycle numbers between entry into DN2a and the onset of Bcl11b expression that our experimental results showed previously¹¹. The LCA-category system is a tool to dissect when the decision to undergo commitment happen, as distinct from the execution of commitment itself. Presumably, no decision should have been made before an LCA-cell, thus, all ‘closed pre-LCA’-cells should have similar gene expressions. By indexing the LCA-cells, it is possible to track how many generations before the observed cell state transition the actual decision was taken and then relate this decision to the most reproducible features of gene expression at this point. One example when the decision to commit would have to be taken very close to the actual cell state transition is if all LCAs of order 2 or higher are similar to ‘closed pre-LCA’ in their gene expression states. Another very different example would be if LCA-cells of order 6 express similar gene network activity traits as open cells, then the decision to commit is taken long before the observed transition. The ‘closed post-LCA’-cells represent those cells that fell short of obtaining completely open Bcl11b regulatory region and does also include cells that exited the T-cell lineage fate.

In silico simulations of lineage trees

We simulated 300 wild-type (WT) T-cell committing colonies, all identically initialised in the ETP cell state, and produced corresponding lineage trees for each colony. Figure 3a shows an example lineage tree and in Supplementary Figs. 5 to 10 a larger subset is presented. The lineage tree in (Fig. 3a) branched seven times, corresponding to seven divisions, yielding 128 final cells. Out of these, 22 are open, all located at the lower half of the tree. Every cell in the branch rooted at cell 10 is open, and the remaining open cells are all located in the branch rooted at cell 8 (from here, a branch rooted at cell n is shorted to branch n). The cells in branch 6 show no sign of

Fig. 3 | Stochastic simulations of cell lineages.
a Lineage tree for a simulated colony. Three different lineages with different fates are marked with blue, orange, and green lineage paths respectively. Black nodes depict cells with the Bcl11b regulatory region closed and X expression greater than 0, red nodes represent cells where the Bcl11b is closed and X is depleted, and white nodes represent cells where the Bcl11b region is open. **b** Each panel shows the gene expression dynamics for each of the three marked lineages respectively. The vertical lines represent cell divisions and are labelled with the cells' corresponding last common ancestor (LCA) categories. **c** Each panel show the expression for Tcf7, PU.1, X and the fraction of open Bcl11b regulatory sites respectively for the three marked cell lineages. The coloured dots indicate cell divisions.



being about to open up, while a few cells in branch 5 have lost X expression but are still closed.

This tree shows clear heterogeneity in the different branches which makes it a prime candidate to investigate what happens during development at different important branching points. The simulated gene expression levels are shown for three chosen cell lineages in Fig. 3b, where each panel

shows the evolution of one lineage. The three lineages are 187 (blue path in Fig. 3a which is an open cell in branch 10; 205 (orange path) which is a closed cell in branch 5, but with X depleted; and 241 (green path) which is a closed cell in branch 6. The vertical lines in Fig. 3b indicate cell divisions and are labelled with the cells' LCA labels. Figure 3c shows the computed expression of Tcf7, PU.1, X and the fraction of open Bcl11b regulatory sites in one panel

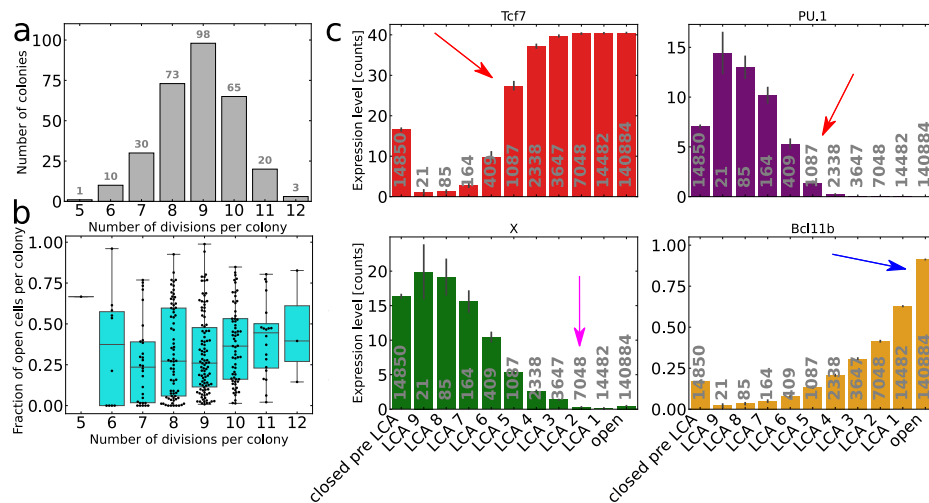


Fig. 4 | Last common ancestor (LCA) statistics. **a** The number of colonies reaching each colony size. **b** Boxplot over the distribution of the fraction of open cells per colony for different-sized colonies. Each dot represents a colony. The centre line of a box is the median, the bounds of the box represent the first and third quartile, and the whiskers extend to 1.5 times the interquartile range. Points falling outside of this range are shown as outliers. **c** Mean expression level for Tcf7, PU.1, X, and the fraction of open Bcl11b regulatory sites for cells belonging to the different LCA

categories. The categories are ordered in an approximate developmental order. The grey numbers indicate the number of cells belonging to each category. The red arrows point out the preparatory switch towards the T-cell fate. The pink arrow indicates the most common LCA stage where X function is lost. The blue arrow shows that the opening of the Bcl11b regulatory region is delayed with two generations. The error bars represent standard deviations.

each to more conveniently compare how the activity of these genes are calculated to differ in the three lineages. Here, the coloured dots indicate cell divisions.

