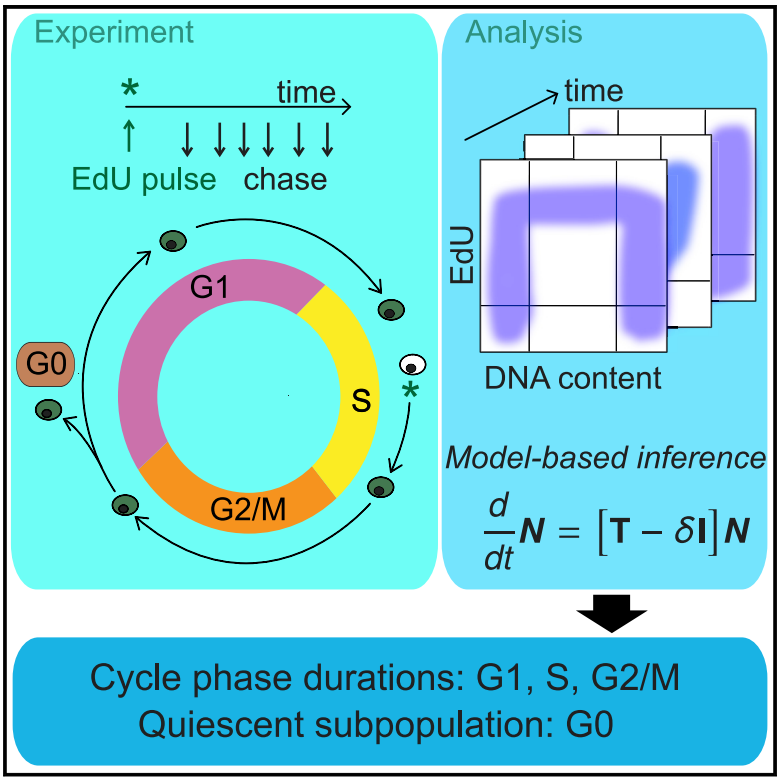
Cell Reports
Methods

Report

CycleFlow simultaneously quantifies cell-cycle phase lengths and quiescence *in vivo*

Graphical abstract



Highlights

- CycleFlow quantifies quiescence and cell cycling in heterogeneous populations *in vivo*
- CycleFlow combines pulse chase with a thymidine analog with mathematical inference
- Accounting for experimental uncertainty yields robust estimates of cell-cycle parameters
- Application to T cell development quantifies thymocyte quiescence

Authors

Adrien Jolly, Ann-Kathrin Fanti, Csilla Kongsaysak-Lengyel, Nina Claudino, Ines Gräßer, Nils B. Becker, Thomas Höfer

Correspondence

nils.becker@dkfz.de (N.B.B.), t.hoefer@dkfz.de (T.H.)

In brief

The switching of cells between cycling and quiescence is a fundamental process in development, tissue homeostasis, and immune responses. Jolly et al. describe a broadly applicable method that combines thymidine-analog labeling with model-based inference to disentangle and quantify quiescence and cycle timing.

Jolly et al., 2022, Cell Reports Methods 2, 100315
October 24, 2022
https://doi.org/10.1016/j.crmeth.2022.100315

CellPress

Cell Reports Methods

CellPress
OPEN ACCESS

Report

CycleFlow simultaneously quantifies cell-cycle phase lengths and quiescence *in vivo*

Adrien Jolly,[1] Ann-Kathrin Fanti,[2] Csilla Kongsaysak-Lengyel,[2] Nina Claudino,[1] Ines Gräßer,[1] Nils B. Becker,[1,*] and Thomas Höfer[1,3,*]

[1]Division of Theoretical Systems Biology, German Cancer Research Center (DKFZ), 69120 Heidelberg, Germany
[2]Division of Cellular Immunology, German Cancer Research Center (DKFZ), 69120 Heidelberg, Germany
[3]Lead contact
*Correspondence: nils.becker@dkfz.de (N.B.B.), t.hoefer@dkfz.de (T.H.)
https://doi.org/10.1016/j.crmeth.2022.100315

MOTIVATION Assaying cell cycling with thymidine analogs *in vivo* is a standard method. However, the cell populations of interest are often heterogeneous, consisting of subpopulations of cycling and quiescent cells. Hence, the correct determination of the cell-cycle duration requires the simultaneous assessment of the quiescent fraction. Here, we show that a straightforward extension of the standard pulse-chase protocol, with a single thymidine analog, allows accurate determination of the quiescent cell fraction and of the cell-cycle phase durations in the proliferating subpopulation. Our method, CycleFlow, relies on measuring at several time points during the chase and interpreting the data with a realistic mathematical model of the cell cycle. CycleFlow yields robust estimates of the cell-cycle characteristics in the face of typical sources of experimental uncertainty.

SUMMARY

Populations of stem, progenitor, or cancer cells show proliferative heterogeneity *in vivo*, comprising proliferating and quiescent cells. Consistent quantification of the quiescent subpopulation and progression of the proliferating cells through the individual phases of the cell cycle has not been achieved. Here, we describe CycleFlow, a method that robustly infers this comprehensive information from standard pulse-chase experiments with thymidine analogs. Inference is based on a mathematical model of the cell cycle, with realistic waiting time distributions for the G_1, S, and G_2/M phases and a long-term quiescent G_0 state. We validate CycleFlow with an exponentially growing cancer cell line *in vitro*. Applying it to T cell progenitors in steady state *in vivo*, we uncover strong proliferative heterogeneity, with a minority of $CD4^+CD8^+$ T cell progenitors cycling very rapidly and then entering quiescence. CycleFlow is suitable as a routine method for quantitative cell-cycle analysis.

INTRODUCTION

The decision of cells to progress through the cell cycle or enter quiescence is fundamental for multicellular organisms. This decision is under strict control during development and tissue regeneration. Long-term quiescence of stem and progenitor cells is thought to be enabled by signals from niche or stroma cells or intrinsic developmental programs in the intact organism (van Velthoven and Rando, 2019), and it becomes deregulated in cancer (Aguirre-Ghiso, 2007; Li and Clevers, 2010). The proliferating fraction sustains cell production, and the proliferation rate can be regulated at various stages of the cell cycle, including the G_1 and S phases (Akinduro et al., 2018; Kretschmer et al., 2020; Gitlin et al., 2015). Hence, characterizing the proliferative activity of a cell population requires knowledge of the durations of the cell-cycle phases and the quiescent fraction. However, simultaneous quantification of long-term quiescence and proliferative activity *in vivo* remains challenging, thus also limiting mechanistic insight.

A major obstacle stems from the fact that inference of quiescence and cycle progression are interdependent. For instance, when estimating proliferative activity based on DNA content, the often-applied rule of thumb that low S and G_2/M fractions indicate slow cell cycles can be misleading: the low fractions may result from slow proliferation in a homogeneous population (i.e., a long G_1 phase) but also from a substantial subpopulation of quiescent cells (large G_0 fraction), or both, which cannot be distinguished based on DNA content. This example demonstrates that in order to obtain accurate cycle times, quiescent versus proliferating subpopulations need to be quantified.

Hitherto, pulse-chase techniques for quantifying cell cycling and methods that estimate the quiescent fraction have been

applied separately. This is problematic, as pulse-chase techniques yield non-representative averages of cell-cycle time when neglecting quiescence. To estimate the quiescent fraction, proxy markers of cycling cells (Ki67, phosphorylated retinoblastoma protein [pRb]) have been used; how to combine such data with pulse-chase data in a statistically consistent manner is an open problem. Moreover, Ki67 continues to be expressed for some time after cells have stopped cycling (Miller et al., 2018), and pRB is absent from both transiently arrested (Spencer et al., 2013) and fully quiescent cells. Another approach, the long-term dilution of cell labels (thymidine analogs or fluorescently tagged histones), has been used in experiments demonstrating hematopoietic stem cell quiescence (Wilson et al., 2008; Foudi et al., 2009), but the quantitative interpretation of these data is not straightforward (Morcos et al., 2020).

The proliferative activity of cells and, in fact, the above-mentioned confounding effects are dynamic in nature. Hence, an adequate mathematical model of cell-cycle dynamics is needed to interpret the experimental data. Mathematical analysis has been applied to estimate cell-cycle progression from pulse-chase experiments, usually by estimating an overall proliferation rate (De Boer and Perelson, 2013; Lahoz-Beneytez et al., 2016) or by using sophisticated double-labeling techniques to estimate the duration of specific cell-cycle phases (Akinduro et al., 2018; Kretschmer et al., 2020). Thus far, these approaches have not considered quiescence. Typically, they also use exponential distributions for the lengths of the total cell cycle or its phases, which recent data show to be rather of Erlang type (Chao et al., 2019).

Here, we describe CycleFlow, an easy-to-use method that simultaneously quantifies proliferation and long-term quiescence in heterogeneous cell populations. Cells arrested transiently as a result of DNA damage (Spencer et al., 2013; Barr et al., 2016; Ryl et al., 2017; Yang et al., 2017; Arora et al., 2017) are considered proliferative, whereas the quiescent fraction comprises cells remaining out of the cell cycle for a longer period of time, at least several cycle durations. CycleFlow combines pulse chase of both a thymidine analog (e.g., 5-ethynyl-2′-deoxyuridine [EdU]) and a DNA stain, with Bayesian parameter inference. The key idea is to label a cohort of S phase cells with a brief pulse of a thymidine analog and then use DNA content as a coordinate to follow the labeled cells through the cell cycle, observing re-entry of divided cells into the cycle or transition into a quiescent state (Figure 1A). The experimental data are interpreted within a generic mathematical model of cell-cycle progression in heterogeneous populations, allowing us to infer simultaneously the fractions of quiescent (G_0) versus proliferating cells and the times that the latter spend in the G_1, S, and G_2/M phases of the cell cycle.

RESULTS

Following a cell cohort labeled in S phase through the cycle

We applied CycleFlow in two standard settings: exponential population growth and population turnover at steady state. For exponential growth, we used a MYCN-driven cancer cell line (TET21N) in culture (Figure 1B, left panel). We had previously quantified cell-cycle parameters of these cells including cycle time and G_1 phase duration by time-lapse microscopy (Kuchen et al., 2020), which provided a "ground truth" to evaluate the results of CycleFlow. For steady-state turnover, we used developing T cells in the thymus *in vivo* (Figure 1B, right panel), a paradigm for studying cell differentiation in vertebrates (Yui and Rothenberg, 2014). We focused on the major thymocyte fraction consisting of cells at the $CD4^+CD8^+$ double-positive (DP) stage, which transition from proliferation to quiescence after successful recombination of the *TCRA* gene (Mingueneau et al., 2013). Quiescent DP thymocytes then further differentiate into single-positive CD4 and CD8 T cells. Although DPs are a key differentiation stage, it is not known how rapidly they cycle before becoming quiescent and how large the quiescent DP fraction is, with estimates in the literature ranging from 30% to 90% (Yates, 2014).

