CycleFlow quantifies cell-cycle heterogeneity *in vivo*

Adrien Jolly[1], Ann-Kathrin Fanti[2], Ines Gräßer[1],
Nils B. Becker[1,*], Thomas Höfer[1,*]

[1]Division or Theoretical Systems Biology

[2]Division of Cellular Immunology

German Cancer Research Center (DKFZ)

Im Neuenheimer Feld 280, 69120 Heidelberg, Germany

*corresponding authors: nils.becker@dkfz.de, t.hoefer@dkfz.de

1

Abstract

While the average cell-cycle length in a cell population can be derived from pulse-chase experiments, proliferative heterogeneity has been difficult to quantify. Here we describe CYCLE-FLOW, a broadly applicable method that applies Bayesian inference to combined measurements of EdU incorporation and DNA content. CYCLEFLOW accurately quantifies the fraction of proliferating versus quiescent cells and the durations of cell-cycle phases of the proliferating cells *in vitro* and *in vivo*.

Main

Owing to advances in live-cell microscopy, we are gaining deep understanding of how eukaryotic cells transition between quiescence and proliferation and control the length of their division cycle[1–10]. These studies rely on direct observation of cell cycles *in vitro*. Similar quantification of cell cycles and quiescence during development, tissue renewal and tumor growth in the intact organism has not been possible by imaging, and existing proliferation assays *in vivo* only partially address the key questions. In particular, the rule of thumb that high S and G_2/M fractions indicate rapid cell proliferation (as measured by DNA staining or single-cell RNA-sequencing (scRNA-Seq) data) can mislead, as we will see below. To quantify cell proliferation, pulse-chase experiments with thymidine analogs or other labels have been widely employed[11] and, recently, improved by using sophisticated double-labeling techniques[12,13]. However, the quiescent fraction is not resolved in these experiments, which also confounds the estimation of cell-cycle duration. Quiescent cells are often distinguished from cycling cells by (negative) staining for Ki67, but the protein continues to be expressed after cells have stopped cycling, complicating the interpretation of the results[14]. An alternative assay for quiescence, long-term retention of a diluting label, also has limitations, as initially quiescent cells will not be labeled when using thymidine analogs, and dilution of fluorescent histone variants is difficult to distinguish from degradation[15]. Also, there is no generic gene-expression signature for quiescence that could be detected in scRNA-Seq data. Thus, quiescence has largely remained a qualitative rather than quantitative concept despite its key role in stem cell differentiation and therapy resistance.

Here we describe CYCLEFLOW, a time-efficient method that jointly quantifies the quiescent fraction in a cell population and the duration of the cell-cycle phases in the proliferating cells, both *in vitro* and *in vivo*. CYCLEFLOW combines two readily applicable experimental techniques — pulse-chase with a thymidine analog (e.g., 5-ethynyl-2'-deoxyuridine, EdU) and total DNA staining [16] — with Bayesian parameter inference. The key idea is to label a cohort of S phase cells with a brief pulse of tymidine analog and then use the DNA content as a coordinate to follow the labeled cells through the cell cycle, eventually observing re-entry of divided

cells into the cell cycle or transition into a quiescent state (Fig. 1a). We model label dynamics during cell-cycle progression mathematically and, confronting the model with a time series of flow-cytometry snapshots, infer the fractions of proliferating versus quiescent (G_0) cells and the time the former spend in the G_1, S and G_2/M phases of the cell cycle.

We first applied CYCLEFLOW to the MYC-driven proliferation of a cancer cell line, TET21N, in culture. This system provides a ground truth for evaluating the performance of CYCLE-FLOW, as cell-cycle duration and absence of quiescence have been established by time-lapse imaging[10]. We grew TET21N cells in the exponential phase, added EdU to the culture medium and washed after one hour; cells were collected at six time points over the course of 18 hours, and analyzed by flow cytometry for DNA and EdU content (Fig. 1b). As expected for an exponentially growing population, the proportions of the cell-cycle phases (Fig. 1c) stayed approximately constant; for brevity, we denote the corresponding gates by $\overline{G01}$ (2n DNA content), $\overline{S}$ (DNA content intermediate between 2n and 4n) and $\overline{G2M}$ (4n DNA content). Three hours after EdU application, all cells in $\overline{S}$ were EdU-positive (Fig. 1d). At 5 hours, cells in the initial stages of $\overline{S}$ were not EdU-labeled, implying that EdU was no longer available for *de novo* labeling. This allowed us to follow the progression of the cohort of initially EdU-labeled cells around the cell cycle; at 5 hours, these cells were predominantly in late $\overline{S}$ and $\overline{G2M}$. The labeled cohort returned to $\overline{G01}$ (14 hours), remaining clearly distinguishable from EdU-negative cells. From $\overline{G01}$, labeled cells re-entered into the cell cycle and progressively appeared again in $\overline{S}$ (14 and 18 hours). Thus, EdU pulse labeling allows following a cohort of cells over one cycle.

However, there are several difficulties in interpreting these data in a straightforward manner. First, the cells within the EdU-labeled cohort do not transit synchronously to G_2, which could be due to variable starting points in S phase at the time of labeling or cell-to-cell variations in the speed of progress. Second, cells within $\overline{G01}$ may be either in G_0 or G_1. Third, the precise time course of EdU availability is not measured. Finally, the assignment of cell-cycle phases based on DNA content has limited resolution: Note that, at 3 hours, some cells at the right-hand edge of $\overline{G01}$ were EdU-positive, implying that these had already entered S phase but were not identified as such by DNA staining (Fig. 1d). Likewise, it is possible that some labeled cells in $\overline{G2M}$ were still in late S phase. To address these difficulties, CYCLEFLOW bases the interpretation of the experimental data on a mathematical model of cell-cycle progression and EdU labeling that takes into account all relevant features (Fig. 1e,f; Supplementary Table 1; Online Methods). Specifically, the model describes the progression of a cell through the cycle as a Markov process, dividing each cycle phase (G_1, S and G_2/M) into subphases in order to account for phase length variability[8] (Fig. 1e). Upon division, cells may continue to cycle or enter quiescence (G_0). Release from G_0 may occur but can be neglected on the short timescale of the experiment. Cells may also be lost from any cycle phase or G_0, e.g., due to cell death or differentiation. As long as EdU is present in the medium (as estimated from the data), cells in S phase can incorporate it and become labeled (Fig. 1f). Finally, the model ex-

3

ploits the joint information on DNA content and EdU labeling available for each cell, rather than DNA-content-based gates only, to assign cells to the cycle phases.