The three lineages only have the first cell in common. It is clear that already during the second generation of cells, the blue path starts to become different from the green and orange paths. PU.1 decreases at the same time as Tcf7 starts to increase. These two components are tightly connected in the GRN (Fig. 1b). PU.1 represses Tcf7 both directly and through repression of the Tcf7-activating Notch signal, while Tcf7 jointly represses PU.1 together with Runx1 and Gata3. Runx1 is stable over time and Gata3 follows Tcf7 but is expressed at a much lower level. Therefore, it is probably a small fluctuation of either decreasing PU.1 or increasing Tcf7 which throws the preparatory switch towards the T-cell fate to start producing an excess of Tcf7. Once there is an abundant amount of Tcf7, PU.1 is firmly repressed while Tcf7 is kept at a high level by both its direct self-activation and its indirect activation through Gata3. The orange lineage also goes through this switch, but at a later time point on day 3, while the green lineage does not go through this switch at all. Once Tcf7 (and Gata3) is highly expressed, X is repressed. For the blue lineage, X is depleted before day 3 while, for the orange lineage, this happens at day 4. On the other hand, X stays highly expressed in the green lineage where PU.1 is also high and Tcf7 is low. For the blue and orange lineages, when Tcf7 has gone through the switch and X starts to decrease, the fraction of open Bcl11b regulatory sites slowly starts to increase. However, when X is completely depleted, the epigenetic remodelling rules promoting the opening of the Bcl11b regulatory sites begin to operate, resulting in Bcl11b’s accessibility making great increasing leaps at every cell division. The blue lineage opens up Bcl11b two generations after X is depleted, exhibiting a delay between the loss of X and the opening of Bcl11b. Presumably, in the orange lineage, Bcl11b would also open up after the next division, if the simulation would have been longer; the orange lineage exhibits similar traits as the blue one but roughly two days delayed. The green lineage is similar to the early behaviour of the blue and orange lineages and would probably also be able to throw the preparatory T-cell fate switch if the simulation was run much longer, this way giving the system a chance to achieve high expression of Tcf7 and low PU.1. Two additional lineages are shown in Supplementary Fig. 1 in purple and red. The purple lineage represents a cell lineage which progress from ‘LCA’ to ‘closed post-LCA’ back to ‘LCA’ again. The red cell lineage goes from ‘LCA’ to ‘closed

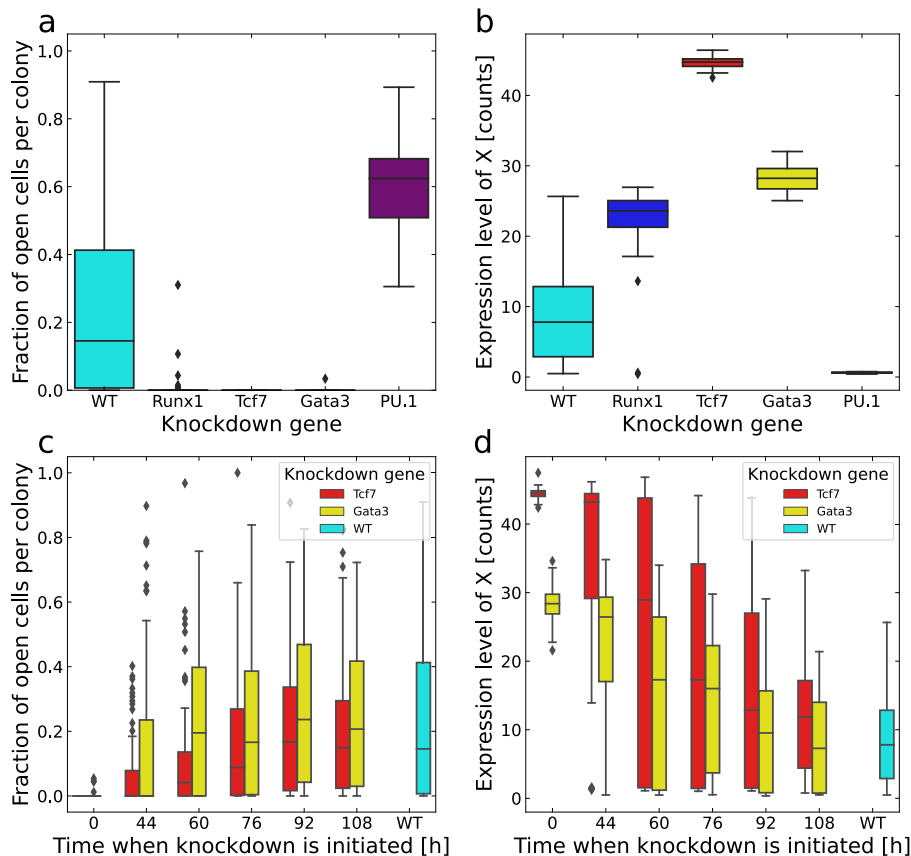
post-LCA’, but does not return to ‘LCA’. Both the lineages undergo the transcriptional switch when they are ‘LCA’. The difference between the lineages is that the red lineage has X expression for a longer time, with the result that the Bcl11b regulatory sites only open up as governed by Eq. (2). Thus, the red lineage does not open up as many sites at once step-wise as the purple lineage does once X has becomes depleted.

Statistics of multiple simulated cell colonies. In the previous section, we dissected the gene expression dynamics for three specific cell lineages in the lineage tree shown in Fig. 3. In this section, we use the LCA labels (Fig. 2b, c) to gauge the commitment dynamics at the population level over all 300 colonies.

The number of divisions in the 300 simulated colonies spanned between five and twelve, with eight to ten divisions being the most common (see Fig. 4a). The colonies vary in sizes since the cell cycle lengths are randomly sampled from the distributions specified in Table 3, which were fitted with experimental data in Olariu et al. (2021)¹¹. Figure 4b shows that the model generated a variation in the fraction of Bcl11b-open cells per colony. The colonies that divided 7, 8 or 9 times had similar median fractions of open cells, while colonies that divided 10 or 11 times had a slightly higher fraction of open cells. The reason for this is that the larger colonies had the possibility to divide more times after cells started to become Bcl11b-open compared to the smaller colonies. There were too few colonies with 5, 6 and 12 divisions to make any remark about their distributions. We also observe that the absolute cell generation number does not seem to be of particular importance for the commitment process. Thus, in order to meaningfully compare the commitment mechanisms between colonies, we need a reference point to count cell divisions from. This is provided by the LCA classification.