The proportions of cell-cycle phases, as determined by DNA content, were very different between the two populations; DP cells contained a much larger fraction with 2*n* DNA content than TET21N cells (Figure 1C). Importantly, 2*n* cells may be quiescent, which we refer to as G_0, or in the G_1 phase of the cell cycle; we denote the 2*n* flow cytometry gate by $\overline{\text{G01}}$. Naively, this observation might suggest that DP cells cycle more slowly than TET21N cells. However, this conclusion would be wrong if only a small fraction of DP cells were cycling, and most were in G_0.

We thus aimed to distinguish cells in G_0 from cells in G_1 within $\overline{\text{G01}}$ and thereby also determine the phase durations of the cycling fraction. To this end, we pulse labeled both cell types with EdU: for TET21N, we added EdU to the culture medium and washed after 1 h, and for DP, we injected a single EdU dose intraperitoneally into young adult mice (Figure 1D). In both cases, a cohort of EdU-labeled cells progressed from $\overline{\text{S}}$ to $\overline{\text{G2M}}$ and, after mitosis, back to $\overline{\text{G01}}$. Already after a few hours, unlabeled cells entered $\overline{\text{S}}$, while labeled cells in $\overline{\text{G01}}$ had not yet re-entered into a second cycle, producing a gap in the EdU-positive gate in early S phase (Figure 1E, 5 h time point for TET21N, 3 and 5 h time points for DP). At later time points, however, labeled cells started to re-enter into a second cycle, closing this gap again (Figure 1E, 14 and 18 h time points for TET21N, and 14 h time point for DP). Thus, EdU pulse labeling allows following a cohort of cells over one cycle.

CycleFlow models cycling and quiescent cell fractions

Notably, in DP, but not in TET21N, a large fraction of cells remained unlabeled (Figure 1E, final time points), which could mean either that these cells were quiescent, or that they were cycling but not in S phase during the brief EdU pulse. To resolve this ambiguity, we designed the mathematical model underlying CycleFlow to infer the true quiescent fraction along with the cell-cycle phase durations of the cycling cells. Following Chao et al. (2019), we described the length of the cell-cycle phases by suitable Erlang distributions (Figure S1A; STAR Methods). We allowed cycling cells to transition irreversibly to G_0 after each cycle. This irreversibility reflects the long-term nature of quiescence (in an extended model, we also considered exit from quiescence; Figure S1C). To

CellPress
OPEN ACCESS

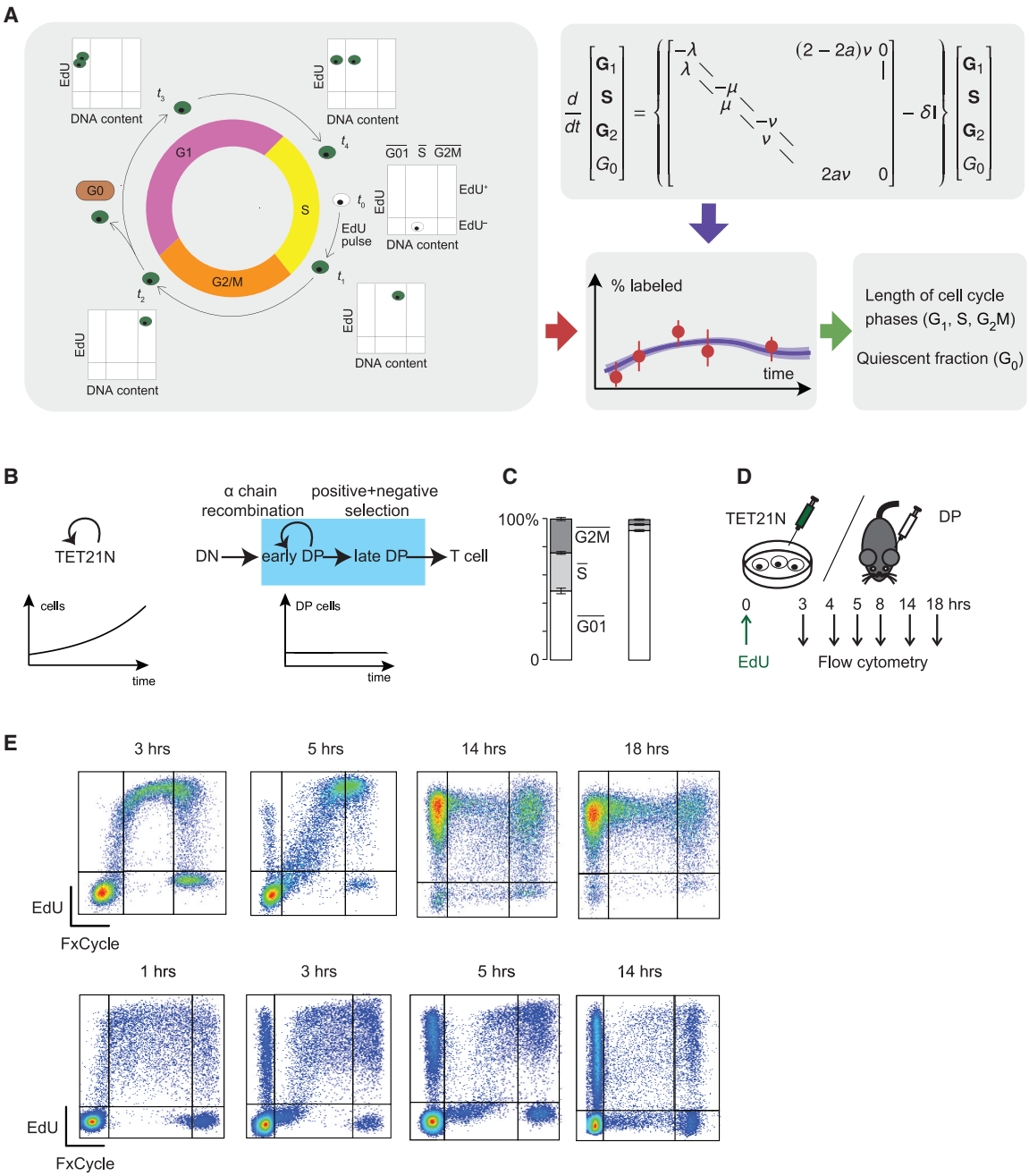


Figure 1. Outline of CycleFlow

(A) Schematic of CycleFlow: 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) Schematic of TET21N cells' exponential growth in culture (left) and T cell differentiation in the thymus (right). The number of TET21N cells grow exponentially, while the number of cells in DP remains constant.

(C) Distribution of cells in the $\overline{G01}$, $\overline{S}$, and $\overline{G2M}$ gates as determined by DNA content averaged over all time points. For TET21N (left barplot, n = 25); for DP thymocytes (right barplot, n = 31). Error bars indicate SEM.

(D) EdU pulse-chase experiment for cells in culture or mouse.

(E) Progression of EdU-labeled TET21N (top row) and DP thymocytes (bottom row) through the cell cycle, as defined by DNA content. In each case, four representative flow cytometry snapshots are shown.

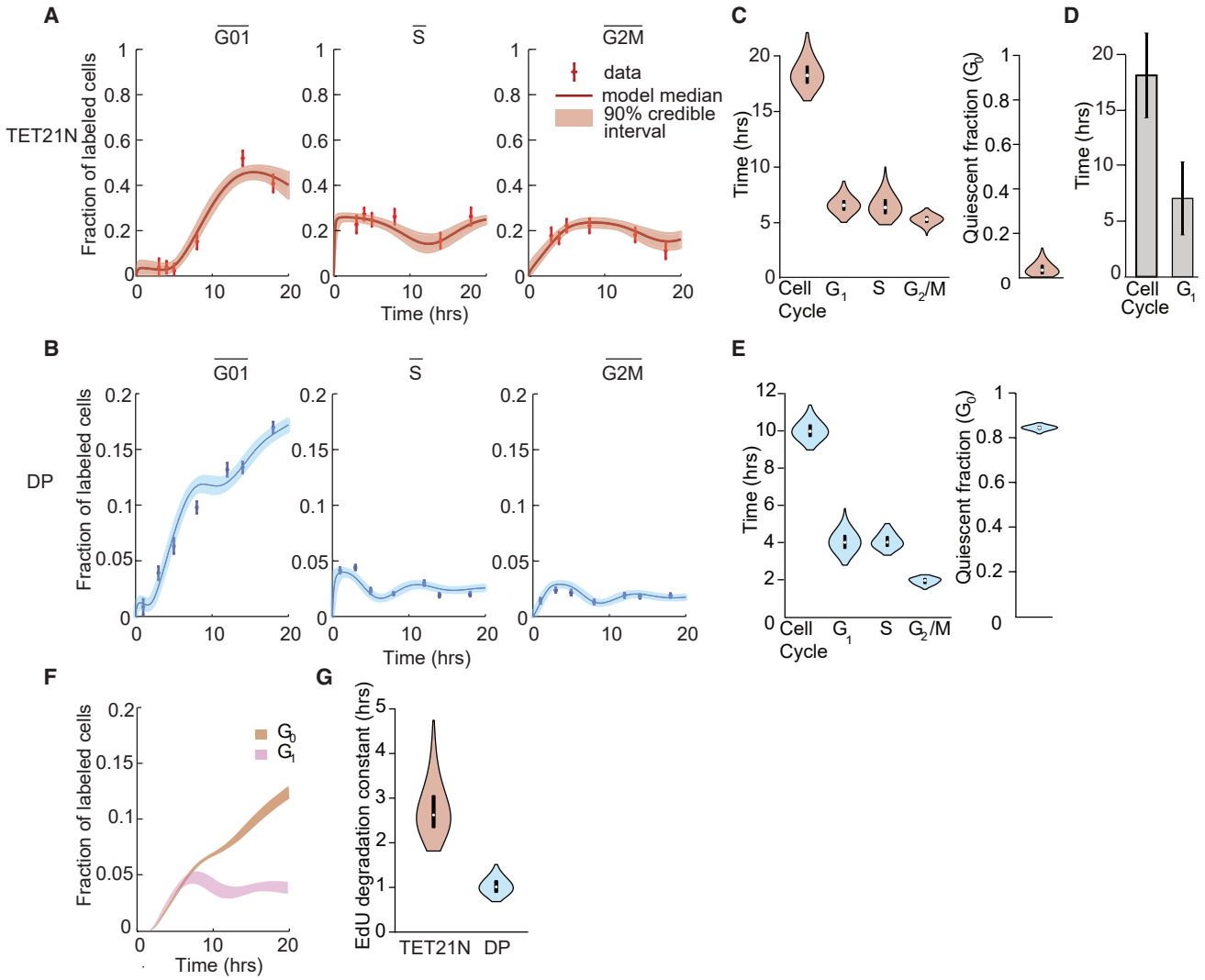


Figure 2. Application of CycleFlow to determine quiescence and proliferation rate of TET21N cells and DP thymocytes *in vivo*

(A and B) Time courses of EdU-labeled TET21N cells (A) and DP thymocytes (B) 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.