Using Bayesian inference, we estimated the model parameters from the data by Monte-Carlo sampling from the posterior distribution. The model accurately fitted the experimental data within measurement error (Fig. 1g,h), inferring a cycle length of 18.4 (16.7, 21.1) hours (posterior median and (5%, 95%) quantiles) and phase lengths for G_1, S and G_2/M of 6.5 (5.5, 8) hours, 6.8 (5.5, 9) hours and 5 (4.3, 5.6) hours, respectively (Fig. 1i). Using direct observation by time-lapse microscopy of TET21N cells expressing the FUCCI Cdt1 degron[10], reliable values were obtained for total cycle and G_1 phase lengths, and these matched the results of CYCLEFLOW (Fig. 1j). Moreover, CYCLEFLOW inferred the fraction of G_0 cells to be negligible (Fig. 1k), as observed[10]. Taken together, these findings validate CYCLEFLOW.

Next, we applied CYCLEFLOW to a developmental process *in vivo*, where the extent of quiescence has been controversial—the differentiation of T cells in the thymus. We focused on the CD4+ CD8+ double-positive (DP) thymocytes, the development of which is accompanied by a transition from proliferation to quiescence[17]. Varying estimates of the quiescent fraction (30-90%) have been put forward but direct quantification has been lacking[18]. To address this question, we injected a single intra-peritoneal dose of EdU into young adult mice and performed EdU and DNA staining at consecutive time points (Fig. 2a). Gating by DNA content showed >90% of DP thymocyte in $\overline{G01}$ (Fig. 2b). We observed rapid termination of EdU labeling (Fig. 2c, 3 hours) and, as before, return of the labeled cohort to $\overline{G01}$ (Fig. 2c, 3 and 5 hours) followed by re-entry into $\overline{S}$ and $\overline{G2M}$ (Fig. 2c, 14 hours). Of note, a large fraction of cells remained unlabeled. All these features were captured by the mathematical model, which reproduced both the steady-state cell-cycle distribution (Fig. 2d) and the dynamics of labeled cells (Fig. 2e).

Interestingly, although the fraction of labeled cells in $\overline{S}$ and $\overline{G2M}$ plateaued at low values already after a few hours, the labeled fraction in $\overline{G01}$ kept increasing for the duration of the experiment (18 hrs), raising the question of which process fuels the continuing label accumulation. $\overline{G01}$ contains cells in G_1 and G_0; the model separates these contributions and shows that, upon exit from mitosis, nearly equal fractions of cells enter quiescence and continue cycling (Fig. 2f). While the fraction of labeled cycling cells quickly plateaus (similar to the TET21N cells, Supplementary Fig. 1), the labeled quiescent cells continue to accumulate. Thus, in DP thymocytes proliferation precedes quiescence, which is consistent with the known developmental progression[18].

