



OPEN ACCESS

Edited by:

Martin Pera,
Jackson Laboratory, United States

Reviewed by:

Stephan Daetwyler,
University of Texas Southwestern
Medical Center, United States
Gennady Bocharov,
Institute of Numerical Mathematics
(RAS), Russia
Vitaly V. Ganusov,
the University of Tennessee,
United States

***Correspondence:**

Ken R. Duffy
Ken.Duffy@mu.ie
Philip D. Hodgkin
hodgkin@wehi.edu.au

[†]These authors have contributed
equally to this work

[‡]Present address:

Giulio Prevedello,
Sony Computer Science Laboratories,
Paris, France
Simone C. Oostindie,
Genmab, Utrecht, Netherlands

Specialty section:

This article was submitted to
Computational Biolmaging,
a section of the journal
Frontiers in Bioinformatics

Received: 10 June 2021

Accepted: 28 September 2021

Published: 26 October 2021

Citation:

Cheon H, Kan A, Prevedello G,
Oostindie SC, Dovedi SJ, Hawkins ED,
Marchingo JM, Heinzl S, Duffy KR and
Hodgkin PD (2021) Cyton2: A Model of
Immune Cell Population Dynamics
That Includes Familial
Instructional Inheritance.
Front. Bioinform. 1:723337.
doi: 10.3389/fbinf.2021.723337

Cyton2: A Model of Immune Cell Population Dynamics That Includes Familial Instructional Inheritance

HoChan Cheon^{1†}, Andrey Kan^{2,3}, Giulio Prevedello^{1‡}, Simone C. Oostindie^{2,3‡}, Simon J. Dovedi⁴, Edwin D. Hawkins^{3,5}, Julia M. Marchingo⁶, Susanne Heinzl^{2,3}, Ken R. Duffy^{1*†} and Philip D. Hodgkin^{2,3*†}

¹Hamilton Institute, Maynooth University, Maynooth, Ireland, ²Immunology Division, the Walter and Eliza Hall Institute of Medical Research, Parkville, VIC, Australia, ³Department of Medical Biology, the University of Melbourne, Parkville, VIC, Australia, ⁴Oncology R&D, AstraZeneca, Cambridge, United Kingdom, ⁵Division of Inflammation, the Walter and Eliza Hall Institute of Medical Research, Parkville, VIC, Australia, ⁶Cell Signalling and Immunology Division, School of Life Sciences, University of Dundee, Dundee, United Kingdom

Lymphocytes are the central actors in adaptive immune responses. When challenged with antigen, a small number of B and T cells have a cognate receptor capable of recognising and responding to the insult. These cells proliferate, building an exponentially growing, differentiating clone army to fight off the threat, before ceasing to divide and dying over a period of weeks, leaving in their wake memory cells that are primed to rapidly respond to any repeated infection. Due to the non-linearity of lymphocyte population dynamics, mathematical models are needed to interrogate data from experimental studies. Due to lack of evidence to the contrary and appealing to arguments based on Occam's Razor, in these models newly born progeny are typically assumed to behave independently of their predecessors. Recent experimental studies, however, challenge that assumption, making clear that there is substantial inheritance of timed fate changes from each cell by its offspring, calling for a revision to the existing mathematical modelling paradigms used for information extraction. By assessing long-term live-cell imaging of stimulated murine B and T cells *in vitro*, we distilled the key phenomena of these within-family inheritances and used them to develop a new mathematical model, Cyton2, that encapsulates them. We establish the model's consistency with these newly observed fine-grained features. Two natural concerns for any model that includes familial correlations would be that it is overparameterised or computationally inefficient in data fitting, but neither is the case for Cyton2. We demonstrate Cyton2's utility by challenging it with high-throughput flow cytometry data, which confirms the robustness of its parameter estimation as well as its ability to extract biological meaning from complex mixed stimulation experiments. Cyton2, therefore, offers an alternate mathematical model, one that is, more aligned to experimental observation, for drawing inferences on lymphocyte population dynamics.

Keywords: immune response, population dynamics, mathematical model, familial correlations, proliferation

1 INTRODUCTION

B and T lymphocytes are central contributors to the adaptive immune response. When exposed to a foreign pathogen with epitopes that are complementary to their B or T cell receptors, they respond by proliferating to create a clone army capable of recognising the threat. These cells differentiate into effector cells to fight the invasion, and into memory cells primed to fend off repeated insults. The population size of their response, the proportion of cells allocated to distinct differentiated effector types, the cytokines that they produce, and other key characteristics of the immune response are known to be heterogeneous but regulable (Kaech et al., 2002; Duffy et al., 2012; Buchholz et al., 2013; Gerlach et al., 2013). Variables that influence the outcome include the affinity of the receptor interaction and the provision of costimulatory signals from other cells (Marchingo et al., 2014). In the quest to better understand immune responses and therapeutic intervention, it remains an essential question to determine how signals are integrated to alter cell fate and how the cells process such information to yield diverse, yet appropriate outcomes. Answering this question requires an understanding of operational aspects of lymphocyte population dynamics, and the influence of signals on individual fates. When known, quantitative models and analytical techniques can be developed and used to monitor lymphocyte control under different conditions; they can recreate, and predict outcomes for complex situations (Duffy et al., 2012; Hodgkin, 2018).

Much of the understanding regarding lymphocyte population dynamics has come from assessing *in vitro* experiments. When isolated *ex vivo*, B and T cells are small, non-dividing, resting cells that die after a period of time if placed unstimulated into culture. The provision of activating signals leads to changes that reprogramme survival times and initiate cell division in quantitative manner (Gett and Hodgkin, 2000; Hawkins et al., 2007). After an initial period of intense transcriptional changes and cellular programming, activated cells initiate and undergo division repeatedly, before their offspring return to a non-dividing, quiescent state followed ultimately by death if no further signals, such as from cytokines, are received. Thus mathematical models for immune dynamics must have features that match biological processes and allow the alteration of division times, the number of cell divisions, the likelihood of cell death, and rules for how these parameters are altered by changes in signalling conditions.

Advances in experimental technologies have provided detailed data on lymphocyte population dynamics that have informed modelling frameworks. A key development came in 1994 with the discovery that cell divisions could be followed and enumerated by flow cytometry with fluorescent dye carboxyfluorescein diacetate succinimidyl ester (CFSE) (Lyons and Parish, 1994), with subsequent developments deriving distinct colours (Quah and Parish, 2012) including CellTrace™ Violet (CTV). After a short period of culture with these dyes cells become intensely fluorescent and measurable by flow cytometry. On division, their offspring inherit half their parent's dye and so fluoresce with half their intensity. That methodology allows up to eight

distinct generations to be measurable within a single culture by flow cytometry before fluorescence falls to a level indistinguishable from background. Data from CFSE and CTV experiments informed, for example, the mathematical models reported in Gett and Hodgkin (2000), Boer and Perelson (2005), Ganusov et al. (2005), Asquith et al. (2006), Hawkins et al. (2007), Luzyanina et al. (2007), Subramanian et al. (2008), Duffy and Subramanian (2009), Hyrien and Zand (2008), Zilman et al. (2010), Banks et al. (2011), Miao et al. (2011), Banks et al. (2012), Shokhirev and Hoffmann (2013), Mazzocco et al. (2017). Many of these models either ignore cell survival or assume that it is a fixed feature that is, independent of the age of cells. Most of these models also assume age independent division times to make stochastic systems Markovian or consider only the evolution of the average system, expressed as ordinary differential equations.

In contrast, directly performing novel experiments for the goal of mathematical model design, Hawkins et al. (2007) measured survival over time and concluded cell age was important to their fate. They also extended earlier work of Gett and Hodgkin (2000) that demonstrated that division and death times could be regulated independently within the same cell. Based on those data, they proposed a model where cell age and stochastic operations govern fate outcomes. Their Cyton model of the cell was named for the putative molecular machinery creating regulable timers for division and death.

In the Cyton model, division and death times are heterogeneous in the cell population and so modelled by random variables whose operation appears independent. Within each cell, the two timers are in competition, where whichever one completes its operation first determines the fate of the cell. This model structure gives rise to the prediction of distinctive correlations that are observed in data (Duffy and Hodgkin, 2012). In the absence of detailed information on individual cells and their offspring, the Cyton model assumed that timers were independently reset at each generation. To complete the Cyton model, an additional component was introduced: the number of divisions cells underwent before cessation of expansion and their return to quiescence. This parameter, termed division destiny (DD), was described by a probability of continuing motivation to divide after each cell division.

Thus, in the Cyton model a cell would divide rapidly for a period when division times outcompeted death times. The fate of a cell that stops dividing by triggering division destiny is then solely governed by its final death time. By adjusting the probability distributions of division, death and destiny, the model recreated typical immune cell population dynamics without further *ad hoc* assumptions (Hawkins et al., 2007; Subramanian et al., 2008; Lee et al., 2009; Wellard et al., 2011). After its development, the Cyton model was successfully used as a tool in important studies that extracted information on key features controlling immune dynamics (Hawkins et al., 2013; Shokhirev and Hoffmann, 2013; Marchingo et al., 2014; Shokhirev et al., 2015; Mitchell et al., 2018). Some of the assumptions on which the Cyton model was based were unobserved facets, and needed further experimental

confirmation for their suitability. In particular, questions of familial correlation needed to be addressed by time-lapse microscopy and other, similarly capable, methods.

Stimulated lymphocytes typically aggregate, adhering together, making individual cell tracking by microscopy difficult or impossible. However, Hawkins et al. (2009) noted that B cells stimulated by the Toll-like receptor agonist CpG DNA exhibited the population dynamics typical of standard immune responses, but remained separated and individually identifiable (Hawkins et al., 2009). Using microscopy, the authors tracked over 180 individual family trees enabling statistical features such as dependencies to be assessed. Strikingly, it became apparent that division and death times of siblings were highly correlated. Further, division destiny, the number of divisions cells undergo before returning to quiescence, was a strongly familial feature (Hawkins et al., 2009). This conclusion, which ran contrary to assumptions underlying all previous mathematical models, was examined and further extended in subsequent studies Duffy and Subramanian (2009); Markham et al. (2010); Wellard et al. (2010); Duffy et al. (2012); Dowling et al. (2014); Shokhirev et al. (2015); Mitchell et al. (2018). In a parallel development, a division dye multiplex method, which provides less lineage information than live cell imaging, but has higher throughput for identifying families, was developed (Marchingo et al., 2016; Horton et al., 2018). When used with antigen stimulated CD8⁺ T cells, similar familial features to those observed directly for B cells were reported.

In addition to those population dynamics studies, the proto-oncogene Myc was identified as a molecular correlate that explained one important aspect of familial sharing of information. Results in Heinzel et al. (2017) established that in B and T cells Myc levels increase in response to mitogenic stimuli, and, so long as levels are sustained above a critical threshold, Myc acts as a license for cells to divide. Over the course of the response Myc levels then fall, and once they drop below the threshold, these cells lose their motivation for further division and re-enter quiescence. Crucially, that experimental work established that the time between cell divisions was uncoupled from the Myc level. Further, importantly, Myc levels altered over time, diminishing late in culture, but the kinetics of change were transmitted to offspring without being affected by mitosis. Taken together, these results indicate that the control of division destiny should be viewed as being timed, rather than counted by cell division (Heinzel et al., 2017). The familial inheritance of division destiny was consistent with the high correlations in fate within clonal families that were reported for both B and T cells (Hawkins et al., 2009; Duffy et al., 2012; Marchingo et al., 2016; Horton et al., 2018; Zhou et al., 2018). Heinzel et al. (2017) also reported evidence that time to death under these conditions was also programmed early in the stimulated cell and passed to descendants without being altered in a manner analogous to the transmission of the division destiny times. As a result, the fate of whole family members can be highly concordant while allowing significant variation of the times between families from an otherwise homogeneous cell population.

Collectively, these findings suggest alterations to current model paradigms are necessary. While the Cyton model was

correct in its assessment of competing timers, assigning them to families rather than individual cells is more consistent with these data. Here, we propose and develop a new Cyton model where familial inheritance of times for destiny and survival fates are included. We examine datasets from time-lapse microscopy of B and CD8⁺ T cell families, and interrogate these data to investigate consistency with timed outcomes. We measure correlations in the likelihood of each alternative fate and determine a suitable class of parametric distributions for their description. The proposed model is constructed such that identifiability is improved while computational model fitting burden over the earlier Cyton model is not increased. We use the model to interrogate CTV stained datasets obtained using flow cytometry, illustrating its utility and efficacy when used with both B and T cells.