(C) Duration of total cell cycle, G_1, S, and G_2/M phases of TET21N cells (left panel), and quiescent fraction (right panel) inferred by CycleFlow. Violin plot indicates parameter distributions; white dots, median values; black bars, interquartile ranges.

(D) Mean duration of cell-cycle and G_1 phase of TET21N, expressing the Cdt1 FUCCI degron, measured with time-lapse microscopy (data taken from Kuchen et al. [2020]). Error bars indicate standard deviation of the mean.

(E) Duration of G_1, S, and G_2/M phases (left panel) and fraction of quiescent cells (right panel) of DP thymocytes inferred by CycleFlow. Violin plot indicates parameter distributions; white dots, median values; black bars, interquartile ranges.

(F) Inferred time course of EdU-labeled G_0 and G_1 DP thymocytes in the $\overline{G01}$ gate.

(G) Samples of the posterior distribution of the EdU degradation time τ_E for TET21N cells (red) and DP thymocytes (blue). Violin plot; white dots, median values; black bars, interquartile ranges.

describe EdU labeling, we exploited the fact that the Erlang distribution can be modeled as a sequence of memoryless steps, allowing cells to acquire label at variable points in S phase (Figure S1B; STAR Methods). As the precise duration of EdU availability cannot be directly measured, we estimate it as a free parameter from the experimental data. We also observed that the assignment of cell-cycle phases based on DNA content has limited resolution. For instance, at 3 h, some TET21N 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 (Figure 1E); likewise, it is possible that some labeled cells in $\overline{G2M}$ were still in late S phase. Hence, CycleFlow also estimates the degree of misassignment of flow cytometry gates to cell-cycle phases from the data (see STAR Methods). After solving this model for the time course of the labeled population in the various gates, information on (1) the quiescent cell fraction, (2) the phase lengths for the cycling cells, and (3) the EdU labeling kinetics

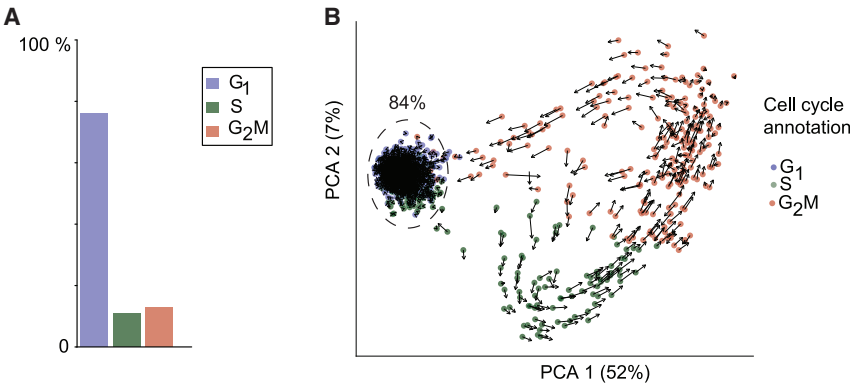


Figure 3. Single-cell transcriptomic analysis of cycle progression

(A) Distribution of cell-cycle phases in DP thymocytes scRNA-seq dataset as estimated by the Seurat package; n =1.

(B) Principal-component analysis of DP scRNA-seq data performed on cell-cycle genes defined by the Seurat package; percentages on each axis denote the variability explained by the principal component. Cells are colored according to cell-cycle phase predicted by Seurat. Arrows indicate RNA velocities projected on the first and second principal components.

can be extracted from the experimental data within a Bayesian setting by sampling from the posterior distribution.

Accurate inference for both growing and steady-state populations

We fitted the model to both datasets, obtaining good agreement between model and data for all cell-cycle phases (Figures 2A, 2B, and S2). The model then also recovered the dynamics of the EdU-negative cell fractions, which were not used for fitting (Figure S3); this result implies that the fractions of cell-cycle phases remained constant throughout each experiment. To evaluate the accuracy of CycleFlow, we first considered the cell-cycle parameters inferred for the TET21N cells. We obtained approximately equal durations of G_1, S, and G_2/M and a very small quiescent fraction (Figure 2C). In particular, most labeled cells accumulating in the $\overline{\text{G01}}$ gate were identified as G_1 (Figure S4), implying that this cancer cell line proliferates homogeneously. Direct observation of TET21N cells expressing the FUCCI Cdt1 degron by time-lapse microscopy (Kuchen et al., 2020) indeed shows practically no quiescent cells. Moreover, the directly measured total cell-cycle duration and G_1 phase duration agrees with the inferred parameters (Figure 2D). Hence, CycleFlow yields correct estimates of the average duration of cell-cycle phases in a cell population containing a negligible fraction of quiescent cells.

Next, we focused on DP thymocytes. Their much lower $\overline{\text{S}}$ and $\overline{\text{G2M}}$ fractions (Figure 1C) could be mistaken to indicate that DPs proliferate at a lower rate compared with the TET21N cells. However, CycleFlow inferred that the proliferating DP thymocytes actually cycle faster than TET21N cells, but this fact is masked by the high fraction of quiescent thymocytes in G_0 (Figure 2E); these constitute 84% (83%, 86%) (median and 90% credible interval) of DP cells. The remaining 16% of cells divide extremely rapidly, with a cycle length of 10 (9.3, 10.9) h. The transition of DP thymocytes into quiescence is seen by comparing the kinetics of labeled cells with TET21N: the key difference to TET21N cells is that in DP thymocytes, the labeled $\overline{\text{G01}}$ fraction increases during the entire experimental time course (Figure 2A). This continued increase is inferred to be due to labeled cells entering quiescence after division instead of progressing to S phase. In this way, CycleFlow disentangles G_0 from G_1 subpopulations (Figure 2F) and reveals a pronounced proliferative heterogeneity in DPs. This finding is consistent with early DP cells first proliferating rapidly and soon becoming quiescent (Yates, 2014); quiescent DPs then constitute the vast majority of the heterogeneous DP compartment.

CycleFlow also infers the duration of EdU availability for labeling. For TET21N cells, which were washed after 1 h, the inferred EdU loss times correspond to a half-life of 1.8 (1.4, 2.8) h (Table S3; Figure 2G). For DP thymocytes *in vivo*, the EdU degradation half-life was 0.7 (0.5, 0.9) h, consistent with known pharmacokinetics of EdU in murine blood (Cheraghali et al., 1995).

Validation of thymocyte quiescence using single-cell transcriptomes

To further validate CycleFlow, we sought independent support for the inferred quiescent DP fraction. We began by performing single-cell RNA sequencing (RNA-seq) of freshly isolated DP cells. We assigned a cycle phase to each cell based on cycle-related transcripts (Stuart et al., 2019), which could not, however, disentangle G_0 and G_1. This procedure overestimated S and G_2 fractions compared with the $\overline{\text{S}}$ and $\overline{\text{G2M}}$ gates defined by DNA content and, hence, underestimated $\overline{\text{G01}}$ (Figure 3A; cf. Figure 1C). The modest agreement of transcriptome-based phase assignment with measurement of DNA content is not surprising, as the gradual transcriptome changes cannot precisely pinpoint the transitions between cell-cycle phases (Hsiao et al., 2020). Recent work using proliferating cell lines *in vitro* has revealed a circular topology of transcriptomes due to cyclic expression (Schwabe et al., 2020). Using principal-component analysis and calculating RNA velocity based on a set of cell-cycle-associated genes for the DP thymocytes (see STAR Methods), we observed a circular flow, which, in contrast to the cell line data (Schwabe et al., 2020), entered a highly populated "sink" region with no apparent efflux (Figure 3B). Given that DP cells exit the cell cycle and enter quiescence, the sink region is likely made up of the quiescent DP cells. This conclusion is in line with these cells being annotated as G_1 by their transcriptome. These observations and the quantitative agreement we found between the fractions of quiescent cells estimated by CycleFlow and cells in the sink region (both 84%) support the validity of CycleFlow in cell populations that contain a quiescent subpopulation.

Distinguishing quiescent cells from transiently arrested cells

Finally, we studied the question whether CycleFlow can effectively distinguish transient cell-cycle arrest, which depends on

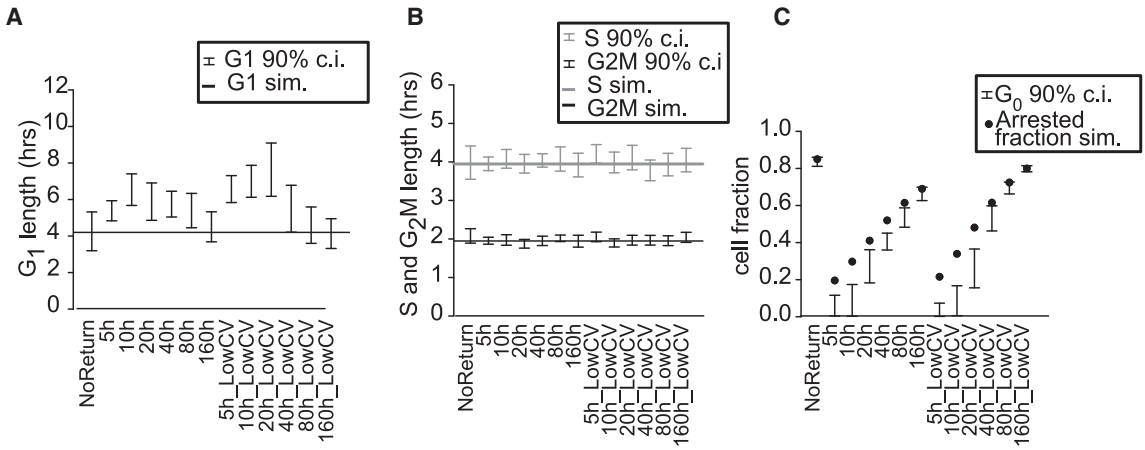


Figure 4. Parameters inferred by CycleFlow from synthetic data including cell-cycle arrest

(A) G_1 length versus mean cell-cycle arrest time used in simulation. "NoReturn" indicates that cells arrest irreversibly. The coefficient of variation (CV) of cell-cycle arrest was set to 1.0 except for the points annotated with "LowCV," where it was 0.13. Error bars indicate 90% credible intervals; the G_1 length used in the simulations is shown as a solid line.