Comparing the experimental data and the inferred cell-cycle parameters for TET21N cells and DP thymocytes illustrates how the simple rule of thumb of judging proliferation rate by S and G_2/M fractions can fail. The cell-cycle phase distributions show much higher $\overline{S}$ and $\overline{G2M}$ fractions for TET21N cells (compare Figs 1c and 2b), which would indicate a higher proliferation rate in this population compared to the DP thymocytes. However, the proliferating DP thymocytes actually cycle faster than TET21N cells, but this is masked by the high frac-
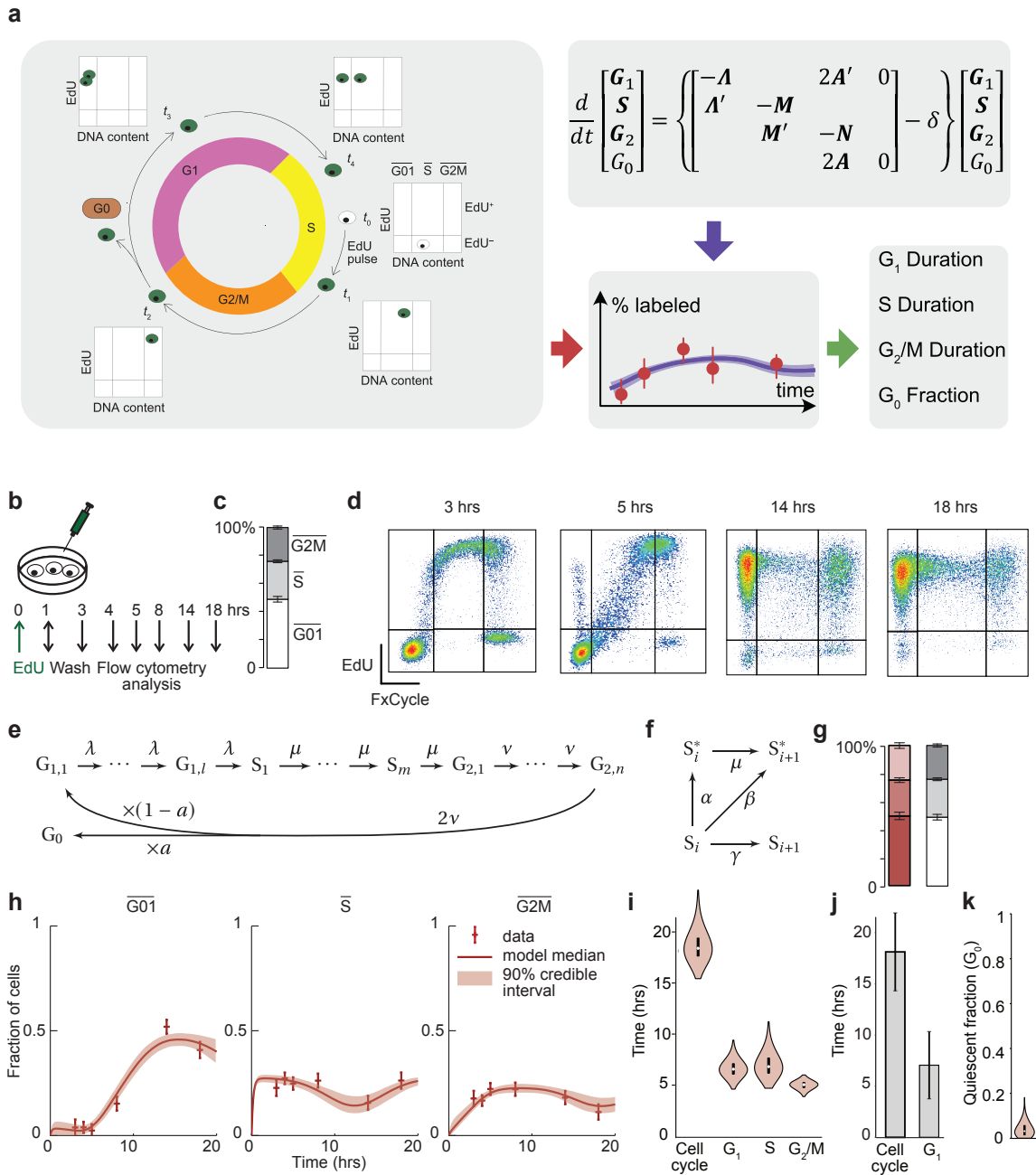
Fig. 1 | Outline of CycleFlow and application to cancer cell proliferation *in vitro*. a, Schematic of CY-CLEFLOW: The progression of EdU labeled cells through the cell cycle is tracked over time, the mathematical model is fitted to the data, and cell-cycle parameters are estimated via Bayesian inference. **b**, EdU pulse-chase experiment for cells in culture. **c**, Distribution of cells in the $\overline{G01}$, $\overline{S}$, and $\overline{G2M}$ gates as determined by DNA content for TET21N (averaged over all time points). Error bars indicate standard error of the mean (SEM); n = 25. **d**, Progression of EdU labeled TET21N cells through the cell cycle, as defined by DNA content. Four representative flow-cytometry snapshots are shown; mean fractions of cell subpopulations (with SEM) over all experimental repeats at the corresponding times are indicated in **h**. **e**, Schematic of mathematical model for cell-cycle progression; cell-cycle phases are divided into sub-phases. Parameters: λ, μ and ν, progression rates through subphases of G_1, S and G_2/M, respectively, with l, k and n, denoting subphase numbers; a, probability of cell cycle arrest. **f**, Mathematical Model of EdU incorporation during cell progression in S phase. $\alpha + \beta$, total EdU incorporation rate; $\beta + \gamma = \mu$, progression rate. **g**, Model fit (left diagram) versus experimental data (right diagram, same as **c**) for the distribution of cells in the $\overline{G01}$, $\overline{S}$, and $\overline{G2M}$ gates. Error bars indicate 90% credible intervals (left diagram). **h**, Time courses of EdU-labeled TET21N cells in $\overline{G01}$, $\overline{S}$, and $\overline{G2M}$ gates compared between experimental data (error bars, pooled SEM; n = 3 to 6 per time point) and model fit. Population sizes are given as fractions of total cells. **i**, Duration of total cell cycle, G_1, S and G_2/M phases of TET21N cells inferred by CYCLEFLOW. White dots, median values; black bars, interquartile ranges. **j**, Mean duration of cell-cycle and G_1 phase of TET21N, expressing the Cdt1 FUCCI degron, measured with time lapse microscopy (data taken from Ref.[10]). **k**, Quiescent fraction of TET21N cells inferred by CYCLEFLOW.

tion of quiescent thymocytes. CYCLEFLOW allows us to see beyond population averages and disentangle cell-cycle heterogeneity by inferring the underlying cell-cycle phase distribution (Fig. 2g), rather than observing gated fractions. Specifically, TET21N cells proliferate uniformly (quiescent fraction 3 (0, 9) %), with 18.4 (16.7, 21.1) hours average cycle length (cf. Fig. 1i,k). By contrast, the large $\overline{\text{G01}}$ fraction in DP thymocytes consists mainly of quiescent cells, which constitute 85 (84, 86) % of all cells (Fig. 2h), corroborating the existence of a large quiescent fraction[18]. The remaining 15 (14, 16) % of cells cycle extremely rapidly, with a cycle length of 10 (9.4, 10.9) hours (Fig. 2i). We note that, whereas Chao et al. (Ref 8) found that cycle length of different cell types was primarily set by the duration of G_1, all phases appear to be shortened in DP thymocytes (Fig. 2i), with a particularly swift passage through G_2/M. This finding indicates that proliferation rate can be regulated at all stages of the cell cycle[19].