2 RESULTS

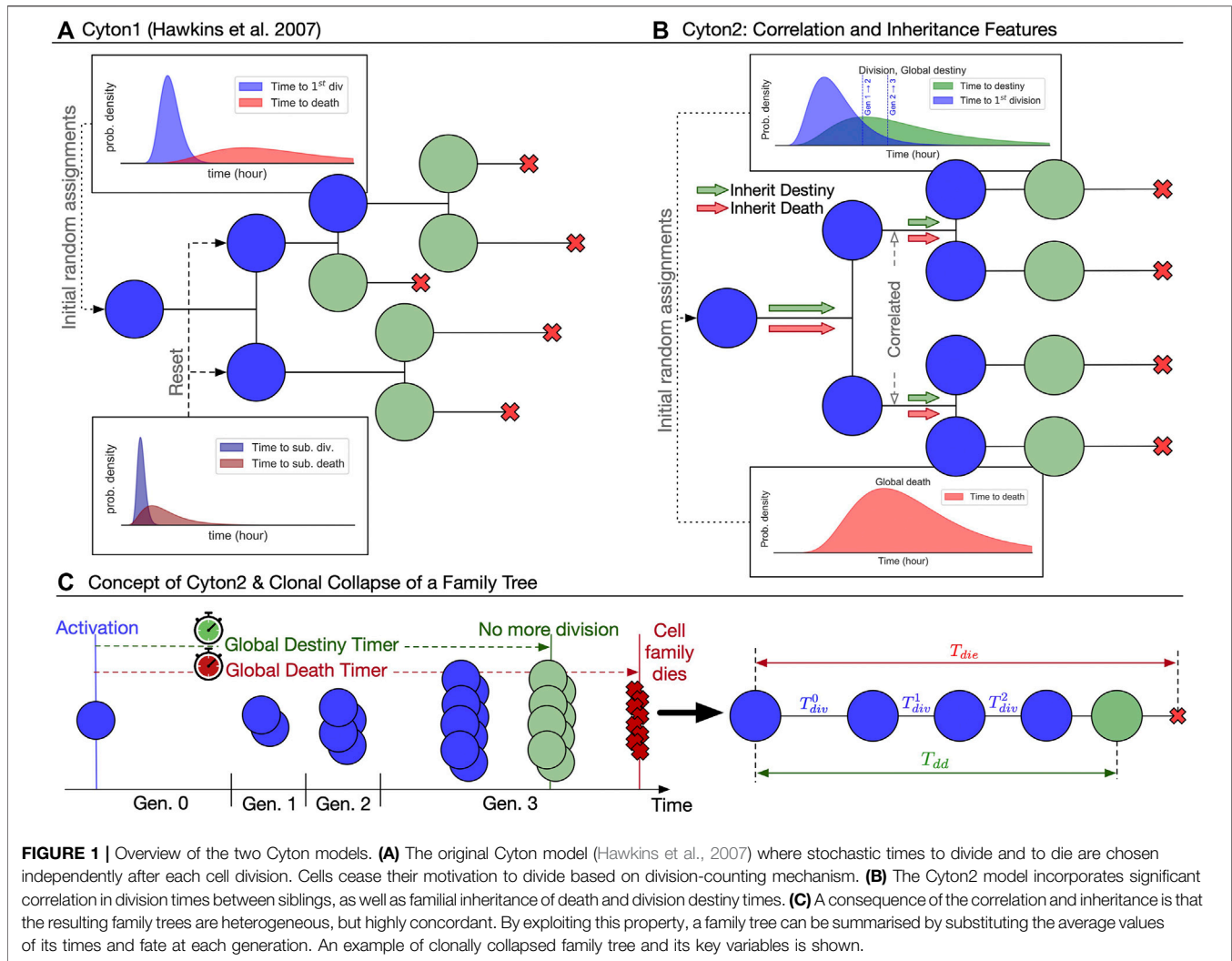
2.1 Cyton2 Model Structure

Recent experimental findings suggest three important modifications to the original Cyton model for stimulus-induced proliferation bursts (**Figure 1A**): 1) division destiny should be converted to a division-agnostic, family-based timed mechanism, replacing the original generation counter; 2) both division destiny and death times should be programmed early after each lymphocyte's activation and applied globally to the ancestor's offspring; and 3) family members of the same generation should have essentially the same division time. As has been observed experimentally (Hawkins et al., 2009; Marchingo et al., 2016; Horton et al., 2018; Mitchell et al., 2018), the resulting family trees of activated lymphocytes derived from a single founder cell, and hence clones, according to Cyton2 rules are largely regular (**Figure 1B**). Thus we posit the new Cyton2 stochastic model using sets of random variables that correspond to a global death timer, a global destiny timer, and division-time machinery (**Figure 1C**).

The development of each family tree in Cyton2 is fully described by a collection of independent, non-negative, real-valued random variables: $(T_{div}^0, \{T_{div}^k\}_{k \geq 1}, T_{dd}, T_{die})$. Three of these describe times from the addition of stimulus: the time to first division T_{div}^0 ; the time to familial division destiny T_{dd} , encapsulating the licence to divide period; and the time to familial death T_{die} . The last set of random variables, $\{T_{div}^k\}_{k \geq 1}$ are the times from each mitosis to the next, should it complete before division destiny or death occurs. From these random variables, a family tree is created according to the following rules:

- Founding cells that give rise to familial clones are initially quiescent, unrelated and autonomous.
- All cells in the family die at T_{die} .
- The family proliferates until $\min(T_{die}, T_{dd})$.
- At time $t < \min(T_{die}, T_{dd})$, cells in the family are in generation $G(t) = \max\{g: \sum_{k=0}^g T_{div}^k < t\}$.

To properly assess the appropriateness of the Cyton2 as a fine-grained description required time-lapse microscopy data. To that



end we re-analysed previously published B cell data sets as well as new, primary CD8⁺ T cell datasets.

2.2 Time-Lapse Microscopy of B and T Cell Families

For B cells, we revisited two datasets for CpG-stimulated B cells published in Hawkins et al. (2009) consisting of 108 clones (B-exp1) and 88 clones (B-exp2), respectively. These datasets had not been analysed for timed global features but had revealed strong familial correlations previously (Duffy and Subramanian, 2009; Hawkins et al., 2009; Markham et al., 2010; Wellard et al., 2010). Thus, to explore familial features we first transformed the data for each family, collapsing the tree into average features (see Methods; **Supplementary Table S1, Supplementary Figures S1, S2** for raw data). This process is illustrated in **Figure 1C** and was applied to each B cell family as shown in **Figure 2A1**. Measurements corresponding to key Cyton2 variables are further illustrated in the cascade plots **Figure 2A2** with the exception of the time to division destiny (T_{dd}) as it cannot be identified directly in data. Instead, the time to last

division (T_{ld}), which is necessarily a lower bound, was used as a proxy for it. These data reconfirmed the well-established understanding that times to first division, ≈ 40 h, are substantially longer than times to subsequent divisions. These data also confirmed the relatively consistent subsequent division times (≈ 10 h) and the strong correlation times between progeny cells within a given generation in each family.

Using these measurements, we evaluated the discrepancy between Cyton2's approximation of perfect within-family correlation in subsequent division time (T_{div}^k), time to last division (T_{ld}), and time to death (T_{die}). We calculated coefficient of variation (CV) per clone, and evaluated the average CV for each variable. For T_{div}^k , T_{ld} and T_{die} , we identified 17.2, 7.3, and 9.4%, respectively, as average CVs for B-exp1. Similar results were found for B-exp2 (see **Supplementary Figure S3A1**). This signifies low variation around the mean times to fates within families, and is consistent with previously reported synchronous behaviour. We then questioned the independence of the variables operating at the clone level using information from the

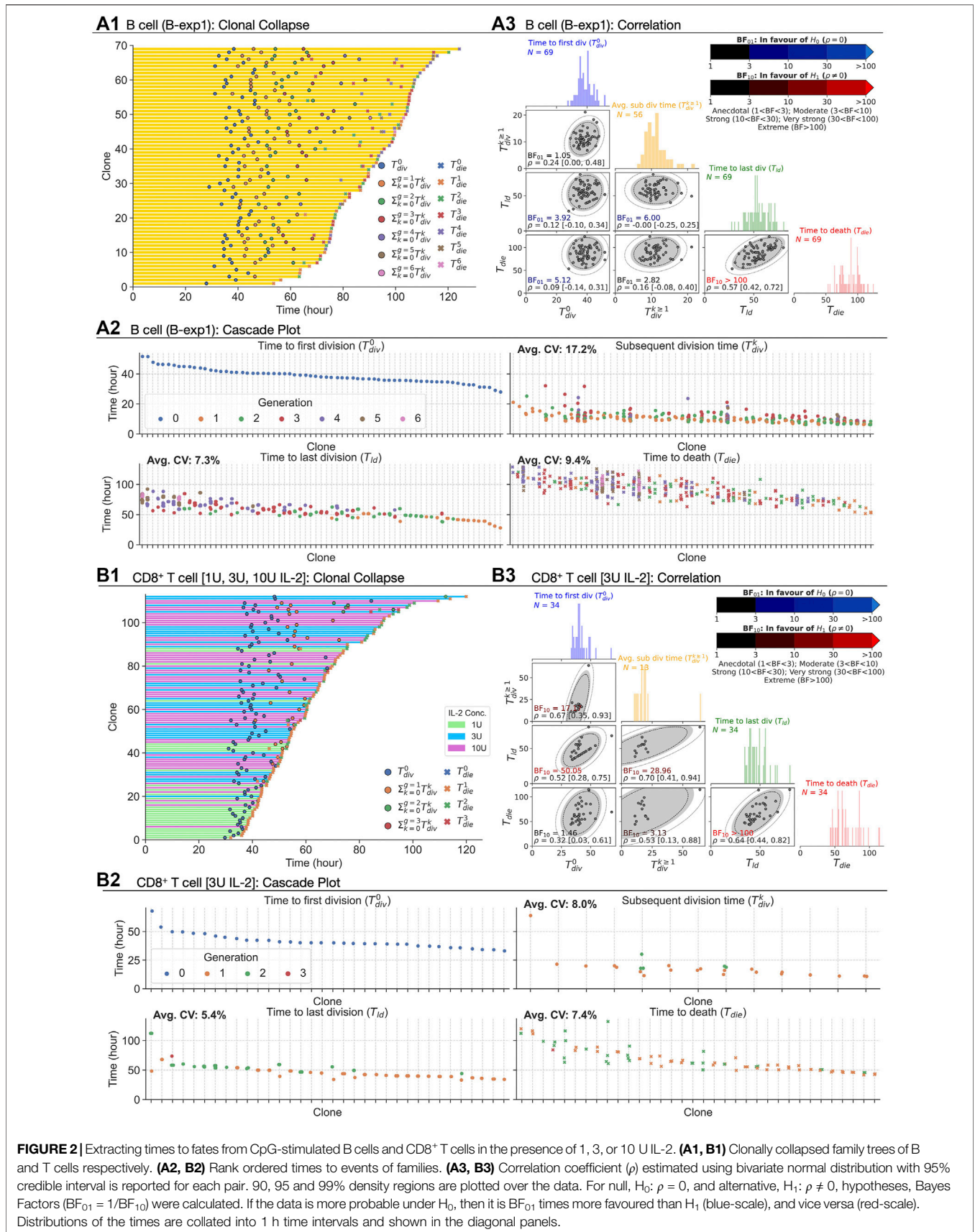


TABLE 1 | Bayesian independence test of the times to fates extracted from B and CD8⁺ T cell filming datasets. For each pair of the times to fates, the correlation coefficient was estimated with 95% credible interval using bivariate normal distribution and Bayes Factor (BF) was calculated. Given two hypotheses ($H_0: \rho = 0$ and $H_1: \rho \neq 0$), if the data is more probable under H_0 , then it is BF_{01} times more favoured than H_1 , otherwise H_1 is BF_{10} times more favoured than H_0 .