(B) Like (A) but for S (gray) and G_2/M (black) phase durations.

(C) G_0 fraction inferred by CycleFlow (90% credible intervals; bars) and simulated arrested fractions (dots).

DNA-damage and repair (Spencer et al., 2013; Barr et al., 2016; Ryl et al., 2017; Yang et al., 2017; Arora et al., 2017), from long-term quiescence as part of a developmental program. We generated synthetic EdU labeling data for a simulated population based on cell-cycle parameters as inferred for DP cells. Instead of entering quiescence irreversibly, the simulated population could now arrest the cycle at *2n* DNA content for variable times and then re-enter into the cycle (Figure S1C). Applying CycleFlow to these data, we found that transient arrests (less than four unarrested cycle durations) were assigned as slow passages through G_1 (Figure 4A). Long-term arrests (exceeding 8 cycles) were assigned to G_0 and thus did not affect inferred G_1 duration (Figure 4C). Arrests between 4 and 8 cycles were split between G_1 and G_0, which reduced the quality of fit. Importantly, the inferred durations of S and G_2/M phases were unaffected by cycle re-entry in all cases (Figure 4B). We conclude that CycleFlow robustly resolves heterogeneity between cycling cells (some of which may undergo transient arrest) and long-term quiescence.

DISCUSSION

By exploiting joint information on DNA content and EdU label progression, CycleFlow determines the long-term quiescent fraction within a cell population along with all the cell-cycle phase durations of the proliferating cell fraction. Our approach overcomes practical problems, including the imperfect correspondence between cell-cycle phases and DNA content gates and the identification of the EdU removal kinetics. CycleFlow is equally applicable *in vitro* and *in vivo* as well as to growing or steady-state cell populations. In particular, CycleFlow enabled us to resolve the cycling characteristics of the DP thymocytes and the relative size of the quiescent subpopulation. Our finding that more than 80% of DP thymocytes are quiescent resolves a long-standing question in the immunological literature that is relevant for understanding the mechanisms of thymocyte selection (Yates, 2014). Application to other cell types is straightforward (e.g., Carvajal Ibañez et al., 2022): the measurements should extend over roughly one cell cycle with several (typically 4–8) time points. The model fitting and parameter inference are facilitated by a publicly available Python package.

Limitations of the study

As with any model-based inference method, the applicability of CycleFlow is limited by the validity of the underlying model assumptions. Specifically, all cycle phase durations are modeled as Erlang distributed, and exit from the cycle into G_0 is assumed to be effectively irreversible. While the Erlang shape of S and G_2 is reasonable for a large class of mammalian cells (Chao et al., 2019), we found that the corresponding substep numbers cannot be identified from data. Here, CycleFlow resorts to fixing substep numbers in order to match typical observed coefficients of variation (Chao et al., 2019). The substep number for G_1 is identifiable from data, and CycleFlow infers it. In some cell populations, transient arrest after division may occur, with a duration similar to the cell cycle. In that situation, the effective distribution of the time until S phase is poorly captured by an Erlang shape, CycleFlow will fit the data worse, and the inferred G_1 substep numbers and G_0 fractions may be distorted (cf. Figure 4). If transient arrest occurred in a large fraction of cycling cells, so that the Erlang assumption will be violated, an extension of the model, as well as additional data separating arrest from quiescence, may be warranted.

Loss of cells due to differentiation or death is modeled as uniform over all phases of the cycle, which is reasonable in absence of more detailed knowledge. If loss is known to be restricted to specific phases, this information can be incorporated straightforwardly by adapting the model. Moreover, CycleFlow could be extended to infer non-uniform loss rates for all phases, but whether these will be identifiable may depend on the particular

dataset. Conversely, gain of cells by differentiation of a progenitor cell type is not included by the CycleFlow model, which describes an essentially self-sustaining population. Incorporating influx would require another extension of the model and additional data, for instance label kinetics of the upstream population. These possible extensions are an interesting avenue for future work.

STAR★METHODS

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

- KEY RESOURCES TABLE
- RESOURCE AVAILABILITY
  - Lead contact
  - Materials availability
  - Data and code availability
- EXPERIMENTAL MODEL AND SUBJECT DETAILS
  - Mice
  - Cell line
- METHOD DETAILS
  - Labeling and analysis of TET21N cells *in vitro*
  - Labeling of thymocytes *in vivo* and analysis
  - Cell sorting for single-cell RNA-seq
  - Single-cell RNA-seq data analysis
  - Parameter inference
  - Model for cell-cycle progression
  - Steady-growth distribution
  - Steady-state distribution
  - Kinetics of labeling
  - Labeled population dynamics
  - Parameter estimation
  - Extended cell cycle model with transient arrest
  - Inference of cell cycle parameters from synthetic data including transient arrests
  - Validation of steady state assumption
- QUANTIFICATION AND STATISTICAL ANALYSIS
  - Implementation and software

SUPPLEMENTAL INFORMATION

Supplemental information can be found online at https://doi.org/10.1016/j.crmeth.2022.100315.

ACKNOWLEDGMENTS

Funding through the Computational Life Sciences program of the Federal Ministry of Education and Research (FKZ 031L0170A, to T.H.) and DKFZ core funding are gratefully acknowledged. We thank Hans-Reimer Rodewald for generous support and discussions, as well as Katrin Busch and all members of the Höfer group for discussions.

AUTHOR CONTRIBUTIONS

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

DECLARATION OF INTERESTS

The authors declare no competing interests.

Received: November 13, 2020
Revised: July 25, 2022
Accepted: September 14, 2022
Published: October 6, 2022

REFERENCES

Aguirre-Ghiso, J.A. (2007). Models, mechanisms and clinical evidence for cancer dormancy. Nat. Rev. Cancer *7*, 834–846.

Akinduro, O., Weber, T.S., Ang, H., Haltalli, M.L.R., Ruivo, N., Duarte, D., Rashidi, N.M., Hawkins, E.D., Duffy, K.R., and Lo Celso, C. (2018). Proliferation dynamics of acute myeloid leukaemia and haematopoietic progenitors competing for bone marrow space. Nat. Commun. *9*, 519.

Arora, M., Moser, J., Phadke, H., Basha, A.A., and Spencer, S.L. (2017). Endogenous replication stress in mother cells leads to quiescence of daughter cells. Cell Rep. *19*, 1351–1364.

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

Busch, K., Klapproth, K., Barile, M., Flossdorf, M., Holland-Letz, T., Schlenner, S.M., Reth, M., Höfer, T., and Rodewald, H.R. (2015). Fundamental properties of unperturbed haematopoiesis from stem cells in vivo. Nature *518*, 542–546.

Carvajal Ibañez, D., Skabkin, M., Hooli, J., Cerrizuela, S., Göpferich, M., Jolly, A., Zumwinkel, M., Bertolini, M., Höfer, T., Kramer, G., et al. (2022). Interferon regulates stem cell function at all ages by orchestrating mTOR and cell cycle. Preprint at bioRxiv. https://doi.org/10.1101/2022.02.03.478954.

Chao, H.X., Fakhreddin, R.I., Shimerov, H.K., Kedziora, K.M., Kumar, R.J., Perez, J., Limas, J.C., Grant, G.D., Cook, J.G., Gupta, G.P., and Purvis, J.E. (2019). Evidence that the human cell cycle is a series of uncoupled, memoryless phases. Mol. Syst. Biol. *15*, e8604.

Cheraghali, A.M., Kumar, R., Knaus, E.E., and Wiebe, L.I. (1995). Pharmacokinetics and bioavailability of 5-ethyl-2'-deoxyuridine and its novel (5R, 6R)-5-bromo-6-ethoxy-5, 6-dihydro prodrugs in mice. Drug Metab. Dispos. *23*, 223–226.

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

Foreman-Mackey, D., Hogg, D.W., Lang, D., and Goodman, J. (2013). emcee: the MCMC Hammer. Publ. Astron. Soc. Pac. *125*, 306–312.

Foudi, A., Hochedlinger, K., Van Buren, D., Schindler, J.W., Jaenisch, R., Carey, V., and Hock, H. (2009). Analysis of histone 2B-GFP retention reveals slowly cycling hematopoietic stem cells. Nat. Biotechnol. *27*, 84–90.

Gitlin, A.D., Mayer, C.T., Oliveira, T.Y., Shulman, Z., Jones, M.J.K., Koren, A., and Nussenzweig, M.C. (2015). T cell help controls the speed of the cell cycle in germinal center B cells. Science *349*, 643–646.

Hsiao, C.J., Tung, P., Blischak, J.D., Burnett, J.E., Barr, K.A., Dey, K.K., Stephens, M., and Gilad, Y. (2020). Characterizing and inferring quantitative cell cycle phase in single-cell RNA-seq data analysis. Genome Res. *30*, 611–621.

Kretschmer, L., Flossdorf, M., Mir, J., Cho, Y.L., Plambeck, M., Treise, I., Toska, A., Heinzel, S., Schiemann, M., Busch, D.H., and Buchholz, V.R. (2020). Differential expansion of T central memory precursor and effector subsets is regulated by division speed. Nat. Commun. *11*, 113.

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

Lahoz-Beneytez, J., Elemans, M., Zhang, Y., Ahmed, R., Salam, A., Block, M., Niederalt, C., Asquith, B., and Macallan, D. (2016). Human neutrophil kinetics: modeling of stable isotope labeling data supports short blood neutrophil half-lives. Blood *127*, 3431–3438.