To conclude, the main advance of CYCLEFLOW is that it disentangles quiescent (G_0) from cycling (G_1) cells with $2n$ DNA content. In turn, this enables determination of the true cell-cycle phase durations of the proliferating cells. In so doing, our mathematical framework overcomes salient problems of previous approaches, including the imperfect overlap between cell-cycle phases and DNA content gates and the gradual removal of EdU. CYCLEFLOW is equally applicable for steady-state (e.g., thymocytes) and expanding (e.g., TET21N cells) cells, both *in vivo* and *in vitro*. Application to other cell types is straightforward: the measurements should extend over roughly one cell cycle with several (typically 4-8) loosely distributed time points.
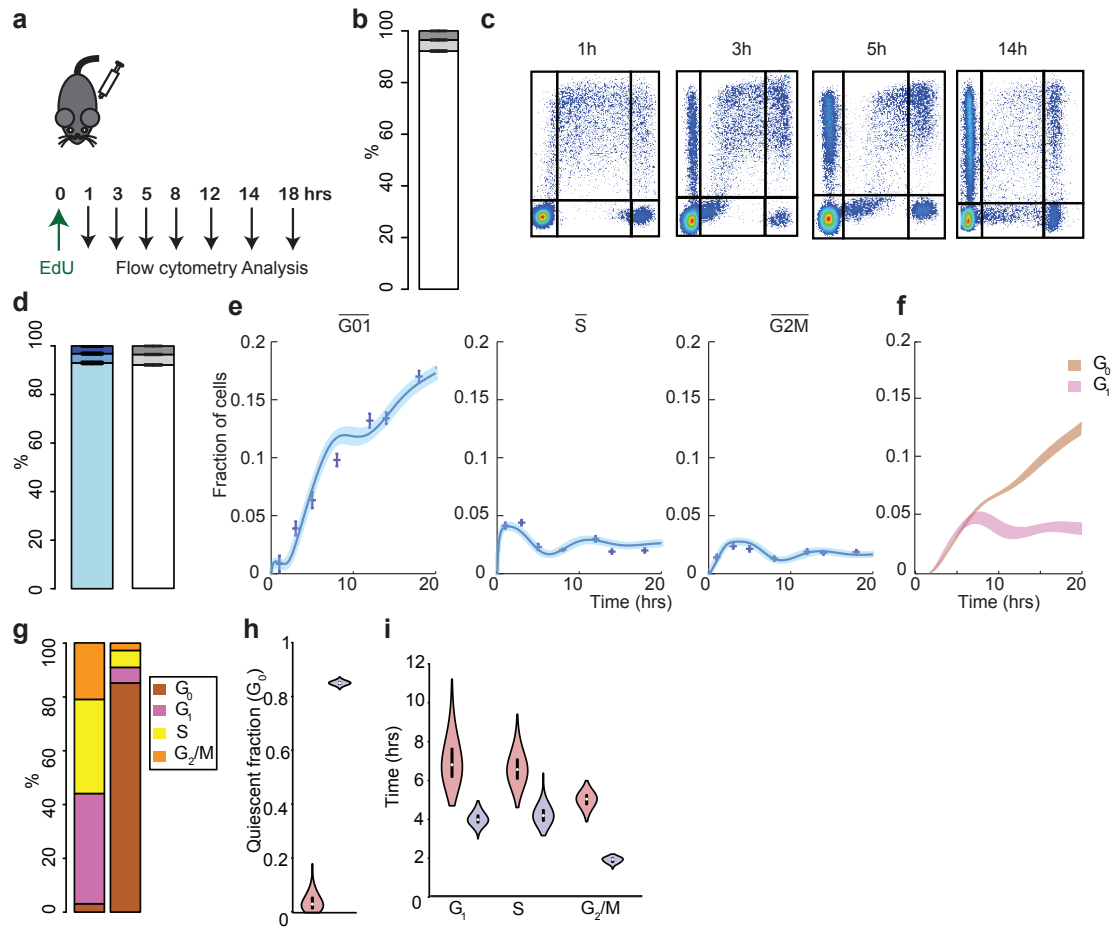
Fig. 2 | Application of CycleFlow to determine quiescence and proliferation rate of DP thymocytes *in vivo*. a, Schematic of the EdU pulse-chase experiment. **b**, Distribution of cells in the $\overline{G01}$, $\overline{S}$, and $\overline{G2M}$ gates as determined by DNA content for DP (averaged over all time points). Errors bars indicate standard error of the mean (SEM); n = 31. **c**, Progression of EdU-labeled DP thymocytes through the cell cycle, as defined by DNA content. Four representative flow-cytometry snapshots are shown; mean fractions of cell subpopulations (with SEM) over all experimental repeats at the corresponding times are indicated in **e**. **d**, Model fit of the distribution of cells in the $\overline{G01}$, $\overline{S}$, and $\overline{G2M}$ gates, error bars indicate 90% credible interval (left bar) and corresponding data (right bar; same as **b**). **e**, Time courses of EdU-labeled DP cells in $\overline{G01}$, $\overline{S}$, and $\overline{G2M}$ gates compared between experimental data (error bars, pooled SEM; n = 3 to 6 per time point) and model fit. Population sizes are given as fractions of total cells. **f**, Inferred fractions of EdU-labeled G_0 and G_1 cells in the $\overline{G01}$ gate. **g**, Inferred distribution of cell-cycle phases including G_0 (median values). Left bar, TET21N cells; right bar, DP thymocytes. **h**, Inferred fraction of quiescent cells in TET21N cells (left) and DP thymocytes (right). White dots, median values; black bars, interquartile ranges. **i**, Duration of G_1, S and G_2/M phases of DP thymocytes (blue) compared to TET21N cells (red) inferred by CycleFlow. White dots, median values; black bars, interquartile ranges.

7

Methods

Experimental

Labeling and analysis of TET21N cells *in vitro*

SH-EP TET21/N (TET21/N) cells were grown in RPMI 1640 medium supplemented with 10% Fetal Calf Serum (FCS) at 37°C, 5% CO_2 and 88% humidity. For each sample, 1.5×10^6 cells were seeded on 15 cm Petri dishes one day before EdU treatment and then treated with EdU (Invitrogen) at a final concentration of 10 μM in the culture medium. Cells were then harvested and fixed in 4% paraformaldehyde solution in PBS and kept in 90% methanol 10% PBS solution at −20°C. For flow cytometry analysis, cells were washed in washing buffer (1% Bovine Serum Albumin(BSA), 0.1% TritonX in PBS) and resuspended in PBS supplemented with 1% BSA. Cells were then permeabilized and the Click-it reaction was performed using the Click-iT Plus EdU Alexa Fluor 488 flow cytometry Kit (Invitrogen) according to the manufacturer's protocol. For total DNA staining, cells were resuspended in FxCycle™ Violet (Thermofisher Scientific) solution (1:1000 in Click-it Permeabilization buffer) prior to flow cytometry measurement. Data were acquired on a MACSQuant VYB (Miltenyi Biotec) and cell populations were analyzed with FlowJo 10.