Cell type	Stim	Number of clones (N) and bayes Factor ^a ($BF_{01} = 1/BF_{10}$) and correlation coefficient [ρ (CI)]						
		$(T_{div}^0, T_{div}^{k \geq 1})$	(T_{div}^0, T_{ld})	(T_{div}^0, T_{die})	$(T_{div}^{k \geq 1}, T_{ld})$	$(T_{div}^{k \geq 1}, T_{die})$	(T_{ld}, T_{die})	
B (B-exp1)	CpG	$N = 56$ $BF_{01} = 1.05$ 0.24 [0.01, 0.48]	$N = 69$ $BF_{01} = 3.92$ 0.12 [-0.10, 0.35]	$N = 69$ $BF_{01} = 5.12$ 0.09 [-0.14, 0.31]	$N = 56$ $BF_{01} = 6.00$ 0.0 [-0.25, 0.25]	$N = 56$ $BF_{01} = 2.82$ 0.16 [-0.08, 0.40]	$N = 69$ $BF_{10} > 100$ 0.57 [0.42, 0.72]	
		CD8 ⁺ T	1U IL-2	$N = 4$ $BF_{01} = 1.70$ -0.01 [-0.86, 0.85]	$N = 28$ $BF_{10} = 39.09$ 0.55 [0.29, 0.78]	$N = 28$ $BF_{01} = 2.53$ 0.19 [-0.15, 0.53]	$N = 4$ $BF_{01} = 1.53$ -0.15 [-0.96, 0.71]	$N = 4$ $BF_{01} = 1.43$ 0.20 [-0.66, 0.98]
		3U IL-2	$N = 13$ $BF_{10} = 17.17$ 0.67 [0.36, 0.93]	$N = 34$ $BF_{10} = 50.05$ 0.52 [0.28, 0.75]	$N = 34$ $BF_{10} = 1.46$ 0.32 [0.03, 0.61]	$N = 13$ $BF_{10} = 28.96$ 0.70 [0.41, 0.94]	$N = 13$ $BF_{10} = 3.13$ 0.52 [0.12, 0.87]	$N = 34$ $BF_{10} > 100$ 0.64 [0.44, 0.82]
		10U IL-2	$N = 16$ $BF_{10} = 7.55$ 0.57 [0.23, 0.86]	$N = 50$ $BF_{10} > 100$ 0.58 [0.41, 0.75]	$N = 50$ $BF_{10} = 15.39$ 0.40 [0.18, 0.62]	$N = 16$ $BF_{10} > 100$ 0.76 [0.55, 0.94]	$N = 16$ $BF_{10} = 1.56$ 0.42 [0.02, 0.79]	$N = 50$ $BF_{10} > 100$ 0.63 [0.46, 0.78]

^aBF Interpretation: Anecdotal ($1 < BF \leq 3$); Moderate ($3 < BF \leq 10$); Strong ($10 < BF \leq 30$); Very strong ($30 < BF \leq 100$); Extreme ($BF > 100$).

collapsed clones. Here, for statistical purposes, we extracted the time to first division, average subsequent division time ($T_{div}^{k \geq 1}$), average time to last division, and average time to death, as the four key variables per clone. For every pair, the correlation coefficient (ρ) and its 95% credible interval were determined using a Bayesian approach. For these data, the Bayes Factor (BF) for competing hypotheses ($H_0: \rho = 0$ and $H_1: \rho \neq 0$) were calculated (see Methods) (Figure 2A3) and tabulated in Table 1. With the exception of the pair (T_{ld}, T_{die}), CpG-stimulated B cells showed little to no correlation between any pair of variables, with H_0 being favoured. While at first glance the exception may appear suggestive of shared regulation, another explanation is possible, which is examined in the next section.

Extending the analysis to T cells, we also interrogated three primary data sets of time-lapse microscopy of murine CD8⁺ T cells not previously published. In each dataset, TCR transgenic OT-I CD8⁺ T cells specific for the SIINFEKL (N4) peptide from the chicken ovalbumin protein (Hogquist et al., 1994) were first stimulated with α CD3 or cognate peptide N4 along with a range of costimulatory signals and strengths for 24 h. In the first dataset (i) the cells were stimulated with α CD3 and co-incubated with 1, 3, or 10 U/mL of the T-cell growth factor IL-2. IL-2 level was buffered by neutralising endogenously produced IL-2 with blocking antibody S4B6, and adding human IL-2 at the nominated concentration (Deenick et al., 2003). By combining datasets obtained from two independent repeats, 109, 90, and 163 clones were recorded. In (ii), the combination of N4, α CD28, and IL-2 were used (T-exp1); and, in experiment (iii) the combination of N4, α CD28 and IL-12 (T-exp2). Details for live imaging and data extraction are given in Methods. In Figure 2B, results from CD8⁺ T cell dataset (i) are aggregated and analysed as for B cells. Similar to B cells, we observed longer times to first division (≈ 40 h) than the subsequent division times (≈ 18 h) for 1, 3, and 10 U of IL-2. Also, the spread of the times within a family show similar or lower average CVs than that of B cells (Figure 2B2 for 3 U; see Supplementary Figures S3B1,C1 for 1 and 10 U). We applied the same calculation to (ii) and (iii) datasets and reached the same conclusions (see Supplementary Figures S4A,B for T-exp1

and T-exp2). Taken together, we conclude CD8⁺ T cells exhibit synchronous fates, similar to observations from B cells. However, in contrast to B cells, moderate to strong correlation coefficients were observed (Figure 2B3). These were further supported by BF calculations, which show strong evidence in favour of H_1 (Table 1). We noticed the same results for T-exp1 and T-exp2 datasets (see Supplementary Table S2).

At face value, as with the pair (T_{ld}, T_{die}) for B cells, these data are suggestive of a lack of stochastic independence between underlying timers. An alternate explanation is, however, possible and we next sought to challenge it.

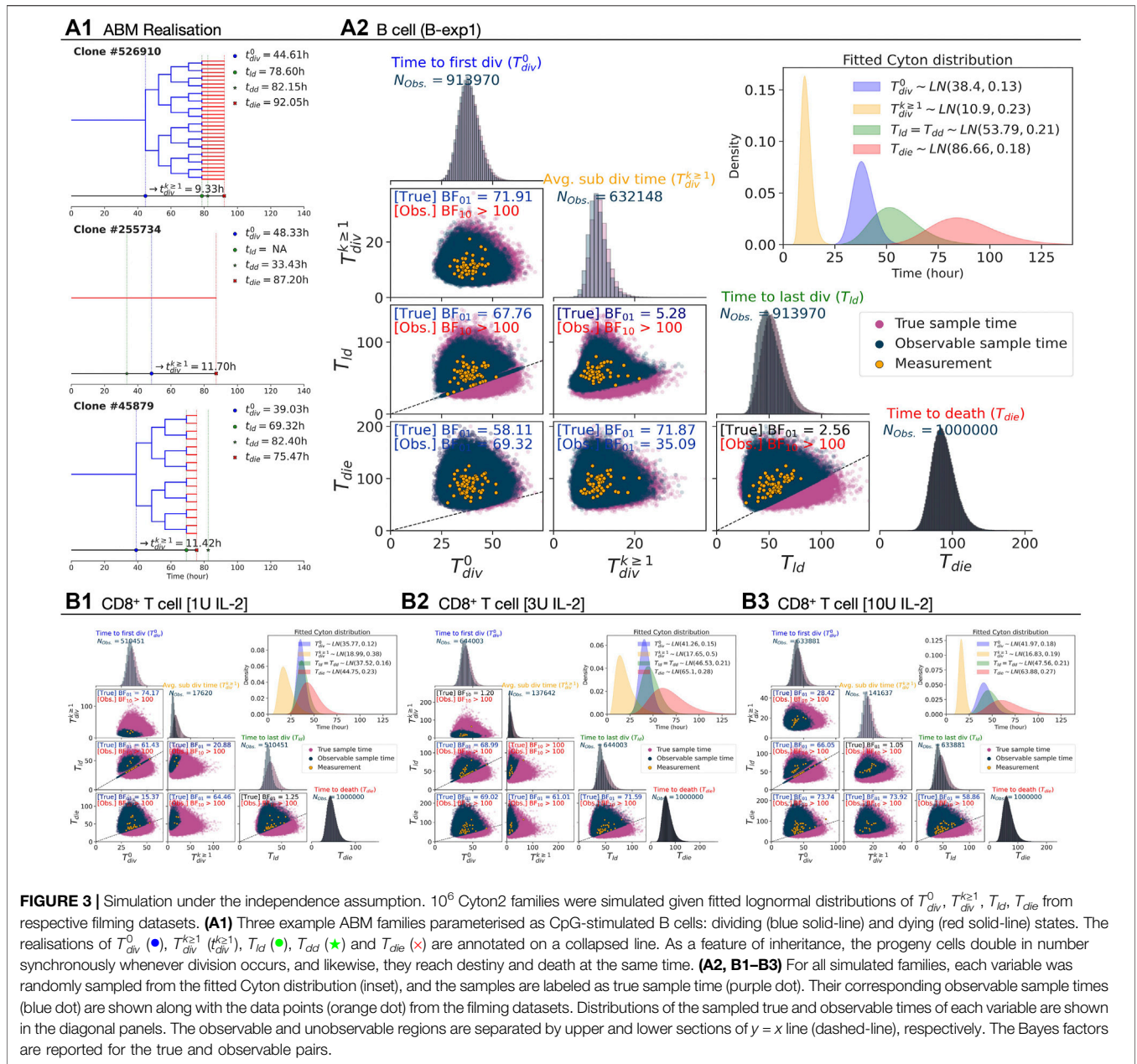
2.3 Induced Dependency Through Right Censoring of Timers

Informed by earlier data, in constructing Cyton2 we assumed that $(T_{div}^0, \{T_{div}^k\}_{k \geq 1}, T_{dd}, T_{die})$ were independent random variables describing times to familial events. In the data, however, not all of them are observable due to a phenomenon called right-censoring. In particular:

- If T_{div}^0 or T_{div}^k is greater than either of T_{dd} or T_{die} , it is not observed in the data.
- If T_{dd} is greater than T_{die} , it is not observed in the data.

Even if the underlying random variables are independent, right-censoring necessarily induces correlation in times observed in data (Duffy et al., 2012; Duffy and Hodgkin, 2012) where the greater the competition in these times, the stronger the observed correlation. While these earlier demonstrations of censorship-induced correlations were seen within one generation, we explored the possibility that heritable fates times across multiple generations could also lead to a similar effects.

In Section 2.2, most of the variable pairs for B cell families were reported to be more probable under the no-correlation hypothesis, while for the CD8⁺ T cell families we found mixed results. The key difference between the two datasets is the depth



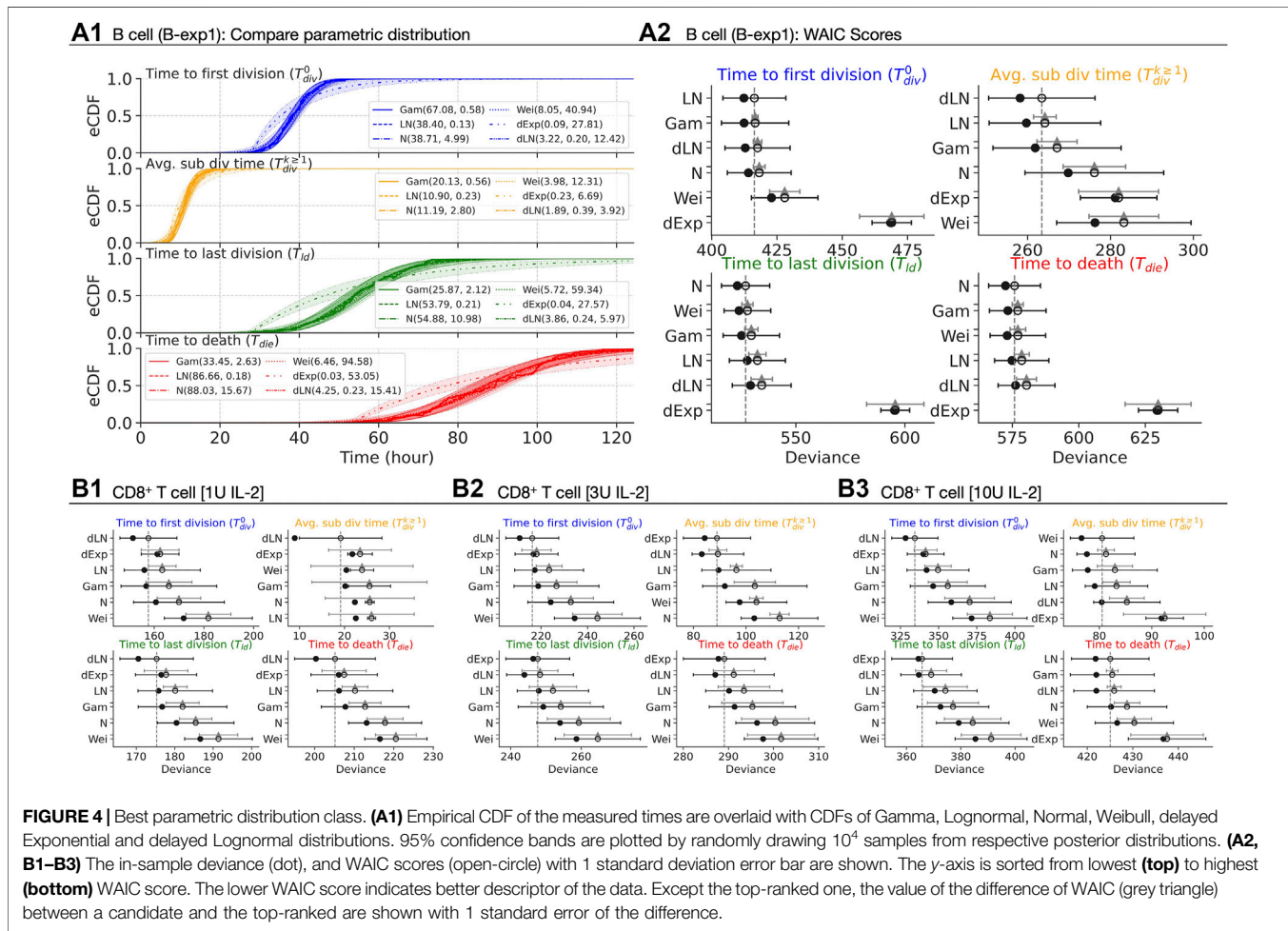
of the trees: many of the B cell families had divided six times, whereas the CD8⁺ T cell families had divided at most three times (Figures 2A2,B2). This suggests the possibility that more of the variables are rendered unobserved for the CD8⁺ T cell families. To challenge that possible explanation, we simulated a Cyton2 process with an agent based model (ABM) (see Methods). As T_{dd} is not directly observable in data, we use T_{ld} as a proxy for it. As $T_{dd} \geq T_{ld}$, this approximation may lead to an increase in the level of induced censorship. Under the assumption that $T_{ld} = T_{dd}$, each variable was independently sampled from respective lognormal distributions that were fit to the data (see Section 2.4). In Figure 3A1, three example realisations of family trees are shown for a parameterisation corresponding to CpG-stimulated B cells.