The colonies were classified according to their LCA properties as described in Sections Last Common Ancestors and Methods and Fig. 6. The mean expression for each gene was calculated for each LCA category, using the cells’ final gene expression values before division. The mean expression levels for Tcf7, PU.1 and X and the fraction of open Bcl11b regulatory sites for the ‘closed pre-LCA’, ‘open’ and ‘LCA n’ categories are shown in Fig. 4c. The categories are sorted in an approximate developmental order. The grey numbers in each bar indicate the number of cells belonging to each LCA category. The mean

Fig. 5 | Knockdown simulations. **a** Statistics on cells with open Bcl11b regulatory regions from KD simulations of Runx1, Tcf7, Gata3 and PU.1 compared to WT simulations where 60 colonies per simulation were considered. Both the genes' initial transcription counts and production rates are reduced to 20% of the original values. **b** Statistics on the expression level of X from the same simulations as in (a). **c, d** Knockdown simulations of Tcf7 and Gata3 at different time points compared to WT simulations with 180 colonies per simulation type. **c** shows the distribution of the fraction of cells with open Bcl11b regulatory regions per colony and **d** shows the expression level of X. The centre line of a box is the median, the bounds of the box represent the first and third quartile, and the whiskers extend to 1.5 times the interquartile range. Points falling outside of this range are shown as outliers.



expressions for all the genes and for the ‘closed post-LCA *m*’ categories are shown in Supplementary Figs. 2 and 3.

Figure 4 c shows the development steps from early closed cells to open cells. A first notable feature is that the early LCAs (LCA 9-6) show very little in common with the later LCAs and open cells. It is first at LCA 5 and later where LCA-cells start to have more similar gene expression to the open cells. It is only the Runx1 level that stays fairly stationary through all stages (Supplementary Fig. 2). At LCA 5, the Tcf7 level becomes substantially higher than it is for ‘closed pre-LCA’, while PU.1 becomes substantially lower, as indicated by the red arrows. The high value of Tcf7 and low value of PU.1 ensures that X, after a few generations of GRN simulated dynamics, will become depleted, see the pink arrow at LCA 2. The division rules governed by the absence of X lead to the descendant cells opening up the Bcl11b regions one or two generations after depletion of X at ‘LCA 2 or 1’, see the blue arrow. This chain of events suggests that the ETP cells acquire properties promoting commitment to the T-cell fate several generations before the observed transition to the DN2b state. This result argues for a decision to enable T-cell commitment that is inherited by LCA descendants before it is executed. The events highlighted in the example lineages in Section In silico simulations of lineage trees and Fig. 3 agree well with the statistics gathered from the 300 simulated lineage trees.

The mean gene expressions for the cells that are closed post-LCA are shown in Supplementary Fig. 3. Note that the ‘closed post-LCA’-cells have two possible future states: they can either continue being closed or become LCA-cells again. Returning to LCA-state is possible if downstream cells in an only-closed-cells branch become open, as illustrated in Supplementary Fig. 1. Supplementary Fig. 3 shows that ‘closed post-LCA 1-3’ cells are all similar to ‘LCA 5-4’ regarding Tcf7 and PU.1 expressions, see the red arrows. These cells can in the following generations either go back to be LCA-cells and eventually open up, or they can continue to stay closed and move to higher ‘closed post-LCA’-orders, as illustrated in Supplementary

Fig. 1d. Depletion of X by the stochastic process probably governs whether the cells return to the LCA-path, or if they stay closed. The cells that stay closed, i.e. high order of ‘closed post-LCA’, return to a ‘closed pre-LCA’-like state.

In silico knockdown simulations. We performed in silico knockdown (KD) simulations of Runx1, Tcf7, Gata3 and PU.1 by reducing the production rate and initial expression level of the respective knocked-down gene. Deterministic simulations showed that knocking down the production rates of each gene to 20% (i.e. 5× reduction) perturbs the system sufficiently to result in a dramatic change of the dynamics (Supplementary Fig. 4). For each KD experiment, we simulated 60 colonies each for every KD gene. Figure 5a and b show the average fraction of open cells and the average expression level of X per colony at day 5, respectively. The results in Fig. 5a show that knockdown of PU.1 leads to more Bcl11b-open cells than wild-type (WT) while knockdown of Runx1, Tcf7 and Gata3 results in very few Bcl11-open cells. More specifically, the KD simulation of Tcf7 yielded no open cells while the knockdown of Gata3 led to only one colony with open cells. The average X expression is increased by KD of Runx1, Tcf7 and Gata3, with the greatest effect from Tcf7. KD of PU.1 instead depletes X. Notably, the distributions of the X expression under KD conditions become more narrow compared to the WT, suggesting that the colonies become less heterogeneous during KD.

To further dissect the commitment mechanism, we performed simulations where Tcf7 and Gata3 were knocked down at different simulation time points. We simulated KD of the two genes after 0, 44, 60, 76, 92, and 108 h for 180 colonies for each KD time point and gene. Figure 5c shows the distribution of the fraction of open cells after 120 hours, across 180 colonies for each KD time considered. The KD of Tcf7 is shown in red and Gata3 in yellow. When KD was initialised from the start, almost none of the colonies had any open cells at the

end of the simulation, in accordance with the results shown in Fig. 5a. The fraction of open cells increases with later initialisation of KD where later KD leads to behaviour more similar to WT. The same effect is seen in Fig. 5d for the average expression of X. X is high when Tcf7 or Gata3 KD is conducted early and it decreases with later KD. These results are in accordance with the observed irreversibility of the T-cell commitment process and is not something that cannot be inferred from the GRN topology alone. This conclusion is also in accord with previous experimental data demonstrating that Tcf7 and Gata3 are stage-dependent and are more important during the DN1 than the DN2 stage, and also that T-cell commitment is severely hindered by Tcf7 knockdown^{5,20}.