Mice

C57BL/6J mice were used, both female and male. Mice were kept in individually ventilated cages under specific pathogen-free conditions in the animal facility at the German Cancer Research Center (DKFZ, Heidelberg). All animal experiments were performed in accordance with institutional and governmental regulations, and were approved by the Regierungspräsidium (Karlsruhe, Germany).

Labeling of thymocytes *in vivo* and analysis

Mice of ages between eleven and sixteen weeks were injected intraperitoneally with 1 mg EdU (Invitrogen) diluted in sterile PBS. At several time points after injection mice were sacrificed and thymi harvested. Thymi were mashed in a 40 μm filter with the plunger of a syringe. To identify dead cells, cells were incubated in Zombie Red™ Fixable Viability dye (Biolegend) solution (1:1000 in PBS). Fc receptors were blocked by incubating cells in PBS supplemented with 5% FCS with 250 μ g.ml^{-1} purified mouse IgG (Jackson ImmunoResearch Laboratories). Antibody stainings were performed in PBS/5% FCS on ice for 30 minutes with optimal dilutions of commercially-prepared antibodies. The following antibodies were used: CD8a allophycocyanin (APC) (53-6.7), TCR gd BV421 (GL3) from eBioscience ; CD4 phycoerythrin (PE) (H129.19), Ter119 BV421 (Ter119) from BD Pharmingen; CD11b BV421 (M1/70), CD19 BV421 (6D5), NK1.1 BV421 (PK136), Gr-1 BV421 (RB6-8C5) from Biolegend. The lineage cocktail was composed of CD11b, CD19, Ter119, NK1.1, Gr-1 and TCR gd. After antibody staining, cells were

8

fixed, permeabilized and the Click-it reaction was performed using the Click-iT Plus EdU Alexa Fluor 488 flow cytometry kit (Invitrogen). For total DNA staining, cells were resuspended in FxCycle™ Violet solution (1:1000 in Click-it Permeabilization buffer) prior to flow cytometry measurement. Data was acquired on a BD LSRFortessa™ cell analyzer (Becton Dickinson) and cell populations were analyzed with FlowJo 10. Double-positive thymocytes were defined as lineage$^-$ CD4$^+$CD8$^+$.

Mathematical inference

Model for cell-cycle progression

Within a growing cell population, each cell progresses through the cycle phases G$_1$, S and G$_2$ (Fig. 1e). Each phase requires a stochastic time to complete. We capture this variability by dividing the cycle phases into subphases: G$_{1,1}$ through G$_{1,l}$, S$_1$ through S$_m$ and G$_{2,1}$ through G$_{2,n}$, respectively. Each subphase corresponds to an approximate progress within the respective cycle phase but has no further biological meaning. We then model the transitions from one subphase to the next (and from the last subphase of a cycle phase to the first subphase of the next) as rate processes with rates λ, μ or ν, for G$_1$, S and G$_2$, respectively. In this model, the times to complete each of the cycle phases are Erlang-distributed random variables, with means $\tau_{G_1} = l/\lambda$, $\tau_S = m/\mu$, and $\tau_{G_2} = n/\nu$, and coefficients of variation (CVs), $1/\sqrt{l}$, $1/\sqrt{m}$ and $1/\sqrt{n}$, respectively. At the end of the cycle, a cell divides to produce two daughters. A daughter cell may then, with probability a, exit the cell cycle into the arrested state G$_0$, or, with probability $1 - a$, restart the cycle in phase G$_{1,}$. We consider entry into G$_0$ to be irreversible on the time scale of the experiment, and therefore exclude any transition from G$_0$ back into the cycle. Finally, cells may be lost from any cycle phase or G$_0$ with rate δ, due to cell death or differentiation.

The dynamics of subpopulations is then described by a linear system of ordinary differential equations. Written in matrix form,

$$\frac{d}{dt}\boldsymbol{N} = \mathbf{T}\boldsymbol{N} - \delta\boldsymbol{N}, \tag{1}$$

where the vector of cell numbers in the various cycle subphases is defined as

$$\boldsymbol{N} = [\mathrm{G}_{1,1}(t), \ldots, \mathrm{G}_{1,l}(t), \mathrm{S}_1(t), \ldots, \mathrm{S}_m(t), \mathrm{G}_{2,1}(t), \ldots, \mathrm{G}_{2,n}(t), \mathrm{G}_0(t)]^\mathsf{T}, \tag{2}$$

9

and the rate matrix

$$
\mathbf{T} = \begin{bmatrix}
-\lambda & & & & & & & & & 2(1-a)\nu & 0 \\
\lambda & \ddots & & & & & & & & & \vdots \\
& \ddots & -\lambda & & & & & & & & \\
& & \lambda & -\mu & & & & & & & \\
& & & \mu & \ddots & & & & & & \\
& & & & \ddots & -\mu & & & & & \\
& & & & & \mu & -\nu & & & & \\
& & & & & & \nu & & & & \\
& & & & & & & \ddots & \ddots & & \\
& & & & & & & & \nu & -\nu & \vdots \\
& & & & & & & & & 2a\nu & 0
\end{bmatrix} . \tag{3}
$$

In a cell population deep in the exponential growth phase, the total cell number $N(t)$ increases exponentially, while the proportions of cells in the various subphases are constant; we call this regime *steady growth*. At steady growth,

$$
\boldsymbol{N}(t) = N(t)\boldsymbol{n} = N(0)e^{(\kappa-\delta)t} \times [\mathrm{g}_{1,1}, \ldots, \mathrm{g}_{1,l}, \mathrm{s}_1, \ldots, \mathrm{s}_m, \mathrm{g}_{2,1}, \ldots, \mathrm{g}_{2,n}, \mathrm{g}_0]. \tag{4}
$$