For B cells, with each point representing a single family, the underlying simulated variable values as well as those that would appear in the data due to the right-censorship described above are shown in Figure 3A2. By construction, the BFs for each pair in the underlying timers favour the null hypothesis ($H_0: \rho = 0$). For the right-censored values that would be observed in practice, however, the BFs are consistent with the experimental data in favouring the alternative hypothesis ($H_1: \rho \neq 0$) for some pairs (Table 2). As the underlying T_{die} distribution is well separated from the T_{div}^0 distribution for these data, it is unlikely that death would censor the time to first division, hence, it explains why the absence of correlation in the observed data is favoured in that case. Similar results were found for B-exp2 (data not shown). We followed a similar protocol for the CD8⁺ T cell data where a

TABLE 2 | Bayesian independence test of the times to fates from simulation. The test was performed with $N_{True} = 10^6$ simulated families via Cyton2-like Agent-Based Model. Each family was assigned randomly sampled times (True), and corresponding observable (Obs.) times were recorded. Depending on the order of the true times to fates, the number of observable times (N_{Obs}) may vary. For both true and observable times, Bayes Factor (BF) was calculated given null and alternative hypotheses ($H_0: \rho = 0$ and $H_1: \rho \neq 0$). Here BF_{01} indicates the simulated data are more probable under H_0 , otherwise it is indicated by BF_{10} .

Cell type	Stim	Percentage of number of observable times (N_{Obs}/N_{True}) and bayes Factor ^a ($BF_{01} = 1/BF_{10}$)											
		$(T_{div}^0, T_{div}^{k \geq 1})$		(T_{div}^0, T_{id})		(T_{div}^0, T_{die})		$(T_{div}^{k \geq 1}, T_{id})$		$(T_{div}^{k \geq 1}, T_{die})$		(T_{id}, T_{die})	
		True	Obs	True	Obs	True	Obs	True	Obs	True	Obs	True	Obs
B (B-exp1)	CpG	$BF_{01} = 71.91$	63.2%	$BF_{01} = 67.76$	91.4%	$BF_{01} = 58.11$	91.4%	$BF_{01} = 5.28$	63.2%	$BF_{01} = 71.87$	63.2%	$BF_{01} = 2.56$	91.3%
			$BF_{10} > 100$		$BF_{10} > 100$		$BF_{10} > 100$		$BF_{10} > 100$		$BF_{10} = 35.09$		$BF_{10} > 100$
CD8 ⁺ T	1U IL-2	$BF_{01} = 74.17$	17.6%	$BF_{01} = 61.43$	51.0%	$BF_{01} = 15.37$	51.0%	$BF_{01} = 20.88$	17.6%	$BF_{01} = 64.46$	17.6%	$BF_{01} = 1.25$	51.0%
			$BF_{10} > 100$		$BF_{10} > 100$		$BF_{10} > 100$		$BF_{10} > 100$		$BF_{10} > 100$		$BF_{10} > 100$
	3U IL-2	$BF_{10} = 1.20$	13.8%	$BF_{01} = 68.99$	64.4%	$BF_{01} = 69.02$	64.4%	$BF_{01} = 13.8\%$	13.8%	$BF_{01} = 61.01$	13.8%	$BF_{01} = 71.59$	64.4%
			$BF_{10} > 100$		$BF_{10} > 100$		$BF_{10} > 100$		$BF_{10} > 100$		$BF_{10} > 100$		$BF_{10} > 100$
	10U IL-2	$BF_{01} = 28.42$	14.1%	$BF_{01} = 66.05$	63.2%	$BF_{01} = 73.74$	63.2%	$BF_{01} = 1.05$	14.1%	$BF_{01} = 73.92$	14.1%	$BF_{01} = 58.86$	63.2%
			$BF_{10} > 100$		$BF_{10} > 100$		$BF_{10} > 100$		$BF_{10} > 100$		$BF_{10} > 100$		$BF_{10} > 100$

^aBF Interpretation: Anecdotal ($1 < BF \leq 3$); Moderate ($3 < BF \leq 10$); Strong ($10 < BF \leq 30$); Very strong ($30 < BF \leq 100$); Extreme ($BF > 100$).



higher degree of right-censorship occurs due to the underlying distributions having greater overlap (Figures 3B1–3). The BFs of all right-censored pairs were in favour of H_1 , indicating strong correlation between times to fates for CD8⁺ T cells are to be expected in the observed data as a result of high degree of right-

censorship. Thus, despite the temporal-correlations observed in the data, right-censorship supports our assumption in the Cyton2 model that the underlying stochastic variables are independent. This statistical conclusion complements the experimental evidence for fate independence obtained by slowing division

times and preventing cell death without altering other outcomes (Heinzel et al., 2017).

2.4 Using Filming Data to Determine Appropriate Distribution Classes for the Timers

In order to fit the model to commonly available non-microscopy data where direct observation of times is not possible, it is necessary to determine appropriate parametric distribution classes that well-capture the structure of the timers. Probability distributions governing the times to first division and to death for B cell cultures have been reported to be well approximated by a right-skewed distribution such as Lognormal, Weibull, Gamma, or Beta (Hawkins et al., 2009). In particular, the time to first division is known to be better described by a Lognormal rather than other skewed distributions, whereas Gamma or Weibull distribution can be used to approximate the time to death distribution (Hawkins et al., 2007).

For B cells, the empirical cumulative distribution function (eCDF) measured times overlaid with CDFs of four candidate distributions with 95% confidence bands in **Figure 4A1**. Each candidate is parameterised by: (i) (α_G, β_G) for Gamma; (ii) (m, s) for median and shape of Lognormal; (iii) (μ, σ) for mean and standard deviation for Normal; (α_W, β_W) for Weibull; (λ, c) rate and shift for delayed Exponential; and, (m_{db}, s_{db}, c) median, scale and shift for delayed Lognormal. Qualitatively, most of the candidates appear to be excellent descriptors for each of the measurements except for the delayed Exponential distribution for B cells. Here, we used the Widely Applicable Information Criterion (WAIC - see Methods) to quantitatively determine the best fit (**Figure 4A2**) (Watanabe and Opper, 2010). For T_{div}^0 in the B cell data, the Lognormal distribution was top-ranked (416.3) while delayed Exponential was least favoured (468.9, $\Delta SE = \pm 12.3$). The delayed Lognormal was most preferred candidate for $T_{div}^{k \geq 1}$ (263.4), but the Lognormal was a close second (264.2, $\Delta SE = \pm 2.7$). While the delayed Exponential was consistently worst fit for all measurements in B cells, the other five candidates well approximate T_{die} measurements as reported in previous studies. Interestingly, the Normal was favoured (526.9), or on par with the Weibull and Gamma, for T_{ld} as indicated by the standard error of the difference. The delayed Lognormal was second least preferred, however, the difference was relatively marginal compared to the Normal (534.3, $\Delta SE = \pm 4.9$). We observed similar results in the repeat of B cell data (B-exp2), except for T_{die} where the Weibull provided the best fit (see **Supplementary Figures S5A1–2**).

For CD8⁺ T cells, we present the rank-ordered WAIC plot for all IL-2 concentrations in **Figures 4B1–3** (see **Supplementary Figures S5B1–3** for corresponding CDF plots). We observed either the delayed Lognormal or the delayed Exponential to be the best descriptor for all measurements with the exception of $T_{div}^{k \geq 1}$ in 10 U IL-2 in which Weibull was top-ranked (80.5), but only marginally so compared to the Normal (81.3, $\Delta SE = \pm 1.6$), the Gamma (83.0, $\Delta SE = \pm 3.3$) and the Lognormal (83.3, $\Delta SE = \pm 2.6$). Note that the estimates of WAIC for $T_{div}^{k \geq 1}$ in 1 U IL-2 are unreliable as there are only four data points due to lack of

progression of cell division in those conditions. We reached similar conclusions for both T-exp1 and T-exp2 datasets where the delayed Lognormal and the delayed Exponential were strongly and consistently preferred or was on par with other candidate distributions (see **Supplementary Figure S6**).

In summary, these data suggest that several parametric classes of distributions are well-suited as descriptors. We will, however, provide one example analysis of flow cytometry data where use of Gaussian distributions offer an interpretative advantage over the right-skewed distributions. Moreover, as time to subsequent division has little variability, when fitting fluorescence-activated cell sorting (FACS) data, we will use a reduced model that assumes it is an unknown constant that is, fit.

3 APPLICATION TO FACS DATA

While information from time-lapse microscopy has informed core elements of the Cyton2 model, in practice higher-throughput methodologies are typically employed in immunological investigation. In particular, it is common to have bulk experiments that start with a large number of initial cells that have been cultured with a division tracking dye and are then exposed to stimuli for a time-course of measurements by flow cytometry. Thus it is essential that any mathematical model can be fit to such data and extract biologically meaningful information from them. To that end, we derived expressions for the expected time-course per generation of the Cyton2 model and a least-squares fitting methodology, as described in Methods, for fitting to such data to challenge Cyton2's applicability. We challenged the model with both B and CD8⁺ T cell datasets.

3.1 B Cell Data: Assessing Model Fits

We interrogated a primary dataset consisting of *in vitro* CpG-stimulated murine Bim^{-/-} B cells with cell numbers recorded in each generation via flow cytometry. Cells taken from this mouse strain are deficient in the pro-apoptotic molecule Bim (B-cell-lymphoma-2-like protein 11, or Bcl2l11). As a result, these cells survive longer in culture without impacting any other population dynamic feature (Turner et al., 2008). Here, we asked if a standard division tracking assay, which typically has three replicates at each of five or six harvested time points, provides sufficient information to well constrain model fits. To that end, the dataset that we used consists of nine replicates, collected at nine distinct time points.

To assess the amount of data required to ensure a constrained model fit, we altered the amount of data used according to two scenarios: (i) varying the number of replicates sampled at all time points; and (ii) removing some of the time points while maintaining the number of replicates. In **Figures 5A1–3**, the best-fit model and the estimated parameters with 95% confidence intervals are shown (see Methods). Qualitatively, we observed that the confidence bands of the model fit get narrower around the mean as we increase the number of replicates, indicating an improvement in the model constraints, albeit with a law of diminishing returns. As estimated model parameters are coupled, we assessed their vector values using principle