Commitment decision inheritance results

The model was constructed to incorporate known sources of noise in the experimental data¹¹. The observed intra-colony heterogeneity for the Bcl11b regulation site states along with the timing for its opening and closing in the simulated data can be explained by multiple stochastic elements. One important source of stochasticity is the intrinsic noise in the governing GRN. Our simulations showed that the fluctuating expression levels of the transcription factors are responsible for activating the switch towards the T-cell fate. Furthermore, the varying cell cycle length of the cells introduces both intra-colony stochasticity through the unsynchronised cell divisions, and inter-colony stochasticity resulting in different-sized colonies. The stochasticity from the cell cycle lengths affects the epigenetic level since cell division accelerates the opening of the Bcl11b regulatory sites once X has become depleted. This can be seen in Fig. 4b where colonies that underwent more cell divisions have a higher fraction of Bcl11b-open cells. Thus the combined effect of the transcriptional and epigenetic noise and varying cell cycle lengths contributes to the unsynchronised opening of the cells and creates heterogenous cell colonies.

A simulated cell cannot commit to the T-cell lineage without first being prepared by activating the transcriptional switch towards the T-cell fate. However, the decision to open Bcl11b and actually commit to the T-cell fate is taken when the function X is depleted. Until then, the cells can still escape from the T-cell lineage fate as shown by the analysis of the 'Closed post-LCA' cells (Supplementary Fig. 3). This is also supported by the knockdown analysis at different time points (Section In silico knockdown simulations and Fig. 5c and d). The fact that early KD of Tcf7 and Gata3 lead to no or just a few open cells along with high X expression shows that the cells indeed need to be prepared by the transcriptional switch in order to open. Additionally, the late knockdown had no effect compared to WT which shows that Tcf7 and Gata3 become redundant after the cells have committed. Since X expression is kept high after early KD and is lost one or two generations before cells open, we can conclude that the loss of X activity is instrumental for commitment. The KD simulations also show that both Tcf7 and Gata3 are required for the ETP cells to be able to commit to the T-cell fate, even though Tcf7 is more strongly expressed and shows a clearer switch-behaviour (Fig. 3b). Therefore, our results show that there is a one or two-generation delay between the decision to commit and the acquisition of the T-cell committed state, i.e. the Bcl11b-open state.

The LCA analysis showed that not all Bcl11b-closed cells have the same properties. Some closed cells are on the path of becoming Bcl11b-open and are very similar to the open cells in terms of gene expressions. These cells may either be between the switch towards the T-cell fate and X depletion (i.e. between preparation and deciding to commit) or after X depletion but before opening (i.e. between deciding to commit and actually committing). Other closed cells are more similar to the starting ETP cell state. Then the cells are either 'Closed pre-LCA' or they can also be 'Closed post-LCA'. For the latter case, the cells have gone through the preparatory switch but X never became depleted, thus not allowing Bcl11b to become open, resulting in the cells eventually escaping from the T-cell lineage fate. However, gaining back X function after complete depletion is a rare event and it is due to the stochastic implementation of the function X as a transcription production function. As shown in Fig. 4c, the statistics over the 300 simulated

colonies clearly reveal that X function is most likely lost at the 'LCA 2' stage which is roughly two generations before the Bcl11b regulatory region is opened.

Discussion

In this study, we have used a new agent-based version of our previously published multi-scale model for early T-cell development to explore the role of inheritance of decisions that empower cells to undergo commitment to the T-cell fate. We developed an analysis tool based on the concept of last common ancestors in phylogenetic analysis²⁶ and applied it to T-cell lineage trees. Here, we defined different LCA categories based on the uncommitted cells' relations to committed DN2b cells. Then we uniquely classified each cell in a simulated lineage tree into the defined categories. We performed single-cell simulations for proliferating wild-type T-cell colonies along with simulations of key gene knockdown. Analysis of both individual colonies as well as statistics for all colonies resulted in an agreeing picture of the T-cell commitment process. The commitment process is a chain of events taking place during multiple cell generations, initiated by the stochastic nature of transcription. The cells first need to be prepared for commitment, which happens if the expression level of PU.1 becomes low, in combination with Tcf7 expression level increase. The governing GRN drives the genes' expression levels towards this state; however, the analysis of individual cell lineages showed that there is a large heterogeneity in the timing of switching to this state. When a cell has been prepared by this switch towards the T-cell fate, the Bcl11b-opposing X function can become depleted. It has been shown both in vitro and in silico^{5,11} that there is a delay between depleting X (speculated to have the same timing as CD25 upregulation) and Bcl11b-opening. However, the role of X depletion was not completely revealed. In ref. 11, we showed that a higher number of cells lose X function while a lower number turn on Bcl11b with a delay. In this study, we learn from the in silico simulations that depleting X is required, and thereby is the decision that unleashes the ability of a cell to commit to the T-cell fate. Since the observed commitment typically happens one or two generations after X depletion when a cell opens its Bcl11b-regulatory region and transitions into the DN2b state, we conclude that there is decision inheritance and it plays an important role in T-cell commitment. If inheritance of commitment decision would not play a role, e.g. if the system would be completely dominated by noise, then the result would instead be completely random lineage trees with a mix between open and closed cells in the same branches. Moreover, before X depletion, a cell can still escape from the T-cell lineage fate. In an extended follow-up study, it would be interesting to expand the model to include alternative lineage fates to study what happens to those cells that escape from the T-cell lineage fate.

T-cell development has been extensively studied both through experimental and computational efforts. This makes it an excellent model system for studying lineage commitment in other biological systems. This study showing that inheritance plays an important role in T-cell commitment decisions paves the way for revealing whether inheritance is present and is important in the commitment of other cell types. Our modelling framework shows that the epigenetic regulatory level is an important source of delay between commitment decision-making and experimentally observing commitment. Moreover, it shows that the initiation of decision-making is linked to the stochasticity of the transcriptional programme. Therefore, transcriptional and epigenetic mechanisms are central to inheritance in cell commitment. It should also be noted that there is no built-in irreversibility in the model, i.e. there is nothing that prevents a cell with depleted X from starting to express X again, or that prevents a Bcl11b-open cell from closing Bcl11b again. Nevertheless, we obtained an almost completely irreversible behaviour with very few exceptions.