Li, L., and Clevers, H. (2010). Coexistence of quiescent and active adult stem cells in mammals. Science *327*, 542–545.

Lutz, W., Stöhr, M., Schürmann, J., Wenzel, A., Löhr, A., and Schwab, M. (1996). Conditional expression of n-myc in human neuroblastoma cells increases expression of α-prothymosin and ornithine decarboxylase and accelerates progression into s-phase early after mitogenic stimulation of quiescent cells. Oncogene *13*, 803–812.

Miller, I., Min, M., Yang, C., Tian, C., Gookin, S., Carter, D., and Spencer, S.L. (2018). Ki67 is a graded rather than a binary marker of proliferation versus quiescence. Cell Rep. *24*, 1105–1112.e5.

Mingueneau, M., Kreslavsky, T., Gray, D., Heng, T., Cruse, R., Ericson, J., Bendall, S., Spitzer, M.H., Nolan, G.P., Kobayashi, K., et al. (2013). The transcriptional landscape of alpha beta T cell differentiation. Nat. Immunol. *14*, 619–632.

Morcos, M.N.F., Zerjatke, T., Glauche, I., Munz, C.M., Ge, Y., Petzold, A., Reinhardt, S., Dahl, A., Anstee, N.S., Bogeska, R., et al. (2020). Continuous mitotic activity of primitive hematopoietic stem cells in adult mice. J. Exp. Med. *217*, e20191284.

Ryl, T., Kuchen, E.E., Bell, E., Shao, C., Flórez, A.F., Mönke, G., Gogolin, S., Friedrich, M., Lamprecht, F., Westermann, F., and Höfer, T. (2017). Cell-cycle position of single MYC-driven cancer cells dictates their susceptibility to a chemotherapeutic drug. Cell Syst. *5*, 237–250.e8.

Schwabe, D., Formichetti, S., Junker, J.P., Falcke, M., and Rajewsky, N. (2020). The transcriptome dynamics of single cells during the cell cycle. Mol. Syst. Biol. *16*, e9946.

Spencer, S.L., Cappell, S.D., Tsai, F.C., Overton, K.W., Wang, C.L., and Meyer, T. (2013). The proliferation-quiescence decision is controlled by a bifurcation in CDK2 activity at mitotic exit. Cell *155*, 369–383.

Srinivas, S., Watanabe, T., Lin, C.S., William, C.M., Tanabe, Y., Jessell, T.M., and Costantini, F. (2001). Cre reporter strains produced by targeted insertion of EYFP and ECFP into the ROSA26 locus. BMC Dev. Biol. *1*, 4.

Stuart, T., Butler, A., Hoffman, P., Hafemeister, C., Papalexi, E., Mauck, W.M., 3rd, Hao, Y., Stoeckius, M., Smibert, P., Satija, R., and Satija, R. (2019). Comprehensive integration of single-cell data. Cell *177*, 1888–1902.e21.

van Velthoven, C.T.J., and Rando, T.A. (2019). Stem cell quiescence: dynamism, restraint, and cellular idling. Cell Stem Cell *24*, 213–225.

Wilson, A., Laurenti, E., Oser, G., van der Wath, R.C., Blanco-Bose, W., Jaworski, M., Offner, S., Dunant, C.F., Eshkind, L., Bockamp, E., et al. (2008). Hematopoietic stem cells reversibly switch from dormancy to self-renewal during homeostasis and repair. Cell *135*, 1118–1129.

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

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

Yui, M.A., and Rothenberg, E.V. (2014). Developmental gene networks: a triathlon on the course to T cell identity. Nat. Rev. Immunol. *14*, 529–545.

CellPress
OPEN ACCESS

STAR★METHODS

KEY RESOURCES TABLE

REAGENT or RESOURCE	SOURCE	IDENTIFIER
Antibodies		
CD8a (clone 53–6.7) allophycocyanin (APC)	eBioscience	Cat#17–0081; RRID: AB_469335
TCRgd (clone GL3) BV421	BioLegend	Cat#118120; RRID: AB_2562566
CD4 (clone H129.19) phycoerythrin (PE)	BD Pharmingen	Cat#553652
Ter119 (clone Ter119) BV421	Biolegend	Cat#116234; RRID: AB_2562917
CD11b (clone: M1/70) BV421	BioLegend	Cat#101251; RRID: AB_2562904
CD19 (clone: 6D5) BV421	BioLegend	Cat#115549; RRID: AB_2563066
NK1.1 (clone PK136) BV421	BioLegend	Cat#108731; RRID: AB_10895916
Gr-1 (clone RB6-8C5) BV421	BioLegend	Cat#108445; RRID: AB_2562903
Critical commercial assays		
Click-iT™ Plus EdU Alexa Fluor™ 488 Flow Cytometry Assay Kit	Invitrogen	Cat#C10632
Zombie Red™ Fixable Viability Kit	BioLegend	Cat#423109
FxCycle™ Violet Stain	Invitrogen	Cat#F10347
Deposited data		
Raw and analysed scRNA-seq data	This paper	GEO: GSE196412
Thymocytes flow cytometry data	This paper	FlowRepository: FR-FCM-Z3D7
TET21/N flow cytometry data	This paper	FlowRepository: FR-FCM-Z3ER
Experimental models: Cell lines		
TET21/N (Lutz et al., 1996)	DKFZ	RRID: CVCL_9812
Experimental models: Organisms/strains		
mouse:C57BL/6-Tie2$^{MCM/+}$ Rosa26$^{YFP/+}$	Busch et al. (2015)	N/A
mouse: C57BL/6-Rosa26^{YFP}	Srinivas et al. (2001)	N/A
mouse: C57BL/6-Cd45.1	In-house colony	N/A
Software and algorithms		
CycleFlow	This paper	https://doi.org/10.5281/zenodo.7037488
Emcee (3.02)	Foreman-Mackey et al. (2013)	https://emcee.readthedocs.io

RESOURCE AVAILABILITY

Lead contact

Further information and requests for resources and reagents should be directed to and will be fulfilled by the lead contact, Thomas Höfer (t.hoefer@dkfz.de).

Materials availability

This Study did not generate new unique reagents.

Data and code availability

- Single-cell RNA-seq data have been deposited at GEO and are publicly available as of the date of publication. Flow Cytometry data have been deposited at FlowRepository and are publicly available as of the date of publication. The Accession numbers are listed in the key resources table.
- All original code has been deposited at Zenodo and is publicly available as of the date of publication. the DOI is listed in the key resources table.
- Any additional information required to reanalyze the data reported in this paper is available from the lead contact upon request.

EXPERIMENTAL MODEL AND SUBJECT DETAILS

Mice

Mice of ages between eleven and sixteen weeks were used, both female and male. No randomization was done, no blinding was done and no animals were excluded from the analysis. 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).

Cell line

SH-EP TET21/N (TET21N) cells were grown in RPMI 1640 medium supplemented with 10% Fetal Calf Serum (FCS) at 37° C, 5% CO_2 and 88% humidity. Versene was used for all harvesting steps. The Cell line was authenticated by the DKFZ in house facility.

METHOD DETAILS

Labeling and analysis of TET21N cells *in vitro*

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.

Labeling of thymocytes *in vivo* and analysis

Mice 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 μg / ml purified mouse IgG (Jackson ImmunoResearch Laboratories). Antibody stainings were performed in PBS/5% FCS on ice for 30 min with optimal dilutions of commercially-prepared antibodies. The following antibodies were used: CD8a allophycocyanin (APC) (53–6.7) from eBioscience; CD4 phycoerythrin (PE) (H129.19) from BD Pharmingen; CD11b BV421 (M1/70), CD19 BV421 (6D5), NK1.1 BV421 (PK136), Gr-1 BV421 (RB6-8C5), TCR gd BV421 (GL3) and Ter119 BV421 (Ter119) from Biolegend. The lineage cocktail was composed of CD11b, CD19, Ter119, NK1.1, Gr-1 and TCR gd. After antibody staining, cells were 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 $lineage^- CD4^+CD8^+$.

Cell sorting for single-cell RNA-seq

Thymocytes were isolated and stained with antibodies as in the above section. Double positive thymocytes were sorted (BD-FACSAriaIII, 100 μm nozzle) into 500 μL PBS containing 50% FCS. The sorted cells were spun down at 300 g for 5 min at 4°C and resuspended in 35 μL FACS buffer. Single cells were captured using the Chromium System and single cell full transcriptome and targeted TCR transcriptome libraries were generated according to the Chromium Single Cell V(D)J and Reagent Kits protocols (10x Genomics). During the library preparation, reverse transcribed cDNA was amplified 14 times in accordance with 10x Genomics protocols. The full transcriptomes were sequenced on an Illumina HiSeq4000 platform (26 + 74 bp read-length).

Single-cell RNA-seq data analysis

Samples were aligned to mouse genome (mm10) and read counting per cell was performed using the software CellRanger3.0.1 from 10x genomics. The raw count matrix produced by CellRanger was then further processed with the Bioconductor pipeline. SingleCellExperiment R objects were generated and empty droplets were excluded using the function emptyDrops() from the package DropletUtils. Cells with low UMI counts, low number of expressed genes and/or high mitochondrial gene expression were also removed using isOutlier from the package scater. The data was then normalized using logNormCounts function from the package scran. Cell cycle annotation was performed using Seurat v3. We then performed principal component analysis (PCA) using exclusively cell cycle genes defined within the Seurat package. In order to perform RNA-velocity, spliced and unspliced matrices were generated using the function run10x from the python implementation of the software velocyto. We performed RNA velocity using the scVelo algorithm implemented in the R package velociraptor on a subset of 1354 genes annotated with Gene Ontology term "Cell Cycle". The result of the RNA velocity analysis was embedded into the PCA using the function embedVelocity from velociraptor.

Parameter inference

CycleFlow addresses the most important challenges when interpreting thymidine-analog labeling time-course data by incorporating known sources of uncertainty into its dynamical model, as detailed below. For a summary, see Table S1.