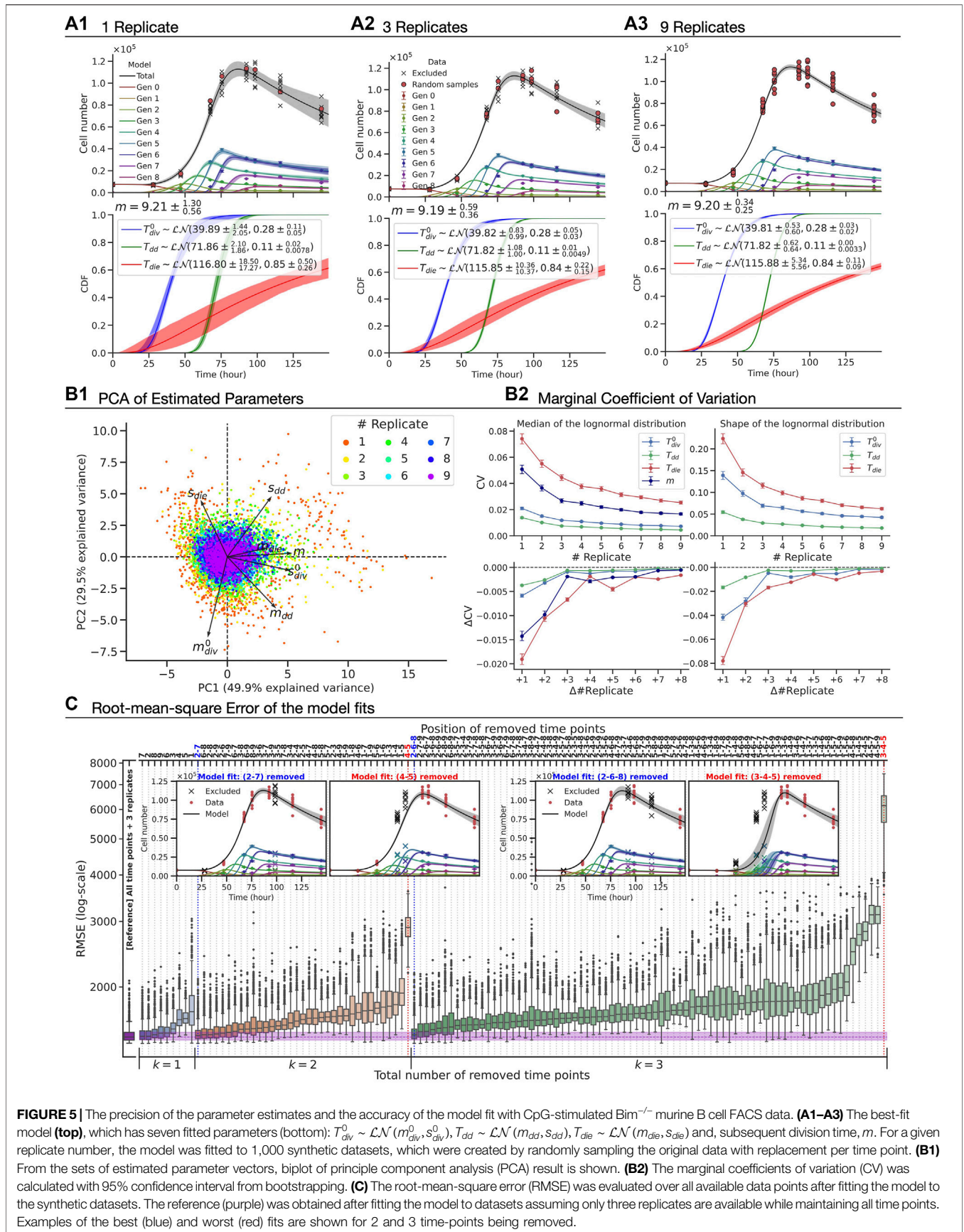


FIGURE 5 | The precision of the parameter estimates and the accuracy of the model fit with CpG-stimulated *Bim*^{-/-} murine B cell FACS data. **(A1–A3)** The best-fit model (top), which has seven fitted parameters (bottom): $T_{div}^0 \sim \mathcal{LN}(m_{div}^0, s_{div}^0)$, $T_{dd} \sim \mathcal{LN}(m_{dd}, s_{dd})$, $T_{die} \sim \mathcal{LN}(m_{die}, s_{die})$ and, subsequent division time, m . For a given replicate number, the model was fitted to 1,000 synthetic datasets, which were created by randomly sampling the original data with replacement per time point. **(B1)** From the sets of estimated parameter vectors, biplot of principle component analysis (PCA) result is shown. **(B2)** The marginal coefficients of variation (CV) was calculated with 95% confidence interval from bootstrapping. **(C)** The root-mean-square error (RMSE) was evaluated over all available data points after fitting the model to the synthetic datasets. The reference (purple) was obtained after fitting the model to datasets assuming only three replicates are available while maintaining all time points. Examples of the best (blue) and worst (red) fits are shown for 2 and 3 time-points being removed.

component analysis (PCA) (Figure 5B1). The PCA result signifies that the first two principle components explain 79% variability in the set and, furthermore, is suggestive that there is no notable correlation amongst components of the parameter vector. To assess the precision of the estimated parameters, we computed coefficients of variation for individual parameters as a function of the number of replicates (Figure 5B2). Again, a law of diminishing returns is observed with no significant benefit in precision of the estimates beyond three replicates. This suggests that the existing operational standard of three replicates offers a good balance between obtaining a precise estimate and managing experimental burden.

We turned our attention to evaluating the model accuracy as time points were removed while maintaining a fixed number of replicates (Figure 5C). For time-point removal, we imposed the following rules to avoid any ambiguity and ensure the feasibility of the model fits:

- 1) At least three time points must be retained.
- 2) Either the first or second time point must remain in order to provide an initial cell number for the model.

Given these rules, there are 366 cases to consider in total. For each case, we constructed 1,000 artificial datasets by randomly sampling three replicates with replacement per remaining time point, fitted the model assuming all random variables were log-normally distributed and then calculated the root-mean-square error (RMSE) of the model-fit from the original unaltered dataset with nine replicates at nine time points. In Figure 5C, we present rank ordered values of the RMSE when one, two or three time-points are removed (see Supplementary Figure S7 for greater than three). The RMSE of model fits using all time points with three replicates is also shown as a reference. Perhaps unsurprisingly, the results showed that capturing a measurement at the time at which cells are expanding is the most important information to be kept for the model accuracy. Intuitively, this would represent a regression of a non-linear curve in which data points around “inflection point” are missing while two ends points are present. Furthermore, we noticed that the first time point is generally more important than the later ones as RMSEs are higher if the first time point was removed. We found little to no difference in the RMSE compared to that of the reference when the positions of the removed time points are sparsely located. As an extreme example with six removed time-points, the model was capable of accurately fitting the data as long as there were three time points that correspond to the early (prior to first division), expansion and contraction phases (e.g., first, fourth and ninth time points, see Supplementary Figure S7B). Knowing in advance those three time points prior to an experiment is unrealistic, and so it represents a lower bound on the number of time-points needed. Removing more than six time-points, the model failed to fit due to the lack of information (results not shown). In summary, this analysis illustrates that Cyton2 is well constrained by data employing standard experimental protocols for following cell expansion by flow cytometry.

3.2 T Cell Data: Assessing Cyton2’s Ability to Draw Biologically Meaningful Inferences

To evaluate the utility of the model in drawing biologically useful inferences, we used it to reassess the non-linear population

dynamics of experiments reported in Marchingo et al. (2014). That study established that CD8⁺ T cells integrated a range of distinct mitogenic stimuli *via* a simple, additive rule for the number of rounds of division they provoked. We questioned how the phenomenon could be understood in light of the new paradigm of familial concordance and global timers as realised in Cyton2.

This data was obtained from *in vitro* CTV-labeled OT-I/Bim^{-/-} CD8⁺ T cells stimulated with the peptide N4 and cultured with co-stimulatory antibodies CD27 (5 μg/ml) and CD28 (2 μg/ml), both alone and in combination. Cells were harvested at 27, 44, 52.5, 66.5, 69, 76.5, 90, 101, and 115.5 h after the stimulation with three replicates at each time point (Marchingo et al., 2014). Mirroring the deduction in the original paper, but expressing it in terms of timers, we sought to ask whether the contribution to division destiny of each co-stimulatory molecule in terms of time could be described by a simple additive process. As there is no simple closed form for the distribution of the sum of two independent lognormally distributed random variables, for this application we instead chose to fit Gaussian distributions. That is, assuming that T_{dd}^{N4} is normally distributed with mean μ_{N4} and variance σ_{N4}^2 , i.e. $\mathcal{N}(\mu_{N4}, \sigma_{N4}^2)$, $T_{dd}^{\alpha CD27}$ is $\mathcal{N}(\mu_{\alpha CD27}, \sigma_{\alpha CD27}^2)$, and $T_{dd}^{\alpha CD28}$ is $\mathcal{N}(\mu_{\alpha CD28}, \sigma_{\alpha CD28}^2)$, if the contributions of $\alpha CD27$ and $\alpha CD28$ to division destiny time were problematically independent and additive, then we would expect that

$$T_{dd}^{\alpha CD27 + \alpha CD28} \sim \mathcal{N}\left(\mu_{N4} + \Delta\mu_{\alpha CD27} + \Delta\mu_{\alpha CD28}, \sqrt{\sigma_{N4}^2 + \Delta\sigma_{\alpha CD27}^2 + \Delta\sigma_{\alpha CD28}^2}\right), \quad (1)$$

where $\Delta\mu_x = \mu_x - \mu_{N4}$ and $\Delta\sigma_x^2 = \sigma_x^2 - \sigma_{N4}^2$ for $x \in \{\alpha CD27, \alpha CD28\}$.

In Figure 6A, we present the total number of cells and the best-fit model with a 95% confidence band around the estimate from the original data. The model was simultaneously fitted to N4, $\alpha CD27$, and $\alpha CD28$ datasets with a shared subsequent division time (see Methods) (Figures 6B1–3), omitting the $\alpha CD27$ plus $\alpha CD28$ dataset for out of sample testing. The estimated m and cumulative distribution function (CDF) of T_{div}^0 , T_{dd} and T_{die} are shown in Figure 6C. In comparison to N4 alone, the addition of $\alpha CD27$ and $\alpha CD28$ extends both means of T_{dd} ($\approx 15\%$) and T_{die} ($\approx 10\%$). Also, we identified $\alpha CD28$ reduces mean of T_{div}^0 (13.3%) while $\alpha CD27$ has minimal impact. Collectively, the compounding effect of these changes results in larger expansion of cell numbers by allowing cells to enter the first division early and to reach destiny and death at later times. Given the parameter estimates for N4, $\alpha CD27$ and $\alpha CD28$, we predicted the number of cells for their combined effect by calculating the T_{div}^0 , T_{dd} and T_{die} according to Eq. 1. Strikingly, this successfully recreated the expansion kinetics of OT-I/Bim^{-/-} CD8⁺ T cells in the presence of both $\alpha CD27$ and $\alpha CD28$ (Figure 6A), supporting the signal integration as a linear sum in a time domain of three dimensions and consistent with the independence of the timers. Additionally, we recapitulated the additive nature of mean division number presented in Marchingo et al. (2014) using ABM given the fitted and predicted parameter estimates (see Supplementary Figure S8). These results illustrate

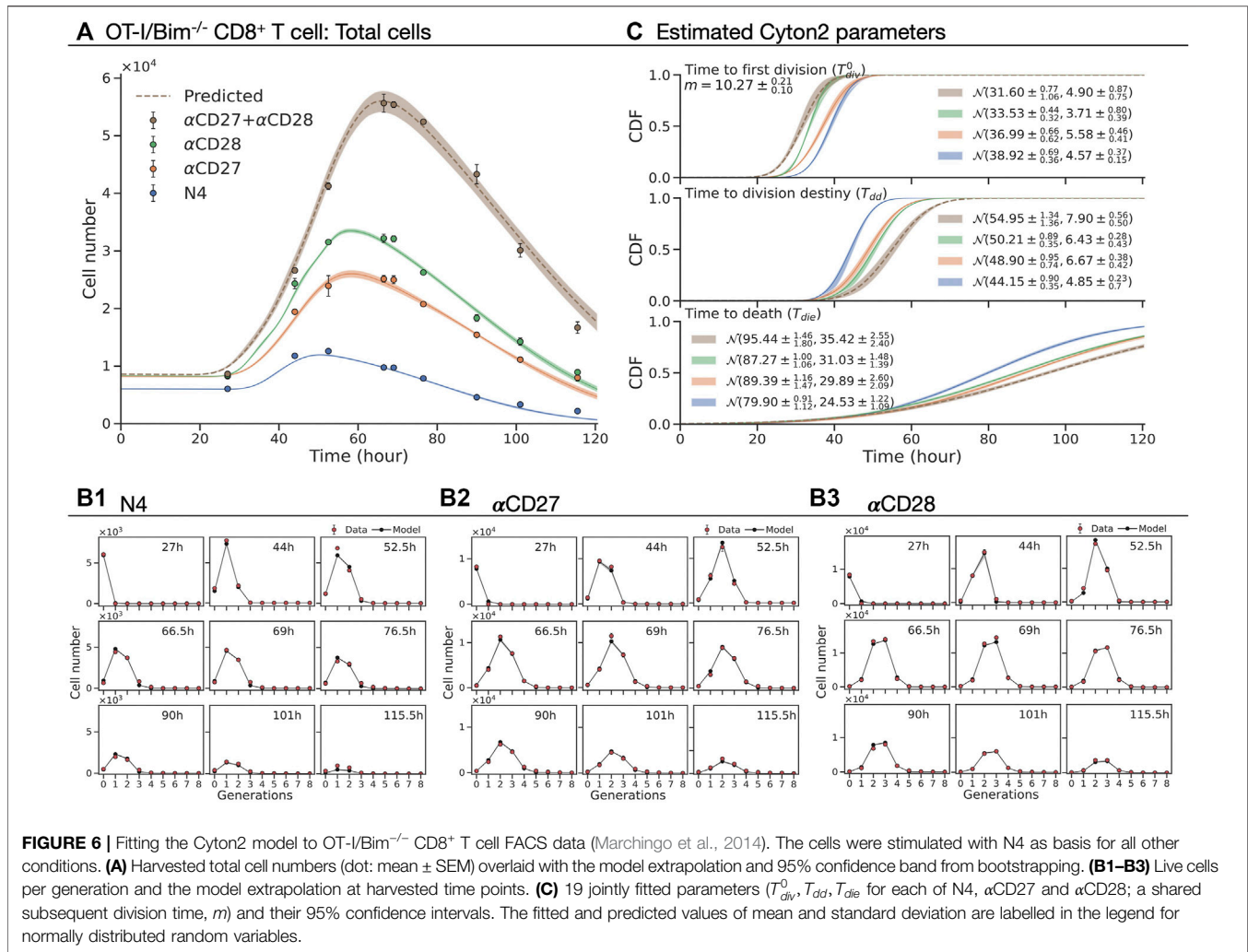


FIGURE 6 | Fitting the Cyton2 model to OT-I/Bim^{-/-} CD8⁺ T cell FACS data (Marchingo et al., 2014). The cells were stimulated with N4 as basis for all other conditions. **(A)** Harvested total cell numbers (dot: mean ± SEM) overlaid with the model extrapolation and 95% confidence band from bootstrapping. **(B1–B3)** Live cells per generation and the model extrapolation at harvested time points. **(C)** 19 jointly fitted parameters (T_{div}^0 , T_{dd} , T_{die} for each of N4, αCD27 and αCD28; a shared subsequent division time, m) and their 95% confidence intervals. The fitted and predicted values of mean and standard deviation are labelled in the legend for normally distributed random variables.

the merit of Cyton2 in uncovering how simple operations can underlie highly non-linear population dynamics.