Here, we have denoted the population growth rate by $(\kappa - \delta)$. The proportions of cells in the subphases are given by the steady-growth distribution $\boldsymbol{n}$; denoting the constant vector of ones by $\mathbf{1} = [1, \ldots, 1]^{\mathsf{T}}$, the normalization of $\boldsymbol{n}$ reads $\mathbf{1}^{\mathsf{T}}\boldsymbol{n} = 1$. Eq. 1 then yields

$$
\tfrac{d}{dt}\boldsymbol{N}(t) = N(t)(\kappa - \delta) \times \boldsymbol{n} = N(t)[\mathbf{T}\boldsymbol{n} - \delta\boldsymbol{n}], \text{ which implies } \kappa\boldsymbol{n} = \mathbf{T}\boldsymbol{n}. \tag{5}
$$

Thus κ is the dominant eigenvalue of the matrix $\mathbf{T}$, and $\boldsymbol{n}$ is the associated normalized eigenvector. The eigenvalue κ can be interpreted as the population growth rate in the absence of loss.

Steady-growth distribution

To find $\boldsymbol{n}$, we start by normalizing Eq. 5:

$$
\kappa = \mathbf{1}^{\mathsf{T}}\mathbf{T}\boldsymbol{n} = [0, \ldots, 0, \nu, 0]^{\mathsf{T}}\boldsymbol{n} = \nu\mathrm{g}_{2,n}. \tag{6a}
$$

Eq. 6a implies that the total rate of production of new cells κN equals the population in the final subphase $\mathrm{G}_{2,n}$ times its rate of completion. Now consider steady growth with production of new cells, where $\mathrm{G}_{2,n}, \kappa > 0$. Starting from Eq. 6a and backsubstituting using Eq. 5, we

10

compute the remainder of the cycle subphase distribution in steady growth as a function of κ:

$$g_0 = g_{2,n} 2a\nu/\kappa = 2a \tag{6b}$$

$$g_{2,n-j} = g_{2,n} \frac{(\kappa + \nu)^j}{\nu^j} = \frac{\kappa}{\nu}\left(\frac{\kappa + \nu}{\nu}\right)^j \qquad 0 < j < n \tag{6c}$$

$$s_m = g_{2,1}\frac{\kappa + \nu}{\mu} = \frac{\kappa}{\mu}\left(\frac{\kappa + \nu}{\nu}\right)^n \tag{6d}$$

$$s_{m-j} = s_m \frac{(\kappa + \mu)^j}{\mu^j} = \frac{\kappa}{\mu}\left(\frac{\kappa + \nu}{\nu}\right)^n\left(\frac{\kappa + \mu}{\mu}\right)^j \qquad 0 < j < m \tag{6e}$$

$$g_{1,l} = s_1\frac{\kappa + \mu}{\lambda} = \frac{\kappa}{\lambda}\left(\frac{\kappa + \nu}{\nu}\right)^n\left(\frac{\kappa + \mu}{\mu}\right)^m \tag{6f}$$

$$g_{1,l-j} = g_{1,l}\frac{(\kappa + \lambda)^j}{\lambda^j} = \frac{\kappa}{\lambda}\left(\frac{\kappa + \nu}{\nu}\right)^n\left(\frac{\kappa + \mu}{\mu}\right)^m\left(\frac{\kappa + \lambda}{\lambda}\right)^j \qquad 0 < j < l \tag{6g}$$

To complete the calculation, the dominant eigenvalue κ can be found numerically by solving Eq. 5 for given transition rates and subphase numbers. For steady growth to be possible, $a \leq 1/2$ is required, since otherwise each cell has less than one proliferating daughter cell on average.

Steady-state distribution

To capture cell populations that are maintained at a fixed size within the model Eq. 1, we consider the loss rate δ to be subject to an implicit homeostatic regulation. This regulation maintains $\delta \equiv \kappa$, so that proliferation and loss balance, and no net growth occurs. In this way, homeostasis is treated as a marginal case of steady growth, Eq. 4 with vanishing effective growth rate, so that the subphase populations $N = N\boldsymbol{n}$ are constant in time. In particular, we can compute the steady-state distribution $\boldsymbol{n}$ by Eqs. 4 with the replacement $\kappa \rightarrow \delta$ everywhere. If desired, one also obtains the homeostatic set point of δ, by solving the eigenvalue problem Eq. 5, with $\kappa \rightarrow \delta$.

Kinetics of labeling

In order to relate the cycling model Eq. 1 to experimental data, we include the dynamics of label incorporation and inheritance. As long as EdU label is available, unlabeled cells in S phase incorporate it, thereby transitioning into a labeled state. The labeled states (denoted S^*, $G^*_{2,k}$ etc.) are defined operationally: Those cells that are gated above the background fluorescence level in the EdU fluorescence channel are considered labeled. In practice we then find that transition to the labeled state requires only a small part of S phase, in other words, the rate of label acquisition $\epsilon > \mu/m$. However it's unclear if label acquisition is also faster than progress from one S subphases to the next, $\epsilon \gtrsim \mu$. Therefore, we consider S phase progress and labeling to be parallel processes, see Fig. 1f. In this scheme, the three new rates α, β and γ are

11

constrained by

$$\epsilon = \alpha + \beta \tag{7}$$

setting the total rate of labeling to be ϵ, and

$$\mu = \beta + \gamma, \tag{8}$$

which states that the total progress rate of unlabeled cells through S phase is undisturbed by label incorporation. We impose a third constraint for convenience,

$$\alpha/\beta = \beta/\gamma, \tag{9}$$

which then uniquely determines the three rates for given ϵ:

$$\alpha = \frac{\epsilon^2}{\epsilon + \mu}; \quad \beta = \frac{\epsilon\mu}{\epsilon + \mu}; \quad \gamma = \frac{\mu^2}{\epsilon + \mu}. \tag{10}$$

Since the chase phase of the experiment remains shorter than two cell cycles, we do not observe nor include in the model any delabeling by dilution of EdU.