Model for cell-cycle progression

Within a growing cell population, each cell progresses through the cycle phases G_1, S and G_2/ M (Figure S1A). 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 (The G_2 subphases also comprise M phase for simplicity of notation). The subphases track the approximate progress of a cell within the respective cycle phase. We 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 and ν, for G_1, S and G_2, respectively. Within 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. Each individual daughter cell may then, with probability a, exit the cell cycle into the quiescent state G_0, or, with probability $1 - a$, restart the cycle in phase $G_{1,1}$. In the main text and in this section, 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. More generally, cells may be allowed to re-enter the cycle from G_0; the implications are discussed further below. Finally, cells may be lost from any cycle phase or G_0 with rate δ, due to cell death or differentiation. The parameter δ is an effective loss rate average over all stages of the cell cycle. In practice, we found that a uniform loss rate was sufficient to fit our data.

Within this model the dynamics of subpopulations is described by a linear system of ordinary differential equations. Written in matrix form,

$$\frac{d}{dt}\boldsymbol{N} = \mathbf{T}\boldsymbol{N} - \delta\boldsymbol{N}, \qquad \text{(Equation 1)}$$

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

$$\boldsymbol{N} = \left[G_{1,1}(t), \ldots, G_{1,l}(t), S_1(t), \ldots, S_m(t), G_{2,1}(t), \ldots, G_{2,n}(t), G_0(t)\right]^T,$$

and the rate matrix

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

CycleFlow presumes that the samples originate from a cell population deep in the exponential growth phase, where 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, we can decompose the cell numbers into an exponentially growing population $N(t)$ and constant cycle fractions $\boldsymbol{n} = [g_{1,1}, \ldots, g_0]$, that is,

$$\boldsymbol{N}(t) = N(t)\,\boldsymbol{n} = N(0)e^{(\kappa-\delta)t}\left[g_{1,1}, \ldots, g_{1,l}, s_1, \ldots, s_m, g_{2,1}, \ldots, g_{2,n}, g_0\right]. \qquad \text{(Equation 3)}$$

Here, we have introduced a new rate κ by denoting the net population growth rate as $\kappa - \delta$. The proportions of cells in the various subphases are given by the steady-growth distribution $\boldsymbol{n}$, which is normalized: $\mathbf{1}^T\boldsymbol{n} = 1$, where we use the abbreviation $\mathbf{1} = [1, \ldots, 1]^T$. The dynamical Equation 1 then yields

$$\frac{d}{dt}\boldsymbol{N}(t) = (\kappa - \delta)N(t)\,\boldsymbol{n} = N(t)[\mathbf{T}\boldsymbol{n} - \delta\boldsymbol{n}], \quad \text{which implies } \kappa\boldsymbol{n} = \mathbf{T}\boldsymbol{n}. \qquad \text{(Equation 4)}$$

Thus κ is seen to equal the dominant eigenvalue of the matrix $\mathbf{T}$, and $\boldsymbol{n}$ is the associated normalized (right) eigenvector. The rate κ is the growth rate the population would have in the absence of loss.

Steady-growth distribution

To find $\boldsymbol{n}$, we start by normalizing Equation 4. Using Equation 2, we obtain

$$\kappa = \mathbf{1}^T\mathbf{T}\boldsymbol{n} = [0, \ldots, 0, \nu, 0]^T\boldsymbol{n} = \nu g_{2,n}. \qquad \text{(Equation 5a)}$$

Equation 5a implies that the total rate of production of new cells κN equals the population in the final subphase $G_{2,n}$ times its rate of completion ν. Now consider steady growth with production of new cells, where $G_{2,n}$ and κ are positive. Starting from Equation 5a and backsubstituting using Equation 4, we compute the remainder of the cycle subphase distribution $\boldsymbol{n}$ in steady growth as a function of κ:

$$g_0 = g_{2,n} 2a\nu/\kappa = 2a \quad \text{(Equation 5b)}$$

$$g_{2,n-j} = g_{2,n}\frac{(\kappa+\nu)^j}{\nu^j} = \frac{\kappa}{\nu}\left(\frac{\kappa+\nu}{\nu}\right)^j \qquad 0<j<n \quad \text{(Equation 5c)}$$

$$s_m = g_{2,1}\frac{\kappa+\nu}{\mu} = \frac{\kappa}{\mu}\left(\frac{\kappa+\nu}{\nu}\right)^n \quad \text{(Equation 5d)}$$

$$s_{m-j} = s_m\frac{(\kappa+\mu)^j}{\mu^j} = \frac{\kappa}{\mu}\left(\frac{\kappa+\nu}{\nu}\right)^n\left(\frac{\kappa+\mu}{\mu}\right)^j \qquad 0<j<m \quad \text{(Equation 5e)}$$

$$g_{1,l} = s_1\frac{\kappa+\mu}{\lambda} = \frac{\kappa}{\lambda}\left(\frac{\kappa+\nu}{\nu}\right)^n\left(\frac{\kappa+\mu}{\mu}\right)^m \quad \text{(Equation 5f)}$$

$$g_{1,l-j} = g_{1,l}\frac{(\kappa+\lambda)^j}{\lambda^j} = \frac{\kappa}{\lambda}\left(\frac{\kappa+\nu}{\nu}\right)^n\left(\frac{\kappa+\mu}{\mu}\right)^m\left(\frac{\kappa+\lambda}{\lambda}\right)^j \qquad 0<j<l \quad \text{(Equation 5g)}$$

To complete the calculation, the dominant eigenvalue κ can be found numerically by solving Equation 4 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. When $a \leq 1/2$ then also $g_0 \leq 1$ as required, Equation 5b.

Steady-state distribution

To capture cell populations that are maintained at a fixed population size within the model Equation 1, we consider the loss rate δ to be subject to implicit homeostatic regulation. The regulation maintains $\delta \equiv \kappa$, so that proliferation and loss balance, and net growth is zero. In this way, homeostasis is treated as a marginal case of steady growth, Equation 3, with vanishing effective growth rate and subphase populations $\boldsymbol{N} = N\boldsymbol{n}$ that are constant in time. We can compute the steady-state distribution $\boldsymbol{n}$ by Equation 5 as before. The homeostatic set point of δ is obtained numerically by solving the eigenvalue problem Equation 4, with $\kappa \to \delta$.

Kinetics of labeling

In order to relate the cycling model Equation 1 to experimental data, we include the dynamics of label incorporation and inheritance (Figure S1B). As long as EdU is available in the medium, unlabeled cells in S phase incorporate it, thereby transitioning into a labeled state at a rate ε. The labeled states (denoted S^*, $G^*_{2,1}$ etc.) are defined operationally: Any cell that is gated above the background fluorescence level in the EdU fluorescence channel is considered labeled. In practice we find that transition to the labeled state requires only a small part of S phase, in other words, the rate of label acquisition $\varepsilon > \mu/m$. It is unclear if label acquisition is also faster than progress from one S subphase to the next, $\varepsilon \gtrless \mu$. We therefore consider S phase progress and labeling to be parallel processes, as shown in Figure S1B. As 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 in cells.

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 removal by an exponential decrease of the total labeling rate

$$\varepsilon = \varepsilon(t) = \varepsilon_0 e^{-t/\tau_E}.$$

Both the initial labeling rate ε_0 and label degradation time scale τ_E are adjustable parameters. In practical parameter inference from data, we find that these two parameters are negatively correlated, and that ε_0 admits only a lower bound, but the label degradation time τ_E remains identifiable.

Labeled population dynamics

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

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

unlabeled cell numbers are denoted similarly by $\boldsymbol{N}^\circ$. The two 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 Equation 3, with a steady-growth distribution $\boldsymbol{n}$ satisfying Equation 4. At times $t > 0$, label acquisition transfers cells into the labeled populations, so that Equation 1 then still 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:

$$(\mathbf{L})_{ij} = \begin{cases} \varepsilon & \text{if } l < j = i \leq l + m \\ 0 & \text{otherwise} \end{cases}.$$

The labeling matrix $\mathbf{L}$ has the same dimensions as $\mathbf{T}$; its nonzero entries are ε on the diagonal for all subphases of S. It can be verified straightforwardly that $\mathbf{L}\boldsymbol{N}^\circ$ is the flux from unlabeled to labeled compartments as shown in Figure S1B. Using this notation,

$$\begin{aligned} \frac{d}{dt}\boldsymbol{N}^* &= (\mathbf{T} - \delta)\boldsymbol{N}^* + \mathbf{L}\boldsymbol{N}^\circ \\ &= (\mathbf{T} - \mathbf{L} - \delta)\boldsymbol{N}^* + \mathbf{L}\boldsymbol{N} \end{aligned} \qquad \text{(Equation 6a)}$$

$$= (\mathbf{T} - \mathbf{L} - \delta)\boldsymbol{N}^* + N\mathbf{L}\boldsymbol{n}. \qquad \text{(Equation 6b)}$$

In Equation 6a, we have used the fact that already-labeled cells follow the undisturbed cycle progression ($\mathbf{T} - \delta$); in Equation 6b, 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, $\boldsymbol{n}^* = [\mathrm{g}^*_{1,1}, \ldots, \mathrm{g}^*_0] = \boldsymbol{N}^*/N$. From Equation 6b and using exponential growth of $N(t)$, we obtain

$$\frac{d}{dt}\boldsymbol{n}^* = (\mathbf{T} - \mathbf{L} - \kappa)\boldsymbol{n}^* + \mathbf{L}\boldsymbol{n}. \qquad \text{(Equation 7)}$$

The appropriate initial condition is $\boldsymbol{n}^*(0) = \mathbf{0}$. The unlabeled fractions are then obtained as $\boldsymbol{n}^\circ = \boldsymbol{n} - \boldsymbol{n}^*$.

Parameter estimation

In experiment, cycle phase is assigned by DNA content. Thus, cells in G_0 or G_1 fall into a single experimental gate, $\overline{\mathrm{G01}}$. Although most DNA-replicating cells are correctly counted in the corresponding $\overline{\mathrm{S}}$ gate, the gating procedure inevitably assigns some cells that have started DNA replication to the $\overline{\mathrm{G01}}$ gate, and some cells that have completed replication to the $\overline{\mathrm{S}}$ gate rather than the $\overline{\mathrm{G2M}}$ gate. When deriving model predictions, we account for this crosstalk by assigning the first $m_{\overline{01}}$ subphases of S phase to $\overline{\mathrm{G01}}$, and the last $m_{\overline{2}}$ subphases to $\overline{\mathrm{G2M}}$, so that only the remaining innermost $m - m_{\overline{01}} - m_{\overline{2}}$ subphases S_i are assigned to the gate $\overline{\mathrm{S}}$, as shown in Equation 8.

$$\underbrace{\mathrm{G}_0;\ \mathrm{G}_{1,1} \cdots \mathrm{G}_{1,l}}_{\overline{\mathrm{G01}}}\ \underbrace{\mathrm{S}_1 \cdots \mathrm{S}_{m_{\overline{01}}} \mathrm{S}_{m_{\overline{01}}+1} \cdots \mathrm{S}_{m-m_{\overline{2}}}}_{\overline{\mathrm{S}}}\ \underbrace{\mathrm{S}_{m-m_{\overline{2}}+1} \cdots \mathrm{S}_m \mathrm{G}_{2,1} \cdots \mathrm{G}_{2,n}}_{\overline{\mathrm{G2M}}} \qquad \text{(Equation 8)}$$

Conversely, misassignment of G_0, G_1 or G_2 cells to the $\overline{\mathrm{S}}$ gate was not observed in our data: At the first time point after labeling, the $\overline{\mathrm{S}}$ gate contained neglibigle amounts of unlabeled cells (Figure 1E).