By introducing the LCA category system, we obtained a common framework which allowed us to compare cells in relation to the commitment event instead of absolute time or generation numbers. This enabled us to study and find the steps of the commitment

mechanisms at population level. Although every cell can be uniquely classified into the introduced LCA categories, it would be possible to define other sets of categories as well, which could be used to study the mechanism from other angles. The concept could also be applied to completely different systems with inheritance.

Our model including the minimal transcriptional network naturally identifies intermediate states between ETP and DN2a with regards to X activity, Bcl11b-epigenetic state and T-cell factor expression. However, our model cannot capture the experimentally observed intermediate states identified in ref. 27, as we only employ the key factors. Moreover, our model shows that cell division is not required for opening Bcl11b, which is in line with data obtained from direct experimental perturbation of the cell cycle in ref. 5. However, cell division acts as a catalyst speeding up the opening¹⁸. Experimental data in ref. 11 suggests that there is a relation between speeding up the cell cycle and turning on Bcl11b. Thus we do capture cell transitions between states without cell division requirements. One might ask if this feature of proliferation aiding commitment also holds for later stages of T-cell development such as beta-selection. This picture was supported in ref. 28, whereas in ref. 29 the two processes appear to be less linked.

In our model simulations, we obtain heterogeneous lineage trees even though we start from homogeneous starting cells across many simulations. It has been shown experimentally though that even the ETP starting cells are actually heterogeneous²⁷. In our model with the minimal GRN, a heterogeneous ETP starting population would only differ with a few counts in gene expression per gene included in the model. Therefore, we anticipate no major differences in simulation outcome since the other stochastic sources downstream in the model would be more prominent.

Stochastic elements are integral parts in all the levels of the multi-scale model and can explain the heterogeneity in the timing of the commitment as well as the fraction of committed cells of the simulated in silico T-cell colonies. Interestingly, the distribution of the X level becomes more narrow when any of the four genes Runx1, Tcf7, Gata3 and PU.1 is knocked down. From a statistical physics point of view, this corresponds to showing that the entropy of the cell is decreased when knocking down a gene. The entropy corresponds to a cell's developmental possibility, where high entropy means that the cell is at a branching point in the developmental path with many options³⁰. Thus, knocking down a gene makes the population more homogeneous which decreases the entropy and closes off potential developmental branches. This directs the cells on a path to commit to the T-cell fate as for PU.1 knockdown or to chose an alternative path as for the other knockdown scenarios.

We developed our framework from in vitro experimental data. However, the model could also be tuned from in vivo data and consequently make in vivo predictions. In Manesso et al. (2013)³¹, a simplified statistical model with both proliferation and commitment components was developed and used to analyse in vivo developmental data³² predicting that commitment increases with number of cell divisions. Our more detailed model here has this feature built in since in divisions the gene expression values and the Bcl11b-epigenetic states are passed down to the daughter cells. Combining our multi-scale agent-based model with the in vivo experimental data^{27,32} for additional parameter tuning would be an interesting future extension of this work.

In summary, our agent-based multi-scale model predicts that inheritance of progression towards T-cell fate plays a significant role in the commitment mechanism. Moreover, the model shows that the commitment takes place over several cell generations. The most striking model prediction is that a cell's decision to commit is actually made one to two generations before observed commitment, i.e. Bcl11b upregulation. Since the T-cell development represents an important model system, these results carry relevance for development in other biological systems.

Methods

The agent-based multi-scale model was implemented from scratch using Python 3.7³³. The code is available at https://github.com/Emil-cbbp/agent-based_multi-scale_model.git. All the tuning of the model was done in our previous study¹¹ and we left the parameter values unchanged. For the independence of this article, we summarise the multi-scale model in the following section.

Model implementation

Level 1: transcriptional level. The GRN in the transcriptional level was modelled using a set of rate equations following the Shea-Ackers formalism³⁴. The GRN includes Runx1, Tcf7, Gata3, PU.1, X and Notch signalling where the concentration levels of these are denoted [R], [T], [G], [P] and [X] respectively. Notch signalling is denoted with N. The rate equations used are

$$\begin{aligned} \frac{\partial[R]}{\partial t} &= \frac{p_1[R]+p_2N}{1+p_1[R]+p_2N} - \gamma_R[R] \\ \frac{\partial[T]}{\partial t} &= \frac{p_3[T]+p_4[G]+\frac{p_5N}{p_6+[P]}}{1+p_3[T]+p_4[G]+p_5N+p_7[P]} - \gamma_T[T] \\ \frac{\partial[G]}{\partial t} &= \frac{p_8[T]+\frac{p_9N}{p_{10}+[P]}}{1+p_8[T]+p_9N+p_{11}[P]} - \gamma_G[G] \\ \frac{\partial[P]}{\partial t} &= \frac{p_{12}[P]}{1+p_{12}[P]+p_{13}[R][G][T]} - \gamma_P[P] \\ \frac{\partial[X]}{\partial t} &= \frac{1}{1+p_{14}[T]+p_{15}[G]} - \gamma_X[X] \\ \frac{\partial N}{\partial t} &= \frac{p_{16}}{1+N}, \end{aligned} \tag{1}$$

with parameter values listed in Table 1. We refer to Olariu et al. (2021)¹¹ for a detailed motivation of the rate equation construction and parameter optimisation. The simulations were conducted using the stochastic Gillespie algorithm²⁵ implemented from scratch.

Level 2: epigenetic level. We used an epigenetic model controlling the Bcl11b regulation adopted from Haerter et al. (2014)¹⁸ and Olariu et al.