4 DISCUSSION

The vast majority of published mathematical models of lymphocyte population dynamics employed assume that a newly born cell's fate is independent of its family's history (Smith and Martin, 1973; Nordon et al., 1999; Revy et al., 2001; Ganusov et al., 2005; Yates et al., 2007; Lee et al., 2009; Banks et al., 2012; Hasenauer et al., 2012; Mazzocco et al., 2017), with a few notable exceptions (Hyrien et al., 2010; Wellard et al., 2010; Zilman et al., 2010; Shokhirev et al., 2015; Yates et al., 2017). These assumptions are adopted, not because they are consistent with experimental data from, for example, filming, FACS and lineage tracing, but for reasons of parsimony, model identifiability and computational ease of fitting (Dowling et al., 2005; Boer et al., 2006). In this work, we have presented a variant of the original Cyton model that encapsulates features of inheritance and correlation structure of cell fates. This was

achieved by introducing new random variables that describe the time to division destiny of a family and a global death time, which describes a single death time for all members in a family tree. Similar to the Cyton model, this variant offers a general tool for analysing lymphocyte proliferation and survival, including from the data obtained from CFSE/CTV-labeled division tracking assays. Despite concerns that the inclusion of familial effects might result in a model with too many parameters or one that is, hard to fit, neither proves to be the case, making the model suitable for general use.

The analysis of the B and CD8⁺ T cell filming data allowed direct tests on the model's assumptions of independent timers. Additionally, it enabled us to assess the suitability of classes of parametric distributions of random variables of the Cyton2 model, which is necessary for when the model is used with commonly available non time-lapse microscopy data. As there is no theoretical reason to favour one distribution class over another and several classes provide good fits to data, for most of our fitting examples we adopted the lognormal distribution class.

To fit Cyton2 to ubiquitous FACS data, we derived an expression for the mean population dynamics with one

constant and three sets of distribution parameters. The random variables represent times to first division, to division destiny and to death, and the constant captures the subsequent division time. In the present work, the model was designed for cell populations exposed to newly available stimuli. In future work we will consider the inclusion of repeated challenges and continuous feedback mechanisms as occur, for example, with autocrine signalling *via* IL-2. Alterations to the model that allow the inclusion of ongoing signalling, as likely occurs when fighting replicating pathogens such as viruses or bacteria, will be the subject of future development. Moreover, the cell population model considered here does not include differentiation or other developments that would create asymmetries through altered division times, destiny or survival. Such alternative fates can arise from analogous competing outcomes promoting differentiation (Duffy et al., 2012), and we anticipate that the basic Cyton2 framework introduced here will be expanded to encompass additional fates as experimental information for the control of differentiation is acquired.

We provided two illustrative uses of the model through the analysis of FACS data from stimulated B and CD8⁺ T cell cultures. The first one provides quantitative support for the standard experiment design of triplicates per time-point, but also elucidates the importance of including time-points around the initial expansion and final contraction of the population. The second example revisits the work of Marchingo et al. (2014) addressing the question of signal integration by T cells. While the original study was informed based on modelling paradigms available at the time, reconsideration of it terms of family-based timers draws similar conclusions on additivity, but with a distinct temporal understanding that will influence all subsequent studies. Taken together these results suggest Cyton2 will prove to be a powerful tool in the quantitative assessment of immune responses. We anticipate the model will be useful in evaluating signal processing and genetic differences in both murine and human T and B cells, and will facilitate comparisons between healthy and unwell individuals.

The model is informed by, and the worked examples are for, data from *in vitro* experiments where stimulation is provided to a group of T or B cells, and the resulting proliferation occurs in a burst that can be followed by division tracking dyes or direct filming. Those population dynamics follows the pattern of an exponential rise, a period of division cessation, and then of cell loss that characterises immune responses *in vivo* (Veiga-Fernandes et al., 2000; Costa Del Amo et al., 2020). As such, as with the original Cyton model (Hawkins et al., 2007; Subramanian et al., 2008; Marchingo et al., 2014), Cyton2 can be successfully fit to *in vivo* data (data not shown). We note, however, that for most *in vivo* data stochastic models offer no advantage over models, such as those based on ODEs, that assume transitions to distinct phases and require fewer parameters (Boer and Perelson, 2005; Ganusov et al., 2005; Boer et al., 2006). Furthermore, these latter models often include parameters for transitioning to the memory phase through a second, slower rate of loss, and this has not been implemented into the Cyton framework yet as the mode of transition is not known. The difference between fitting

parsimonious models to *in vitro* and *in vivo* data may eventually be reconciled as the community continues to improve methods that introduce differentiation, memory formation and reveal additional features of responding cell phenotype, such as cell cycle status, as originally envisaged by Antia et al. (2003).

5 METHODS

5.1 Mice

All mice were maintained under specific pathogen-free conditions in the WEHI animal facilities (Parkville, Victoria, Australia) and used at 5–12 weeks of age. All experiments were performed under the approval of the WEHI Animal Ethics Committee. FUCCI red/green (RG) mice were acquired by crossing FUCCI Red (B6.B6D2-Tg (FUCCI)639Bsi) with FUCCI Green (B6.B6D2-Tg (FUCCI)492Bsi) mice, both obtained from Riken BioResource Centre (Sakaue-Sawano et al., 2008). FUCCI RG mice were then crossed to OT-I or OT-I/Ly5.1 mice to obtain OT-I/FUCCI RG and OT-I/FUCCI RG-Ly5.1 respectively (Dowling et al., 2014). In one experiment (stimulation with N4, α CD28 and IL-12) cells from C57BL/6 mice that were irradiated and reconstituted with bone marrow from OT-I/FUCCI RG-Ly5.1 were used.

5.2 CD8⁺ T Cell Isolation

OT-I CD8⁺ T cells were isolated from single cell suspensions prepared from lymph nodes (axillary, branchial, inguinal) by negative selection using EasySep Mouse CD8 α^+ T cell Isolation kit (StemCell technologies) according to the manufacturer's protocol. Purity (CD8⁺ V α 2⁺) was typically between 80 and 95%. Splenocytes were used for the isolation of CD8⁺ V α 2⁺Ly5.1⁺ T cells from C57BL/6 mice that were irradiated and reconstituted with bone marrow from OT-I/FUCCI RG-Ly5.1. Purified CD8⁺ T cells were labelled with 5 μ M CellTrace™ Violet (CTV, Invitrogen) to track and monitor cell division in parallel bulk cultured by flow cytometry. Labelling was performed for 20 min at 37°C in PBS + 0.1% BSA.

5.3 *In Vitro* Cell Culture

All T cell cultures were prepared using filming medium (GIBCO advanced RPMI 1640 without phenol red + 5% GIBCO FCS) at 37°C and 5% CO₂ in a humidified atmosphere. All cell cultures contained 25 μ g/ml anti-mouse IL-2 antibody (S4B6: WEHI antibody facility) which neutralises mouse IL-2 but does not recognise human IL-2 (Deenick et al., 2003).

Cells were either stimulated with plate bound anti-CD3 (α CD3: WEHI antibody facility, clone 145-2C11: 10 μ g/ml) or with the peptide for the OT-I TCR, SIINFELK (N4) (Auspep) at 0.01 μ g/ml.

For stimulation with α CD3, CD8⁺ T cells were cultured on 24-well plates coated with α CD3 at 40,000 cells in 1 ml per well in the presence of 1, 3.16, or 10 U/ml recombinant human IL-2 (Peprotech) and 25 μ g/ml S4B6. After 24 h of culture cells were harvested, washed twice with filming medium, counted and resuspended at 5,000–10,000 cell/ml in filming medium

supplemented with 25 $\mu\text{g/ml}$ S4B6 and 1, 3.16, or 10 U/ml recombinant human IL-2. The units of IL-2 were abbreviated to 1U, 3U and 10U for each concentration throughout the paper.

In experiments using N4 peptide CD8⁺ T cells were cultured with 0.01 $\mu\text{g/ml}$ N4 at 2×10^4 cells per ml in 200 μL of a 96-well U-bottom plate in the presence of 25 $\mu\text{g/ml}$ S4B6. 2 $\mu\text{g/ml}$ αCD28 (clone 37.51, WEHI antibody facility) or 1 ng/ml IL-12 (Peprotec) were added as indicated. Cells were cultured for 24 h, washed and resuspended at 5,000–10,000 cells/ml in filming medium containing 25 $\mu\text{g/ml}$ S4B6.

For one experiment CD8⁺ T cells cultured for 24 h with N4 in presence or absence of αCD28 were split and supplemented or not with 1 U/ml rhIL-2 before replating for filming.

In one experiment CD8⁺ T cells were cultured with N4 alone and addition of either αCD28 , IL-12, or both for 24 h, then washed, resuspended, and replated for filming without any further stimuli added.

For filming, 250 μL cell suspension was added per chamber of an 8 well μ -Slide chamber (Ibidi) containing 125 μm (MGA-125-01) or 70 μm (MGA-7-01) microgrids (Daniel Day, Microsurfaces). These conditions resulted in a significant portion of microwells containing 1 cell per well. Before the start of filming, cells were incubated for ≈ 2 h at 37°C with 5% CO₂ in a humidified atmosphere. Slide chambers were then transferred to an environmentally controlled microscope (Carl Zeiss) and incubated at 37°C with 5% CO₂ in a humidified atmosphere.

5.4 Live Cell Imaging and Cell Tracking

For single cell filming, microgrids (70/125 μm , Daniel Day, Microsurfaces) were placed into an 8 well chamberslide (Ibidi μ -slide). Chambers were UV sterilised with 40 μL 100% ethanol in a laminar-flow cabinet for at least 30 min until dry. Another 40 μL ethanol was added and rinsed 10x with filming medium (advanced RPMI 1640 without phenol red). 250 μL filming medium was added to each chamber and left in the incubator at 37°C overnight to dissolve air bubbles. To reduce background fluorescence of the medium, chambers (grids) were bleached for 2 h using a 470 nm LED, just prior to adding the cells.

Live cell imaging was conducted on an environmentally controlled 37°C + 5% CO₂ humidified Zeiss Axiovert 200 M microscope. Brightfield images were captured with a Zeiss AxioCam MRm (1.4 megapixels) attached to a 0.63x C-mount, using a Plan-Apochromat 20x objective (0.8 n.a.). A GFP/DsRed-A (Semrock) filter block (excitation LED 470/555 nm set at 25%/100%, respectively, with an exposure time of 200 μs) was used for detecting green and red fluorescence. Red, green, bright-field and out-of-focus images were taken at 165 s intervals for 5 or 6 days. Bright-field and fluorescent raw images of single cells in microgrids were digitally processed resulting in overlaid red/green images corrected for background noise of the medium.

Cell tracking was performed using the image processing package FIJI (Schindelin et al., 2012). Lineage Tracker plug-in was used for cell segmentation and tracking (Downey et al., 2011). Gaps or mistakes in segmentation and tracking were adjusted

manually to ensure data accuracy. Cells in each well were followed until they either died, indistinguishable from nearby cells or the experiment ended.

For CD8⁺ T cell data in the presence of IL-2, measurements from two independent experiments were aggregated.