Labeled population dynamics

Finally, we collect the previous expressions into a system of linear ODEs for the dynamics of all, labeled and unlabeled subpopulations. The vector of average labeled cell numbers is

$$\boldsymbol{N}^* = [\mathrm{G}_{1,1}^*(t), \ldots, \mathrm{G}_{2,n}^*(t), \mathrm{G}_0^*(t)]^\mathsf{T};$$

unlabeled cell numbers are denoted by $\boldsymbol{N}^\circ$. They add up to the total cell numbers $\boldsymbol{N} = \boldsymbol{N}^\circ + \boldsymbol{N}^*$.

We require that steady growth or steady state has been attained before labeling begins. Thus up to the label application at $t = 0$, $\boldsymbol{N}^\circ = \boldsymbol{N}$ follows Eq. 4, with steady-growth distribution $\boldsymbol{n}$ satisfying Eq. 5. Labeling transfers cells into the labeled populations, so that at times $t > 0$, Eq. 1 holds for $\boldsymbol{N}$ but not for $\boldsymbol{N}^\circ$ nor $\boldsymbol{N}^*$.

To describe the dynamics of the labeled population $\boldsymbol{N}^*$, we encode the labeling kinetics in matrix form:

$$(\mathbf{L})_{i,j} = \begin{cases} \alpha\delta_{i,j} + \beta\delta_{i,j+1} & \text{if } l < j \le l + m \\ 0 & \text{otherwise} \end{cases}. \tag{11}$$

The labeling matrix $\mathbf{L}$ has the same dimensions as $\mathbf{T}$; its nonzero entries are α on the diagonal for all subphases of S, and β below any α. Here, α and β are given by Eqs. 10. It can be verified straightforwardly that $\mathbf{L}\boldsymbol{N}^\circ$ is the flux from unlabeled to labeled compartments as shown in

12

Fig. 1f. Using this notation,

$$\tfrac{d}{dt}\boldsymbol{N}^* = (\mathbf{T} - \delta)\boldsymbol{N}^* + \mathbf{L}\boldsymbol{N}^{\circ} \tag{12a}$$

$$= (\mathbf{T} - \delta)\boldsymbol{N}^* + \mathbf{L}(\boldsymbol{N} - \boldsymbol{N}^*)$$

$$= (\mathbf{T} - \mathbf{L} - \delta)\boldsymbol{N}^* + N(t)\mathbf{L}\boldsymbol{n}. \tag{12b}$$

In Eq. 12a, we have used the fact that already-labeled cells follow the undisturbed cycle progression $(\mathbf{T} - \delta)$; in Eq. 12b, we have inserted the undisturbed steady-growth expansion of the total population. Finally, we rewrite this system in terms of the dynamics of the (time-dependent, non-normalized) labeled fractions, $\boldsymbol{n}^* = [g^*_{1,1}, \ldots, g^*_0] = \boldsymbol{N}^*/N$. From Eq. 12b and using exponential growth of $N(t)$, we obtain

$$\tfrac{d}{dt}\boldsymbol{n}^* = (\mathbf{T} - \mathbf{L} - \kappa)\boldsymbol{n}^* + \mathbf{L}\boldsymbol{n}. \tag{13}$$

Importantly, in experiments EdU supply is stopped at the end of an initial labeling pulse, and thereafter, the availability of free EdU decreases gradually as it is consumed by cells or otherwise degraded. We model this decrease by an exponential decrease of the total labeling rate

$$\epsilon = \epsilon(t) = \epsilon_0 e^{-t/\tau_E}, \tag{14}$$

where τ_E is an adjustable parameter. In this way α, β (Eq. 10) and therefore $\mathbf{L}$ (Eq. 11) become time-dependent.

Parameter estimation

Experimentally, cycle phase is assigned by DNA content. Thus, cells in G_0 or G_1 fall into a single experimental gate, $\overline{\text{G01}}$. Although most DNA-replicating cells are correctly counted in the corresponding $\overline{\text{S}}$ gate, the gating procedure inevitably assigns some cells that have started DNA replication to the $\overline{\text{G01}}$ gate, and some cells that have completed replication to the $\overline{\text{S}}$ gate rather than the $\overline{\text{G2M}}$ gate. When deriving model predictions, we account for this crosstalk by assigning the first $m_{\overline{01}}$ subphases of S phase to $\overline{\text{G01}}$, and the last $m_{\overline{2}}$ subphases to $\overline{\text{G2M}}$, so that only the innermost $m - m_{\overline{01}} - m_{\overline{2}}$ subphases S_i are assigned to the gate $\overline{\text{S}}$, as shown in Eq. 15.

$$\underbrace{G_0;\ \ G_{1,1} \cdots G_{1,l}\ S_1 \cdots S_{m_{\overline{01}}} }_{\overline{\text{G01}}} S_{m_{\overline{01}}+1} \cdots \underbrace{S_{m-m_{\overline{2}}}}_{\overline{\text{S}}}\ \underbrace{S_{m-m_{\overline{2}}+1}\ \cdots\ S_m G_{2,1} \cdots G_{2,n}}_{\overline{\text{G2M}}} \tag{15}$$

In order to compute model predictions, we first solve Eq. 5 for the steady-growth distribution $\boldsymbol{n}$. From $\boldsymbol{n}$ we obtain the time-independent total cell fractions in each experimental gate,

e.g.

$$\overline{g2}_{\text{pred}} = \sum_{i=m-m_{\overline{2}}+1}^{m} s_i + \sum_{i=1}^{n} g_{2,i}.$$

We then solve system Eq. 13 numerically, for the initial condition $\boldsymbol{n}^*(0) = \boldsymbol{0}$. From $\boldsymbol{n}^*(t)$, we obtain predictions for the time-dependent labeled fractions, *e.g.*

$$\overline{g01}^*_{\text{pred}}(t) = g_0^*(t) + \sum_{i=1}^{l} g_{1,i}^*(t) + \sum_{i=1}^{m_{\overline{01}}} s_i^*(t).$$