Table 1 | Parameter values for GRN in the transcriptional level

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

Table 2 | Parameter values for epigenetic model

Parameter	Value	Probability	Transition	
k_1	0.28	k_1X	$O \rightarrow I$	
		k_1X	$I \rightarrow C$	
k_2	0.20	$k_2N + k_3R$	$C \rightarrow I$	
k_3	0.20	$k_2N + k_3R$	$I \rightarrow O$	
α	0.002	α	$O \rightarrow I$	if mediated by C
β	0.002	β	$I \rightarrow C$	if mediated by C
γ	0.0005	γ	$I \rightarrow O$	if mediated by O
δ	0.0005	δ	$C \rightarrow I$	if mediated by O
ϵ	0.002	ϵ	$I \rightarrow C$	if mediated by I

Parameters k_1 , k_2 and k_3 represent with what rate the input signals from level 1 open or close the regulatory sites. Parameters α - ϵ represent the probability for transitions in need of a mediator to happen. The parameters corresponding reactions are illustrated in Level 2 of Fig. 1b.

(2016)³⁵. The model consists of 500 regulatory sites where each can be in one of three states: closed (C), intermediate (I) or open (O). The Bcl11b region is considered to be open if 75 % of the sites or more are open. The epigenetic level receives input signals from the transcriptional level where Runx1 and Notch signalling work toward opening the regulatory sites and X closes them. The states of the regulatory sites are controlled by a set of rate equations,

$$\begin{aligned} \frac{\partial C}{\partial t} &= (k_2N + k_3[R])I - k_1[X]O + IO - OC \\ \frac{\partial I}{\partial t} &= k_1[X]O + (k_2N + k_3[R])C - (k_1[X] + k_2N + k_3[R])I + OC + CO - IO - IC - II \\ \frac{\partial O}{\partial t} &= k_1[X]I - (k_2N + k_3[R])C + IC + II - CO, \end{aligned} \tag{2}$$

where C, I and O denotes the number of sites in the respective state (see left half of Table 2 for the parameter values¹¹). Not all transitions between states are possible, a site cannot transition directly between O and C without going through I. Some transitions can only take place by the help of a mediator site with a certain state. The right part of Table 2 lists all the possible transitions. The simulations were performed stochastically with an extended version of the Gillespie algorithm²⁵. The first extension is that the number of regulatory sites is constant, so when a site transitions from e.g. $O \rightarrow I$ is added to I and subtracted from O. Next, when a transition has been chosen, a specific regulatory site is chosen at random. If the site's state does not match the chosen transition, nothing happens. If one of the five transitions that require mediation is chosen, an additional mediator site is chosen by random. That site also needs to match in order for the transition to take place.

Level 3: proliferation level. The cell cycle time for each cell is individually drawn from a distribution depending on the generation number of the cell. The lifetimes of the cells get shorter for higher generation number and the distribution get narrower. The division times (T_{div}) is drawn from a Gaussian distribution

$$T_{div} \sim \mathcal{N}(\mu_g, \sigma_g), \tag{3}$$

with mean (μ_g) and standard deviation (σ_g) given per generation in Table 3. These distributions were adapted to experimental data in Olariu et al. (2021)¹¹.

When a cell divides, the gene expression levels are copied from the mother cell to the two daughter cells. The states of the Bcl11b regulatory sites

Table 3 | Parameter values for proliferation model

Generation	μ_g [h]	σ_g [h]
$g = 0$	34	13
$g = 1$	15	5
$g = 2$	13	5
$g = 3$	12	4
$g \geq 4$	12	3

Table 4 | Initial conditions for a simulated T-cell colony

Type	Value
Runx1	1
Tcf7	2
Gata3	1
PU.1	5
X	8
Notch	7
Closed	500
Intermediate	0
Open	0

are passed on differently depending on the expression level of X. If $X \geq 0$, then the regulatory sites are copied from the mother to the daughters. If $X = 0$, the states are passed on by the following rules: all $C \rightarrow I$, $I \rightarrow I$ or O with equal probability of each, and all $O \rightarrow O$. Since our adaptation of the epigenetic regulation model from Haerter et al. (2014) does not consider any spatial relation between the regulatory sites, it is only the number of sites with each state that is passed on at division. At the initialisation of a cell, the regulatory region is created with the correct number of sites of each state and every site is treated individually.

Agent-based model. The agent-based model was realised by implementing each cell as an instance of a class. For each cell object, a division time is generated from the distribution given by Eq. (3) and Table 3 and the cell is assigned initial values of gene expression and Bcl11b regulatory site states as outlined in Section Level 3: proliferation level. The rate equations for the transcriptional level (Eq. (1)) are numerically evolved first, followed by evolving the epigenetic rate equations (Eq. (2)) since the epigenetic level depends on the transcriptional level. Each cell object contains a list of all its ancestors and where all its descendants are recorded, respectively.

A simulation of a cell colony is initialised with one single cell with its initial conditions specified in Table 4. A colony simulation was terminated when all cells in a generation had passed 120 hours in simulated time.

Tree plotting

The lineage tree plots for a simulated colony were produced using the Python package ETE 3³⁶. The initial cell is placed in the root of the tree. For each cell division, the tree branches in two. The radial length of each branch is proportional to the time span between cell division, i.e. the time-axis points radially outwards. The circular node representing a cell is coloured in black, red or white depending on the cell's properties. If the cell has an open Bcl11b regulatory region at the time of division, the node is coloured white. If the Bcl11b regulatory region is closed and the level of X at the time of division is above 0, the node is coloured black. If, instead, the level of X is 0, the node is coloured red.