5.5 Data Selection and Tree Collapse

For each family tree $c \in \mathbb{N}_{\geq 0}$, the times to divide $\{T_{div}^x\}_c$, to die $\{T_{die}^x\}_c$ and to loss $\{T_{loss}^x\}_c$ of all cells were recorded using time-lapse microscope. All of these variables were measured from $t = 0\text{h}$, i.e., the beginning of the experiment. T_{loss} is defined as the time at which the cell becomes indistinguishable to the nearby cells, or survives until the end of given experiment time frame, thus, were lost from the experiment. In order to keep track of the cells' relation, a unique label was given to each cell by x . Let \mathcal{X}_c be the collection of all x for a family c , where $x = x_1, x_2, \dots, x_j$ with $x_j \in \{1, 2\}$ is a finite and ordered sequence of 1 and 2. Beginning with a founder cell, defined as $x = 0$, we denote its first and second daughter cells in generation 1 by $x = 1$ and $x = 2$, respectively. In general, x_1, x_2, \dots, x_j represents the x_j^{th} daughter of the ... of the x_2^{th} daughter of the x_1^{th} daughter of the founder cell (see Harris, 1963, Ch.6). For example, $x = 1, 1, 2$ denotes the second daughter of the first daughter of the first daughter of the founder cell. Given a unique identifier of the cell, the generation k is noted $g(x) = k$ with $g(0) = 0$. With this construct, we define the raw measurement of times as a set $\mathcal{T}_c = \{T_{div}^x, T_{die}^x, T_{loss}^x : x \in \mathcal{X}_c\}$.

For the analyses in Section 2.2 and Section 2.4, we filtered for families that had at least divided once and satisfied the condition $\max(\mathcal{T}_c) = T_{die}^x$. In essence, we eliminated incomplete family trees that contain unusually long-surviving cells, but allowed lost cells to be in place as long as the last observed event is death in a given family. Indeed, there is an increasing chance of observing more lost cells as the family gets larger. However, it was previously shown that the regularity of a family is a result of correlated cell divisions as a biological feature inherited within the family even when considering the unrecovered samples (Marchingo et al., 2016). Therefore, it is highly likely that the lost cells due to indistinguishable circumstances might had undergone similar fates with its sibling, thereby maximising the number of data points while reducing any potential selection bias, whereas it is difficult to weigh how including the long-surviving cells might affect all the other analyses. We noted the long-surviving and “no division” cells constitute approximately 23 and 29% on average, respectively, across all the experiments (data not shown).

Given the heritable feature, we summarise a family tree by collapsing it to a single representative line (Figure 1C). By collapsing, we mean substitute average time to divide (and to die) of the cells in a given generation k . We also enumerated all dead cells within a family and calculated mean time to last division (T_{ld}) as a proxy to the division destiny time. In summary, we represent a single family by $\mathbf{T}(c) = (T_{div}^0, \dots, T_{div}^k, T_{die}^0, \dots, T_{die}^k)$ so long as we observed division or death events in each generation k . Table 3 shows the number of retained clones used in all analyses presented in this paper after applying the filtering rule.

TABLE 3 | Number of clones used in the analysis of the filtering datasets. The numbers were obtained by filtering on clones that had divided at least once and whose last event was a death event (not a loss or a division). The percentage is expressed in relation to the total number of clones in the experiment. For a given cell type and stimulation, the values of each Cyton2 variable were extracted for the statistical analyses.

Cell type	Stimulation	Number of clones			
		Time to first div. (T_{div}^0)	Subsequent div. Time (T_{div}^k)	Time to last div. (T_{ld})	Time to death (T_{die})
B (B-exp1)	CpG	69 (63.9%)	56 (51.9%)	69 (63.9%)	69 (63.9%)
B (B-exp2)	CpG	73 (83.0%)	63 (71.6%)	73 (83.0%)	73 (83.0%)
CD8 ⁺ T	1U IL-2	28 (25.7%)	4 (3.7%)	28 (25.7%)	28 (25.7%)
	3U IL-2	34 (37.8%)	13 (14.4%)	34 (37.8%)	34 (37.8%)
	10U IL-2	50 (30.7%)	16 (9.8%)	50 (30.7%)	60 (30.7%)
CD8 ⁺ T (T-exp1)	N4	20 (44.4%)	0 (0%)	20 (44.4%)	20 (44.4%)
	N4 + α CD28	19 (46.3%)	1 (2.4%)	19 (46.3%)	19 (46.3%)
	N4 + IL-2	28 (75.7%)	4 (10.8%)	28 (75.7%)	28 (75.7%)
	N4 + α CD28 + IL-2	33 (89.2%)	13 (35.1%)	33 (89.2%)	33 (89.2%)
CD8 ⁺ T (T-exp2)	N4	27 (73.0%)	12 (32.4%)	27 (73.0%)	27 (73.0%)
	N4 + α CD28	17 (44.7%)	12 (31.6%)	17 (44.7%)	17 (44.7%)
	N4 + IL-12	13 (44.8%)	9 (31.0%)	13 (44.8%)	13 (44.8%)
	N4 + α CD28 + IL-12	11 (30.6%)	7 (19.4%)	11 (30.6%)	11 (30.6%)

5.6 Agent-Based Model

We developed an agent-based model (ABM) to simulate cells in a single family with the correlated structure proposed for the Cyton2 model. Each realisation of the simulation represents one clonal family. Upon initialisation, the founder cell is assigned time to first division, global destiny and global death times, which are drawn randomly from three independent lognormal distributions. Also, the subsequent division times are randomly sampled for each generation from a lognormal distribution, but the progeny cells of the same generation share the division time. If the founder cell reaches time to first division, it creates two daughter cells, which inherit global destiny and death times. If the cell reaches its division destiny, we immediately classify it as a destiny cell and prevent it from further division. When the cells reach death time, they are removed from the simulation. The model was implemented in Python (version 3.8.6).

5.7 Statistical Analysis: Bayesian Framework

In Section 2.2, 2.3, the correlations of all possible pairs between time to first division (T_{div}^0), average subsequent division time ($T_{div}^{k \geq 1}$), time to last division (T_{ld}) and time to death (T_{die}) were estimated using Bayesian inference. For a given pair of variables and its observed data, say $d_i \in \mathcal{D} = \{(x_i, y_i) : i = 1, 2, \dots, n\}$ where n is the number of observations, we used bivariate normal distribution to estimate the correlation coefficient (ρ). This entails $x_i \sim \mathcal{N}(\mu_x, \sigma_x)$ and $y_i \sim \mathcal{N}(\mu_y, \sigma_y)$. With uninformative priors on the hyper-parameters $\mu_x, \mu_y \sim U(0, 1000)$, $\sigma_x, \sigma_y \sim U(0, 1000)$ and $\rho \sim U(-1, 1)$, we define the bivariate normal distribution.

$$d_i \sim \mathcal{N}(\boldsymbol{\mu}, \boldsymbol{\Sigma}),$$

where $\boldsymbol{\mu} = (\mu_x, \mu_y)$ is a vector of means for x_i and y_i , and $\boldsymbol{\Sigma} = \begin{bmatrix} \sigma_x^2 & \rho\sigma_x\sigma_y \\ \rho\sigma_x\sigma_y & \sigma_y^2 \end{bmatrix}$ is a covariance matrix. We used an extension of the Hamiltonian MCMC algorithm, No-U-Turn Sampler

(Hoffman and Gelman, 2014), implemented in PyMC3 (version 3.9.3) (Salvatier et al., 2016) to obtain the marginal posterior distributions of $\rho, \mu_x, \sigma_x, \mu_y, \sigma_y$. Given these distributions, we calculated 95% credible interval for ρ , and 90, 95, and 99% density regions of (x, y) . In addition, we formulated Bayesian hypothesis testing, where the null hypothesis is $H_0: \rho = 0$ and alternative hypothesis is $H_1: \rho \neq 0$ (which translates to $H_1: \rho \sim U(-1, 1)$) (Jeffreys, 1961). This is formally stated as a ratio of likelihoods of hypotheses given the data,

$$\frac{P(H_0|\mathcal{D})}{P(H_1|\mathcal{D})} = \frac{P(H_0)}{P(H_1)} \times \frac{P(\mathcal{D}|H_0)}{P(\mathcal{D}|H_1)}$$

In order to grade if the data is more probable under H_0 or H_1 , the Bayes factor $BF_{01} = P(\mathcal{D}|H_0)/P(\mathcal{D}|H_1)$ was used given priors of $P(H_0)$ and $P(H_1)$. When $H_1: \rho \sim U(-1, 1)$, it can be computed by evaluating the following integral (Jeffreys, 1961; Wagenmakers et al., 2016):

$$BF_{01} = 1/BF_{10}, \text{ where } BF_{10} = \frac{1}{2} \int_{-1}^1 \frac{(1-\rho^2)^{\frac{n-1}{2}}}{(1-\rho r)^{n-\frac{3}{2}}} d\rho,$$

where r denotes for the sample correlation defined as $r = \sum_{i=1}^n (x_i - \bar{x})(y_i - \bar{y}) / \sqrt{\sum_{i=1}^n (x_i - \bar{x})^2 \sum_{i=1}^n (y_i - \bar{y})^2}$. For the interpretation of Bayes factor, we adopted the discrete categories of evidential strength proposed in Jeffreys (1961) (Table 4).

In Section 2.4, six probability distributions were assessed for $T_{div}^0, T_{div}^{k \geq 1}, T_{ld}, T_{die}$ under the Bayesian framework in a similar manner to estimating the correlation coefficient. Table 5 shows the list of candidate distributions and the uninformative priors prescribed for respective hyper-parameters.

Given posterior distributions of the parameters, we adopted WAIC (Watanabe and Opper, 2010) score to quantitatively assess the candidates, which is estimated as follows:

TABLE 4 | Bayes factor interpretation.

Bayes factor: \mathbf{BF}_{01} (\mathbf{BF}_{10})	Interpretation
>100	Extreme evidence for H_0 (H_1)
30–100	Very strong evidence for H_0 (H_1)
10–30	Strong evidence for H_0 (H_1)
1–3	Anecdotal evidence for H_0 (H_1)
1	No evidence

$$\text{WAIC}(z, \Theta) = -2 \left(\sum_{i=1}^n \log \left[\frac{1}{S} \sum_{s=1}^S P(z_i | \Theta_s) \right] - \sum_{i=1}^n \text{Var}_{s=1}^S \log(P(z_i | \Theta_s)) \right), \tag{2}$$

where z is the data with n independent number of observations, Θ is the posterior distribution, Θ_s is the s th set of sampled parameter values in the posterior distribution with S number of samples and $\text{Var}_{s=1}^S a_s = \frac{1}{S-1} \sum_{s=1}^S (a_s - \bar{a})^2$ denotes for the sample variance (see McElreath, 2020, Ch.7; Vehtari et al., 2017). The first and the second terms in Eq. 2 are known as the log-pointwise-predictive-density (lppd) and the penalty term, respectively. For direct comparison of the candidates, we computed the standard error by calculating the variance over the individual observations instead of their summation under the assumption of normality of WAIC.

$$\text{se}(\text{WAIC}) = \sqrt{n \times \text{Var}_{i=1}^n \left(-2 \left(\log \left[\frac{1}{S} \sum_{s=1}^S P(z_i | \Theta_s) \right] - \text{Var}_{s=1}^S \log(P(z_i | \Theta_s)) \right) \right)}.$$

Let us denote WAIC_i to be the term in $\sqrt{(\cdot)}$ such that $\text{WAIC} = \sum_{i=1}^n \text{WAIC}_i$, then the standard error of the difference of WAIC between, for instance, candidate A and B can be calculated,

$$\text{se}(\text{WAIC}^A - \text{WAIC}^B) = \sqrt{n \times \text{Var}_{i=1}^n (\text{WAIC}_i^A - \text{WAIC}_i^B)}.$$

5.8 Equations for Dynamic Evolution of the Mean

In commonly employed division diluting dye experiments, individual families are not observed and initial cell numbers are typically in their thousands suggesting the use of mean system behaviour as an appropriate descriptor. Thus to fit the

model to such data we derive equations for the mean population dynamics per generation for Cyton2.

Let $Z_g(t)$ denote the number of cells alive in generation $g \in \{0, 1, \dots, G\}$ at time $t \geq 0$. Then, $Z_g(t)$ can be expressed with the variables shown in Section 2.1 for any chosen probability density functions for the random variables. Here, we separately derived $\mathbb{E}[Z_g(t)]$ for $g = 0$ and $g > 0$ cases as lymphocytes generally take longer to divide for the first time than at later generations. In essence, we begin the derivation with parameters $\theta = (T_{div}^0, \{M_g\}_{g \geq 1}, T_{dd}, T_{die})$ denoting time to first division, subsequent division time per generation, time to destiny and time to death, respectively.