In order to then compare model predictions with experiments, we evaluate the negative log-likelihood function in a standard way, as

$$\chi^2 = \frac{1}{2} \sum_{\overline{g}=\overline{g01},\overline{g2}} \frac{\left(\overline{g}_{\text{pred}} - \overline{g}_{\text{obs}}\right)^2}{\sigma_{\overline{g}}^2} + \frac{1}{2} \sum_{i=1...I} \sum_{\overline{g}^*=\overline{g01}^*,\overline{s}^*,\overline{g2}^*} \frac{\left(\overline{g}^*_{\text{pred}}(t_i) - \overline{g}^*_{\text{obs}}(t_i)\right)^2}{\sigma_{\overline{g}^*}^2}. \tag{16}$$

Here, the first sum collects time independent terms; the experimental observations $\overline{g}_{\text{obs}}$ are averages over all time-points and experimental repeats, and the associated uncertainties $\sigma_{\overline{g}}$ are standard errors of these averages. Only two gates are included, since normalization fixes the third. The second sum runs over the I experimental time points and is not constrained by normalization. The corresponding uncertainties are estimated as standard errors over experimental repeats of the mean labeled gate fractions; for robustness, we pool these errors and assign the same $\sigma_{\overline{g}^*}$ for all time points.

Based on the log-likelihood Eq. 16, we then perform Markov Chain Monte Carlo sampling to evaluate the posterior distribution over model parameters. The full set of model parameters and the allowed ranges of their uniform prior distributions are given in Supplementary Table 2. The credible intervals resulting from sampled posterior distributions for DP and TET2lN are shown in Supplementary Table 3.

Implementation and Software

The model was simulated using the Python programming language with the package Scipy v.0.13.0 and Monte Carlo Markov Chain (MCMC) sampling was performed with the package Emcee v.3.02 with default settings. In accordance with the program's documentation, a chain was deemed to have converged when the autocorrelation times for every parameters exceeded 50 times the length of the chain. Results of the MCMC sampling were then processed with Matlab R2013B (MathWorks) and R to produce graphical representations.

Acknowledgments

Contributions

T.H., A.J. and N.B. conceived the Study; N.B. and T.H. supervised the project; A.J and A-K.S. conceived and conducted the experiments on mice; A.J and I.G. conceived and conducted the experiments on TET21N; A.J. and N.B. developed the mathematical model; A.J. implemented the model and A.J. and N.B. analyzed the data; T.H., N.B. and A.J wrote the manuscript and all the coauthors contributed to the manuscript.

Competing Interests

The authors declare no competing interests.

Data availability

All raw Data and codes are available upon reasonable request to the authors.

References

1. Sakaue-Sawano, A. *et al.* Visualizing spatiotemporal dynamics of multicellular cell-cycle progression. *Cell* **132**, 487–498 (2008).

2. Duffy, K. R. *et al.* Activation-induced B cell fates are selected by intracellular stochastic competition. *Science* **335**, 338–341 (2012).

3. Spencer, S. L. *et al.* The proliferation-quiescence decision is controlled by a bifurcation in CDK2 activity at mitotic exit. *Cell* **155**, 369 – 383 (2013).

4. Barr, A. R., Heldt, F. S., Zhang, T., Bakal, C. & Novák, B. A dynamical framework for the all-or-none G1/S transition. *Cell systems* **2**, 27–37 (2016).

5. Barr, A. R. *et al.* DNA damage during s-phase mediates the proliferation-quiescence decision in the subsequent G1 via p21 expression. *Nature Communications* **8**, 14728 (2017).

6. Ryl, T. *et al.* Cell-cycle position of single MYC-driven cancer cells dictates their susceptibility to a chemotherapeutic drug. *Cell Systems* **5**, 237 – 250.e8 (2017).

7. Yang, H. W., Chung, M., Kudo, T. & Meyer, T. Competing memories of mitogen and p53 signalling control cell-cycle entry. *Nature* **549**, 404–408 (2017).

8. Chao, H. X. *et al.* Evidence that the human cell cycle is a series of uncoupled, memoryless phases. *Molecular systems biology* **15**, e8604 (2019).

9. Loeffler, D. *et al.* Asymmetric lysosome inheritance predicts activation of haematopoietic stem cells. *Nature* **573**, 426–429 (2019).

10. Kuchen, E. E., Becker, N. B., Claudino, N. & Höfer, T. Hidden long-range memories of growth and cycle speed correlate cell cycles in lineage trees. *eLife* **9**, e51002 (2020).

11. De Boer, R. J. & Perelson, A. S. Quantifying T lymphocyte turnover. *J. Theor. Biol.* **327**, 45–87 (2013).

12. Akinduro, O. *et al.* Proliferation dynamics of acute myeloid leukaemia and haematopoietic progenitors competing for bone marrow space. *Nat Commun* **9**, 519 (2018).

13. Kretschmer, L. *et al.* Differential expansion of T central memory precursor and effector subsets is regulated by division speed. *Nat Commun* **11**, 113 (2020).

14. Miller, I. *et al.* Ki67 is a graded rather than a binary marker of proliferation versus quiescence. *Cell Rep* **24**, 1105–1112 (2018).

15. Morcos, M. N. F. *et al.* Continuous mitotic activity of primitive hematopoietic stem cells in adult mice. *J. Exp. Med.* **217** (2020).

16. Dolbeare, F., Gratzner, H., Pallavicini, M. G. & Gray, J. W. Flow cytometric measurement of total DNA content and incorporated bromodeoxyuridine. *Proc. Natl. Acad. Sci. U.S.A.* **80**, 5573–5577 (1983).

17. Mingueneau, M. *et al.* The transcriptional landscape of alpha beta T cell differentiation. *Nat. Immunol.* **14**, 619–632 (2013).

18. Yates, A. J. Theories and quantification of thymic selection. *Front Immunol* **5**, 13 (2014).

19. Gitlin, A. D. *et al.* T cell help controls the speed of the cell cycle in germinal center B cells. *Science* **349**, 643–646 (2015).