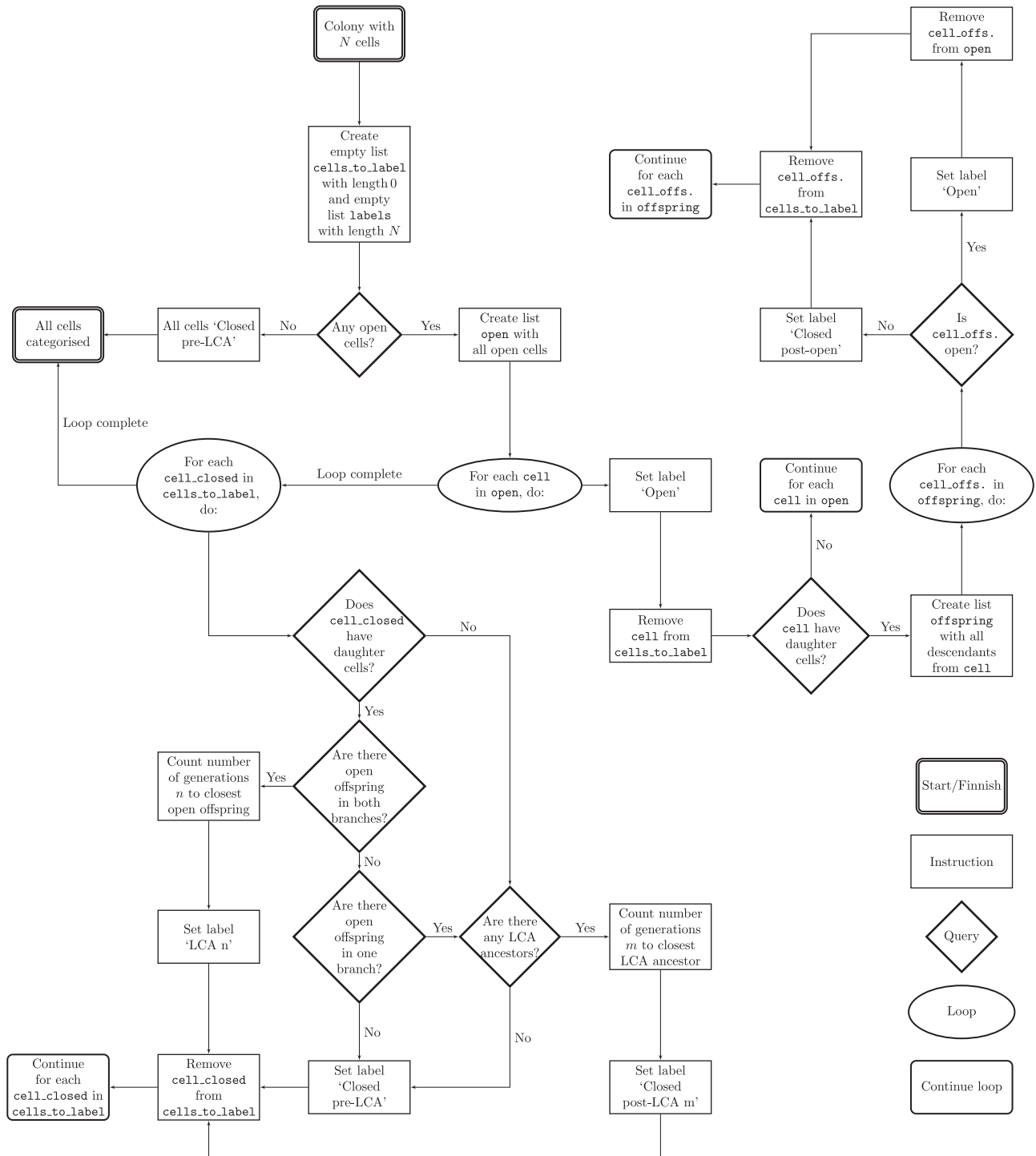


Fig. 6 | Last common ancestor flowchart. Flowchart over the classification process of the last common ancestor (LCA) categories which starts at the top in the centre.

Last Common Ancestor categories

The definitions of the LCA categories are stated in Fig. 2b, c. The precise way of how the classification of each cell in a colony is implemented is summarised in the flowchart shown in Fig. 6.

Knockdown simulations

Knockdown was implemented by multiplying the production rate of a gene by a knockdown factor. We performed two kinds of knockdowns: initial knockdown and delayed knockdown. When initiating knockdown from the

start of the simulation, the initial level of the KD gene was also multiplied by the same knockdown factor. When simulating delayed knockdown, the simulations were carried out the standard way up until the point where the knockdown was initiated, i.e. after 44, 60, 76, 92, or 108 hours in simulated time.

Code availability

The original code is available at https://github.com/Emil-cbbp/agent-based_multi-scale_model.git.

Received: 15 November 2023; Accepted: 8 April 2024;