In order to compute model predictions, we first solve Equation 4 for the constant steady-growth distribution $\boldsymbol{n}$. From $\boldsymbol{n}$ we obtain the time-independent total cell fractions in each experimental gate, *e.g.*

$$\overline{\mathrm{g2}}_{\mathrm{pred}} = \sum_{i=m-m_{\overline{2}}+1}^{m} \mathrm{s}_i + \sum_{i=1}^{n} \mathrm{g}_{2,i}.$$

We then solve system Equation 7 numerically. From $\boldsymbol{n}^*(t)$, we obtain predictions for the time-dependent labeled fractions, *e.g.*

$$\overline{\mathrm{g01}}^*_{\mathrm{pred}}(t) = \mathrm{g}^*_0(t) + \sum_{i=1}^{l} \mathrm{g}^*_{1,i}(t) + \sum_{i=1}^{m_{\overline{01}}} \mathrm{s}^*_i(t).$$

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

$$\chi^2 = \frac{1}{2}\sum_{\overline{g}=\overline{g01},\overline{g2}} \frac{\left(\overline{g}_{\text{pred}} - \overline{g}_{\text{obs}}\right)^2}{\sigma_{\overline{g}}{}^2} + \frac{1}{2}\sum_{i=1\ldots I}\sum_{\overline{g}^*=\overline{g01}^*,\overline{s}^*,\overline{g2}^*} \frac{\left(\overline{g}^*_{\text{pred}}(t_i) - \overline{g}^*_{\text{obs}}(t_i)\right)^2}{\sigma_{\overline{g}^*}{}^2}. \qquad \text{(Equation 9)}$$

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 (double) 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 Equation 9, 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 Table S2. As the substep numbers m and n of S and G_2 phase, respectively, were only lower-bounded by this procedure, we fixed them to a value $n = m = 15$, well above the lower bound. The credible intervals resulting from sampled posterior distributions for DP and TET21N are shown in Table S3.

The model assumes a cell population at steady growth, and we include the steady-state fractions directly as the first term in Equation 9. The steady-growth assumption can be checked for consistency by comparing total cycle fractions at successive time points. Only the labeled fractions are included directly in Equation 9, which implies that the residuals of the corresponding predicted unlabeled fractions $(\overline{g}^{\circ}_{\text{pred}}(t_i) - \overline{g}^{\circ}_{\text{obs}}(t_i))$ can provide another consistency check. If these residuals were to exceed the fitted residuals strongly, this would indicate a possible lack of steady growth.

Extended cell cycle model with transient arrest

We now consider an extension of the CycleFlow model where cells are allowed to arrest the cell cycle temporarily but then return into $G_{1,1}$ (Figure S1C). We extend the transition scheme by introducing a reversible arrest phase A with p subphases $A_1, \ldots, A_p$ in analogy to the other cycle phases, which replaces the irreversible entry into the quiescent state G_0. The vector of cell numbers becomes $\boldsymbol{N} = [G_{1,1}(t), \ldots, S_1(t), \ldots, G_{2,n}(t), A_1(t), \ldots, A_p(t)]^T$. The mean arrest time is p/π, where π is the transition rate between subphases of A. An exponential arrest time with a CV of 1 is achieved by setting $p = 1$, and more controlled arrest times are achieved by higher p; high values $p > 10$ resemble a fixed delay. The transition rate matrix is extended accordingly,

$$\mathbf{T} = \begin{bmatrix}
-\lambda & & & & & & & & & & & & & & \pi \\
\lambda & \ddots & & & & & & & & & & & & & 0 \\
 & \ddots & -\lambda & & & & & & & & & & & & \vdots \\
 & & \lambda & -\mu & & & & & & & & & & & \\
 & & & \mu & \ddots & & & & & & & & & & \\
 & & & & \ddots & -\mu & & & & & & & & & \\
 & & & & & \mu & -\nu & & & & & & & & \\
 & & & & & & \nu & \ddots & & & & & & & \\
 & & & & & & & \ddots & -\nu & & & & & & \\
 & & & & & & & & \nu & -\nu & & & & & \\
 & & & & & & & & & 2a\nu & & -\pi & & & \\
 & & & & & & & & & & & \pi & \ddots & & \vdots \\
 & & & & & & & & & & & & \ddots & -\pi & 0 \\
 & & & & & & & & & & & & & \pi & -\pi
\end{bmatrix}.$$

The steady state can be expressed as a function of the steady growth rate in a similar way as in the model without cycle re-entry. Equation 5a still holds unchanged in steady growth:

$$\kappa = \mathbf{1}^T\mathbf{T}\boldsymbol{n} = \nu g_{2,n}. \qquad \text{(Equation 10a)}$$

Indeed, the production rate of new cells κN still equals the population in subphase $G_{2,n}$ times its rate of completion ν, irrespective of return from A. Return from A does increase growth as expected intuitively, because it leads to a higher influx back into G_1, which entails a higher population in $G_{2,n}$ and thereby higher growth rate κ. Again backsubstituting using Equation 4, one finds the general rule $\varphi_{i-1}r_{\rightarrow i} = \varphi_i(r_{i\rightarrow} + \kappa)$. Here, φ_i for $2 \le i \le I + m + n + p$ denotes a subphase with influx and outflux rates $r_{\rightarrow i}$ and $r_{i\rightarrow}$, respectively. Application of the rule then yields the steady state fractions

$$a_1 = g_{2,n}\frac{2a\nu}{\kappa + \pi} = \frac{2a\kappa}{\kappa + \pi} \qquad \text{(Equation 10b)}$$

$$a_{1+j} = a_1 \frac{\pi^j}{(\kappa+\pi)^j} = \frac{2a\kappa\pi^j}{(\kappa+\pi)^{j+1}} 0 < j < p. \quad \text{(Equation 10c)}$$

The remaining steady state subphase fractions remain exactly as given in Equations 5c, 5d, 5e, 5f, and 5g, which shows that re-entry into the cycle affects the non-arrested cell cycle phases' relative occupancies only through the change in the steady growth rate κ. (κ is be determined numerically, as before.)

Because $\kappa \lesssim 1/T_c$ where T_c is the average cell cycle time, whereas the individual subphase transition rates $\lambda, \mu, \nu \sim (l + m + n + p)/T_c$, the factors $\kappa + \lambda$ etc. appearing in Equations 5 and 10 are dominated by the subphase transition rates and change only slightly when κ is varied moderately. This fact explains our numerical observation that the relative subphase occupations among the cycling phases are not changed appreciably by the moderate increase in κ due to cycle re-entry.

Inference of cell cycle parameters from synthetic data including transient arrests

We performed stochastic simulations of the extended model with reentry of cells from A to G_1. We chose an example population with $l = 2$, $m = n = 15$ cycle subphases and either $p = 1$ A subphase, corresponding to an exponential arrest time (with coefficient of variation (CV) 1), or $p = 60$, corresponding to a nearly deterministic arrest time, with CV 0.13.

Each simulation was initalized with 10,000 EdU negative cells distributed according to steady state determined as above. We applied cell dilution with rate $\delta \equiv \kappa$ to ensure that the cell number remained constant on average. The transition rates λ, μ and ν were set to the median values inferred for the DP thymocytes; the label kinetics were given by $\varepsilon_0 = 5/\text{h}$ and $\tau = 1\text{h}$. For each of the two cases (arrest time CV 0.13 or 1), we varied π to obtain mean arrest durations of 5h, 10h, 20h, 40h, 80h and 160h. We assigned the subphases to EdU positive and negative $\overline{\text{G01}}$, $\overline{\text{S}}$ and $\overline{\text{G2M}}$ gates with perfect gating, without misclassified S phase cells. For each set of parameters, from 6 simulation runs we computed the mean fractions and their standard errors at 3h, 5h, 8h, 12h, 14h and 18h, to reproduce the experimental data acquisition. We then carried out parameter inference using CycleFlow, i.e. without return from G_0. Here we assumed perfect gating, setting $m_{\overline{01}} = m_{\overline{2}} = 0$.

Validation of steady state assumption

CycleFlow is applicable to biological systems a steady state or steady growth. This condition can be tested by measuring the total fraction of cells in each cell cycle phase over time; these must remain constant to satisfy the steady state condition.

Alternatively this condition can be tested by comparing the predicted time-dependent fraction of unlabeled cell to actual measurements not used for fitting. Figure S3 shows that for both Tet21N A and DP thymocytes B, the predicted fractions of unlabeled cells match the data, thereby confirming that the steady state condition is fulfilled.

QUANTIFICATION AND STATISTICAL ANALYSIS

The number of replicates (n) for each experiment and interpretation of error bars are given in the figure legends. Parameters were estimated by Markov chain Monte Carlo (MCMC) sampling from their respective posterior distributions; numerical values of parameter medians and credible intervals are listed in Table S3.

Implementation and software

The CycleFlow model was simulated using the Python programming language v.3.8.5 with the package Scipy v.0.13.0 and 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 v.4.0.2 to produce graphical representations.

RNA-seq data count matrices were generated using CellRanger v3.0.1 and velocyto.py v0.17.17. The data was then analyzed with R v4.0.2 and the following R packages: DropletUtils v1.10, scater v1.18.6, scran v1.18.7, Seurat v3 and velociraptor v1.0.0.