Published online: 17 April 2024

References

- Yui, M. A. & Rothenberg, E. V. Developmental gene networks: a triathlon on the course to T cell identity. *Nat. Rev. Immunol.* **14**, 529–545 (2014).
- Rothenberg, E. V., Ungerback, J. & Champhekar, A. Forging T-lymphocyte identity: intersecting networks of transcriptional control. *Adv. Immunol.* **129**, 109–174 (2016).
- Del Real, M. M. & Rothenberg, E. V. Architecture of a lymphomyeloid developmental switch controlled by PU. 1, Notch and Gata3. *Development* **140**, 1207–1219 (2013).
- Hosokawa, H. et al. Transcription factor PU.1 represses and activates gene expression in early T cells by redirecting partner transcription factor binding. *Immunity* **48**, 1119–1134 (2018).
- Kueh, H. Y. et al. Asynchronous combinatorial action of four regulatory factors activates Bcl11b for T cell commitment. *Nat. Immunol.* **17**, 956–965 (2016).
- Hosokawa, H. et al. Cell type-specific actions of Bcl11b in early T-lineage and group 2 innate lymphoid cells. *J. Exp. Med.* **217**, e20190972 (2019).
- Ikawa, T. et al. An essential developmental checkpoint for production of the T cell lineage. *Science* **329**, 93–96 (2010).
- Li, L., Leid, M. & Rothenberg, E. V. An early T cell lineage commitment checkpoint dependent on the transcription factor Bcl11b. *Science* **329**, 89–93 (2010).
- Li, P. et al. Reprogramming of T cells to natural killer-like cells upon Bcl11b deletion. *Science* **329**, 85–89 (2010).
- Yui, M. A., Feng, N. & Rothenberg, E. V. Fine-scale staging of T cell lineage commitment in adult mouse thymus. *J. Immunol.* **185**, 284–293 (2010).
- Olariu, V. et al. Multi-scale dynamical modeling of T cell development from an early thymic progenitor state to lineage commitment. *Cell Rep.* **34**, 108622 (2021).
- Weber, B. N. et al. A critical role for TCF-1 in T-lineage specification and differentiation. *Nature* **476**, 63–68 (2011).
- Scripture-Adams, D. D. et al. GATA-3 dose-dependent checkpoints in early T cell commitment. *J. Immunol.* **193**, 3470–3491 (2014).
- Shin, B. & Rothenberg, E. V. Multi-modular structure of the gene regulatory network for specification and commitment of murine T cells. *Front. Immunol.* **14**, 1108368 (2023).
- Shin, B., Zhou, W., Wang, J., Gao, F. & Rothenberg, E. V. Runx factors launch T cell and innate lymphoid programs via direct and gene network-based mechanisms. *Nat. Immunol.* **24**, 1458–1472 (2023).
- Nguyen, P., Pease, N. A. & Kueh, H. Y. Scalable control of developmental timetables by epigenetic switching networks. *J. Royal Soc. Interface* **18**, 20210109 (2021).
- Pease, N. A. et al. Tunable, division-independent control of gene activation timing by a polycomb switch. *Cell Rep.* **34**, <https://doi.org/10.1016/j.celrep.2021.108888> (2021).
- Haerter, J. O., Lövkvist, C., Dodd, I. B. & Sneppen, K. Collaboration between CpG sites is needed for stable somatic inheritance of DNA methylation states. *Nucleic Acids Res.* **42**, 2235–2244 (2014).
- Ng, K. K. et al. A stochastic epigenetic switch controls the dynamics of T-cell lineage commitment. *Elife* **7**, e37851 (2018).
- Zhou, W., Gao, F., Romero-Wolf, M., Jo, S. & Rothenberg, E. V. Single-cell deletion analyses show control of pro-T cell developmental speed and pathways by Tcf7, Spi1, Gata3, Bcl11a, Erg, and Bcl11b. *Sci. Immunol.* **7**, eabm1920 (2022).
- Rosenbauer, F. et al. Lymphoid cell growth and transformation are suppressed by a key regulatory element of the gene encoding PU. 1. *Nat. Genet.* **38**, 27–37 (2006).
- Romero-Wolf, M. et al. Notch2 complements Notch1 to mediate inductive signaling that initiates early T cell development. *J. Cell Biol.* **219**, e202005093 (2020).
- Shin, B. et al. Runx1 and Runx3 drive progenitor to T-lineage transcriptome conversion in mouse T cell commitment via dynamic genomic site switching. *Proc. Natl Acad. Sci.* **118**, e2019655118 (2021).
- Hosokawa, H. et al. Stage-specific action of Runx1 and GATA3 controls silencing of PU.1 expression in mouse pro-T cells. *J. Exp. Med.* **218**, e20202648 (2021).
- Gillespie, D. T. Exact stochastic simulation of coupled chemical reactions. *J. Phys. Chem.* **81**, 2340–2361 (1977).
- Baum, D. et al. Reading a phylogenetic tree: the meaning of monophyletic groups. *Nat. Educ.* **1**, 190 (2008).
- Zhou, W. et al. Single-cell analysis reveals regulatory gene expression dynamics leading to lineage commitment in early T-cell development. *Cell Syst.* **9**, 321–337 (2019).
- Kreslavsky, T. et al. β -Selection-induced proliferation is required for $\alpha\beta$ T cell differentiation. *Immunity* **37**, 840–853 (2012).
- Dose, M. et al. c-Myc mediates pre-TCR-induced proliferation but not developmental progression. *Blood* **108**, 2669–2677 (2006).
- Wiesner, K., Teles, J., Hartnor, M. & Peterson, C. Haematopoietic stem cells: entropic landscapes of differentiation. *Interface Focus* **8**, 20180040 (2018).
- Manesso, E., Chickarmane, V., Kueh, H. Y., Rothenberg, E. V. & Peterson, C. Computational modelling of T-cell formation kinetics: output regulated by initial proliferation-linked deferral of developmental competence. *J. R. Soc. Interface* **10**, 20120774 (2013).
- Porritt, H. E., Gordon, K. & Petrie, H. T. Kinetics of steady-state differentiation and mapping of intrathymic-signaling environments by stem cell transplantation in nonirradiated mice. *J. Exp. Med.* **198**, 957–962 (2003).
- Van Rossum, G. & Drake, F. L. *Python 3 Reference Manual* (CreateSpace, Scotts Valley, CA, 2009).
- Ackers, G. K., Johnson, A. D. & Shea, M. A. Quantitative model for gene regulation by lambda phage repressor. *Proc. Natl Acad. Sci.* **79**, 1129–1133 (1982).
- Olariu, V., Lövkvist, C. & Sneppen, K. Nanog, Oct4 and Tet1 interplay in establishing pluripotency. *Sci. Rep.* **6**, 25438 (2016).
- Huerta-Cepas, J., Serra, F. & Bork, P. ETE 3: reconstruction, analysis, and visualization of phylogenomic data. *Mol. Biol. Evol.* **33**, 1635–1638 (2016).

Acknowledgements

The authors thank Dr. Wen Zhou and Dr. Mary A. Yui for helpful discussions. The authors gratefully acknowledge the support of the United States Public Health Service (NIH) (USPHS grant R01HL119102 to E.V.R. and C.P.). This work was partially supported by the Wallenberg AI, Autonomous Systems and Software Program (WASP) funded by the Knut and Alice Wallenberg Foundation.

Author contributions

E.A., V.O., E.V.R. and C.P. designed the study. E.A. implemented the agent-based multi-scale model and performed all the analysis. E.A. and V.O. wrote the manuscript. All authors provided inputs and comments on the manuscript.

Funding

Open access funding provided by Lund University.

Competing interests

The authors declare no competing interests.

Additional information

Supplementary information The online version contains supplementary material available at <https://doi.org/10.1038/s41540-024-00368-y>.

Correspondence and requests for materials should be addressed to Victor Olariu.

Reprints and permissions information is available at <http://www.nature.com/reprints>

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

© The Author(s) 2024