Cell Reports Methods, Volume *2*

Supplemental information

CycleFlow simultaneously quantifies cell-cycle phase lengths and quiescence *in vivo*

Adrien Jolly, Ann-Kathrin Fanti, Csilla Kongsaysak-Lengyel, Nina Claudino, Ines Gräßer, Nils B. Becker, and Thomas Höfer

Challenge	Addressed by
Cells labeled at any stage of S phase.	Subphases track progress within cycle phases.
Variable cycle progression speed.	Distributions for phase durations.
Ambiguity between G_0 and G_1.	Model states are unambiguous cycle phases, not gates.
EdU is consumed gradually.	Decrease in labeling efficiency is fitted from data.
DNA content reflects cycle phase with limited accuracy.	Cells at boundaries of DNA staining gates assigned to cycle phases via model fit to data.

Table S1 | Model features introduced to address the main challenges in the interpretation of thymidine analog labeling, Related to STAR Methods.

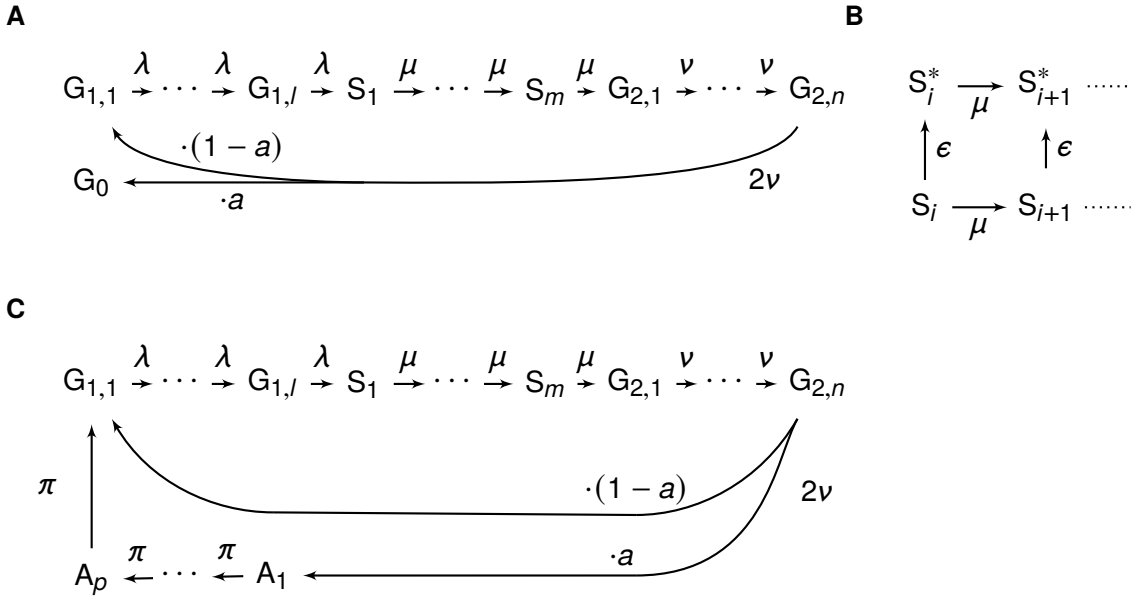


Figure S1 | Dynamical models, Related to STAR Methods. **A** Cells progress through the subphases of G_1, S and G_2 phase. The average and variability of the time to complete each cycle phase is governed by the progression rate and substep number. At the end of the cycle, cells divide. With probability a, daughter cells arrest and enter state G_0; otherwise they continue the cycle. From any cycle state, cell may be lost at rate δ (not shown). **B** Cells in subphase S_i can acquire the label to end up in state S^*i and transition to the next subphase. **C** Extended cell cycle model, not used in CycleFlow. Upon division, cells enter arrest with probability a. The arrested phase A is subdivided into subphases $A_1, \ldots, A_p$, after which cells return into the cycle at $G_{1,1}$. The remaining transitions are as in **A**.

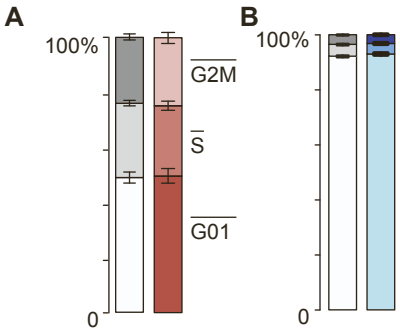


Figure S2 | Measured and inferred steady-state cell-cycle fractions based on DNA content, Related to Figure 2. **A** TET21N, experimental data averaged over all time points (left diagram, error bars indicate standard error of the mean) vs Model fit (right diagram, error bars indicate 90% credible interval) **B** same as **A** for DP thymocytes

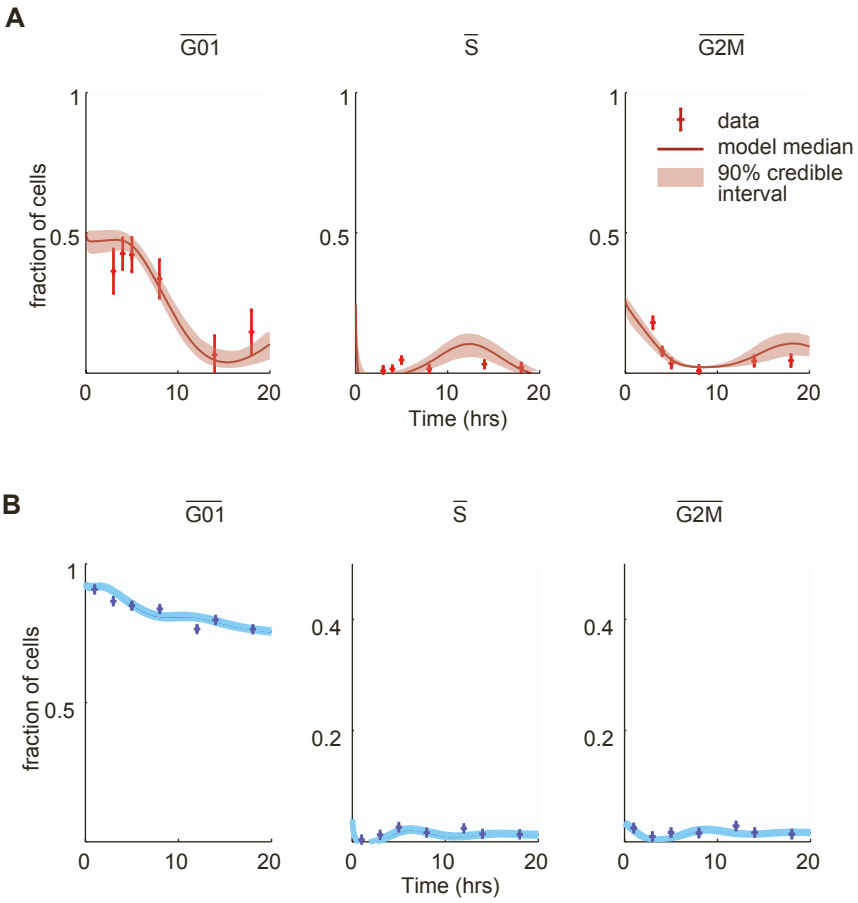


Figure S3| Time courses of EdU-unlabeled cells in $\overline{\text{G01}}$, $\overline{\text{S}}$, and $\overline{\text{G2M}}$ gates compared between experimental data and model prediction, Related to Figure 2. A TET21N and **B** DP cells (error bars, pooled SEM; n = 3 to 6 per time point) and model fit. Population sizes are given as fractions of total cells.

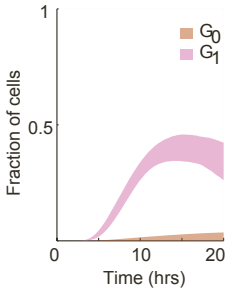


Figure S4| Inferred time course of EdU-labeled G_0 and G_1 TET21N cells in the $\overline{\text{G01}}$ gate, Related to Figure 2. Population sizes are given as fractions of total cells.

Parameter	Description	Unit	Prior range
λ	g_1 subphase progress rate	h^{-1}	$[0.01, 6]$
μ	S subphase progress rate	h^{-1}	$[1, 5]$
ν	G_2 subphase progress rate	h^{-1}	$[1, 10]$
a	Cycle arrest probability		$[0, 1]$
τ_E	EdU degradation time	h	$[0.5, 5]$
ϵ_0	initial labeling rate	h^{-1}	$[1, 15]$
l	g_1 subphase number		$[2, 30]$
m	S subphase number		15, fixed
n	G_2 subphase number		15, fixed
$m_{\overline{01}}$	S subphases assigned to $\overline{\mathsf{G01}}$		$[0, 5]$
$m_{\overline{2}}$	S subphases assigned to $\overline{\mathsf{G2M}}$		$[0, 5]$

Table S2 | Model parameters with allowed ranges, Related to STAR Methods.

Parameter	Description	Unit	TET21N	DP
λ	g_1 subphase progress rate	h^{-1}	$3.6(1.7, 4.9)$	$0.49(0.4, 0.6)$
μ	S subphase progress rate	h^{-1}	$2.3(1.8, 2.9)$	$3.7(3.1, 4.2)$
ν	G_2 subphase progress rate	h^{-1}	$2.8(2.6, 3.2)$	$7.5(6.9, 8.8)$
a	Cycle arrest probability		$0.016(0.002, 0.046)$	$0.42(0.415, 0.43)$
τ_E	EdU degradation time	h	$2.6(2, 4)$	$1(0.8, 1.3)$
ϵ_0	initial labeling rate	h^{-1}	$8.6(3, 14.3)$	$5(2.2, 14)$
l	g_1 subphase number		$24(11, 30)$	$2(2, 2)$
$m_{\overline{01}}$	S subphases assigned to $\overline{\mathsf{G01}}$		$2(0, 3)$	$4(3, 5)$
$m_{\overline{2}}$	S subphases assigned to $\overline{\mathsf{G2M}}$		$1(0, 3)$	$0(0, 1)$
l/λ	g_1 duration	h	$6.6(5.5, 7.9)$	$4(3.3, 4.9)$
m/μ	S duration	h	$6.4(5.2, 8.2)$	$4(3.5, 4.8)$
n/ν	G_2 duration	h	$5.3(4.6, 5.8)$	$2(1.7, 2.1)$
$2a$	G_0 fraction		$0.03(0, 0.09)$	$0.84(0.83, 0.86)$

Table S3 | Model parameters medians and 90% credible intervals, Related to STAR Methods.