Generation Zero

We assume we are following the activation dynamics of a set of resting cells and these cells are provided with signals that program a limited proliferative response. For the purposes here, we also assume that all cells are activated at time $t = 0$, erasing the prior cell programming and survival characteristics. Situations where only a proportion of cells are activated, or where the activated cells take some extended time to transition to the new programming, leading to some early cell death are useful modifications suited to particular applications. Such modifications are discussed further in **Supplementary Material**.

For a given family tree, the number of live cells dividing, dying or reaching destiny in generation $g = 0$ at time t is given by

$$Z_0(t) = \mathbb{1}_{\{T_{die} > t\}} \mathbb{1}_{\{\min(t, T_{dd}) < T_{div}^0\}}, \tag{3}$$

where $\mathbb{1}$ is an indicator function. Assuming that the random variables T_{die} , T_{dd} and T_{div}^0 are independent of each other as we established in Section 2.2, the expected number is given by

$$\mathbb{E}[Z_0(t)] = P(T_{die} > t) P(\{\min(t, T_{dd}) < T_{div}^0\}).$$

We expand the second term, and by the law of total probability we obtain

$$P(\{\min(t, T_{dd}) < T_{div}^0\}) = P(T_{div}^0 > t) P(T_{dd} > t) + P(T_{div}^0 > T_{dd}) P(T_{dd} \leq t).$$

Thus, the expected number of cells in generation zero is,

$$\mathbb{E}[Z_0(t)] = P(T_{die} > t) \left[P(T_{div}^0 > t) P(T_{dd} > t) + \int_0^t dP(T_{dd} \leq \tau) P(T_{div}^0 > \tau) \right]. \tag{4}$$

TABLE 5 | List of candidate parametric distribution classes.

Candidates	Priors	Target distribution
A	$\alpha_G, \beta_G \sim U(0, 200)$	$T_{div}^0, T_{div}^{k \geq 1}, T_{id}, T_{die} \sim \text{Gamma}(\alpha_G, \beta_G)$
B	$m, s \sim U(0, 200)$	$T_{div}^0, T_{div}^{k \geq 1}, T_{id}, T_{die} \sim \text{LN}(m, s)$
C	$\mu, \sigma \sim U(0, 200)$	$T_{div}^0, T_{div}^{k \geq 1}, T_{id}, T_{die} \sim N(\mu, \sigma)$
D	$\alpha_W, \beta_W \sim \text{HalfNormal}(500)$	$T_{div}^0, T_{div}^{k \geq 1}, T_{id}, T_{die} \sim \text{Weibull}(\alpha_W, \beta_W)$
E	$\lambda \sim U(0, 2), c \sim U(0, \infty)$	$T_{div}^0, T_{div}^{k \geq 1}, T_{id}, T_{die} \sim \text{Delayed Exp}(\lambda, c)$
F	$m_d, s_d \sim U(0, 200), c \sim U(0, \infty)$	$T_{div}^0, T_{div}^{k \geq 1}, T_{id}, T_{die} \sim \text{Delayed LN}(m_d, s_d, c)$

This equation can be interpreted as follows: a cell in generation zero remains alive when $T_{die} > t$, and it is sorted either in initial state or in destiny state. The cell in the initial state can divide, reach destiny or die whichever event comes first. However, the destiny cell can no longer divide but only awaits for death.

Generations > 0

To calculate the expected number of live cells for $g > 0$, we limit the windows of cells being in generation g by constraining with $t \in [T_{div}^0 + \sum_{k=1}^{g-1} M_k, T_{div}^0 + \sum_{k=1}^g M_k)$, that is

$$Z_g(t) = 2^g \mathbb{1}_{\{T_{die} > t\}} \mathbb{1}_{\left\{T_{div}^0 + \sum_{k=1}^{g-1} M_k \leq \min(t, T_{dd}) < T_{div}^0 + \sum_{k=1}^g M_k\right\}} \quad (5)$$

The factor 2^g is required to include the effect of clonal expansion of the cells that have divided g times. Assuming T_{die} , T_{dd} , M_k and T_{div}^0 are independent of each other, the expected value is

$$\mathbb{E}[Z_g(t)] = 2^g P(T_{die} > t) P\left(T_{div}^0 + \sum_{k=1}^{g-1} M_k \leq \min(t, T_{dd}) < T_{div}^0 + \sum_{k=1}^g M_k\right).$$

Defining $X_g = T_{div}^0 + \sum_{k=1}^{g-1} M_k$ and $Y_g = M_g$, then, similarly to the $g = 0$ case, we expand and employ the law of total probability to obtain

$$P(X_g \leq \min(t, T_{dd}) < X_g + Y_g) = P(X_g \leq t < X_g + Y_g)P(T_{dd} > t) + P(X_g < T_{dd} < X_g + Y_g)P(T_{dd} \leq t).$$

Hence, the expected number of cells in $g > 0$ is

$$\mathbb{E}[Z_g(t)] = 2^g P(T_{die} > t) \times \left[P(T_{dd} > t) \int_0^t dP(X_g \leq \tau) P(Y_g > t - \tau) + \int_0^t dP(T_{dd} \leq \tau) P(X_g < \tau < X_g + Y_g) \right]. \quad (6)$$

Together with Eqs 4, 6, we can calculate the average number of live cells for a family in generation g at time t for any distribution class of T_{div}^0 , $\{M_g\}_{g \geq 1}$, T_{dd} and T_{die} . Since the equations are equally applicable for N_0 number of initial founder cells, we generalise these by multiplying N_0 such that

$$y_g(t; \theta) := \mathbb{E}[N_0 Z_g(t; \theta)] = N_0 \mathbb{E}[Z_g(t; \theta)],$$

where $\theta = (T_{div}^0, \{M_g\}_{g \geq 1}, T_{dd}, T_{die})$ are the parameters of the Cyton2 model. Typically, the random variables are equipped with a lognormal distribution, which has two additional parameters, thus, we have total of 6 + 2g free parameters to estimate.

Reduced Cyton2 Model

To fit FACS data, we simplify the model by assuming that the subsequent division time is a constant rather than a set of random variables, that is, $M_g = m \in \mathbb{R}_{>0}$ for all $g > 0$. This is based on the empirical observation made from filming data that, after the first division, the cells divide at a consistent rate with little inter- and intra-clonal variability (Figures 2A2, B2). This step drastically reduces the number of free parameters, and it no longer depends on the number of generations but purely on the choices of probability density function of T_{div}^0 , T_{dd} and T_{die} . Essentially, the reduced model has $\tilde{\theta} = (T_{div}^0, m, T_{dd}, T_{die})$ parameters. Since

Eq. 4 does not depend on the subsequent division time, it remains the same:

$$\mathbb{E}[\tilde{Z}_0(t)] = \mathbb{E}[Z_0(t)] = P(T_{die} > t) \left[P(T_{div}^0 > t)P(T_{dd} > t) + \int_0^t f_{T_{dd}}(\tau)P(T_{div}^0 > \tau)d\tau \right],$$

where $f_{T_{dd}}$ is the probability density function of T_{dd} . However, Eq. 6 can be further simplified to

$$\mathbb{E}[\tilde{Z}_g(t)] = 2^g P(T_{die} > t) \times \left[P(T_{dd} > t)P(t - gm < T_{div}^0 < t - (g-1)m) + \int_0^t f_{T_{dd}}(\tau)P(\tau - gm < T_{div}^0 < \tau - (g-1)m)d\tau \right].$$

We used the reduced Cyton2 model for all our analyses of FACS data presented in this paper.

$$\tilde{y}_g(t; \tilde{\theta}) := N_0 \mathbb{E}[\tilde{Z}_g(t; \tilde{\theta})]. \quad (7)$$

5.9 Fitting the Cyton2 Model

Division structured population datasets obtained from FACS were fitted to the reduced Cyton2 model (i.e. Eq. 7). In total, there are 7 parameters to be estimated for each dataset assuming that the random variables are lognormally or normally distributed, thus if we have N number of conditions, we have a maximum of $7N$ free parameters to be fitted. For all conditions, we always used cell numbers at the beginning of the stimulus (typically at $t = 0$) as a fixed initial cell number.

For each set of cell numbers $\{n_{g,r}(t_i)\}$ from the data, where $i \in \{0, 1, \dots, I\}$, $g \in \{0, 1, \dots, G\}$ and $r \in \{0, 1, \dots, R\}$ are time, generation and replicate indices, respectively, we obtained point estimates of the parameters. To achieve this, we used least-squares method with Levenberg-Marquardt (Marquardt, 1963) optimisation algorithm implemented in Python library LMFIT (version 1.0.2) (Newville et al., 2014). We defined the residual sum of squares (RSS) as our cost function,

$$C(\tilde{\theta}) = \sum_{i=0}^I \sum_{g=0}^G \sum_{r=0}^R \left(n_{g,r}(t_i) - \tilde{y}_g(t_i; \tilde{\theta}) \right)^2,$$

such that we find an approximate minimum,

$$\{\tilde{\theta}^*\} \in \arg \min_{\tilde{\theta}} C(\tilde{\theta}).$$

As the algorithm requires a set of starting parameter values, we prescribed 100 sets of initial values drawn uniformly at random from the appropriate parameter ranges, and recorded RSS for each set to identify the best fitted parameters by the lowest RSS. For fitting multiple datasets simultaneously, which requires an extra sum over all datasets in the cost function, the algorithm needs to explore higher dimension of the parameter space compared to fitting one dataset at each iteration. Therefore, we used 200 sets of initial values to increase range of the exploration. After identifying the best fit, we performed bootstrap method (Efron, 1979) with an artificial dataset that was resampled with replacement (per time point) from the original measured data. We repeated this process 1,000 times, which resulted in 1,000 additional estimates for each parameter. This allowed us to calculate 95% confidence intervals on the best fitted parameter values. Additionally, we also obtained confidence bands for

extrapolated cell numbers by calculating 95% percentile range at each of discretised time point from the model.

DATA AVAILABILITY STATEMENT

The datasets presented in this study can be found in online repositories. The names of the repository/repositories and accession number(s) can be found below: <https://github.com/hodgkinlab/cyton2-paper>.

ETHICS STATEMENT

The animal study was reviewed and approved by the Animal Ethics Committee Walter and Eliza Hall Institute.

AUTHOR CONTRIBUTIONS

HC, KD, and PH analysed and interpreted experiments, and wrote the manuscript. SO, EH, JM, and SH planned, analysed, and performed the experiments. AK and GP contributed to data analysis. AK, GP, and SD contributed to the data interpretation. All authors read and contributed revisions to the manuscript.

REFERENCES

- Antia, R., Bergstrom, C. T., Pilyugin, S. S., Kaech, S. M., and Ahmed, R. (2003). Models of CD8+ Responses: 1. What Is the Antigen-independent Proliferation Program. *J. Theor. Biol.* 221, 585–598. doi:10.1006/jtbi.2003.3208
- Asquith, B., Deback, C., Florins, A., Gillet, N., Sanchez-Alcaraz, T., Mosley, A., et al. (2006). Quantifying Lymphocyte Kinetics *In Vivo* Using Carboxyfluorescein Diacetate Succinimidyl Ester (CFSE). *Proc. Biol. Sci.* 273, 1165–1171. doi:10.1098/rspb.2005.3432
- Banks, H. T., Sutton, K. L., Thompson, W. C., Bocharov, G., Roose, D., Schenkel, T., et al. (2011). Estimation of Cell Proliferation Dynamics Using CFSE Data. *Bull. Math. Biol.* 73, 116–150. doi:10.1007/s11538-010-9524-5
- Banks, H. T., Thompson, W. C., Peligero, C., Giest, S., Argilague, J., and Meyerhans, A. (2012). A Division-dependent Compartmental Model for Computing Cell Numbers in CFSE-Based Lymphocyte Proliferation Assays. *Math. Biosci. Eng.* 9, 699–736. doi:10.3934/mbe.2012.9.699
- De Boer, R. J., and Perelson, A. S. (2005). Estimating Division and Death Rates from CFSE Data. *J. Comput. Appl. Math.* 184, 140–164. doi:10.1016/j.cam.2004.08.020
- De Boer, R. J., Gansov, V. V., Milutinović, D., Hodgkin, P. D., and Perelson, A. S. (2006). Estimating Lymphocyte Division and Death Rates from CFSE Data. *Bull. Math. Biol.* 68, 1011–1031. doi:10.1007/s11538-006-9094-8
- Buchholz, V. R., Flossdorf, M., Hensel, I., Kretschmer, L., Weissbrich, B., Gräf, P., et al. (2013). Disparate Individual Fates Compose Robust CD8+ T Cell Immunity. *Science* 340, 630–635. doi:10.1126/science.1235454
- Costa Del Amo, P., Debebe, B., Razavi-Mohseni, M., Nakaoka, S., Worth, A., Wallace, D., et al. (2020). The Rules of Human T Cell Fate *In Vivo*. *Front. Immunol.* 11, 573. doi:10.3389/fimmu.2020.00573
- Deenick, E. K., Gett, A. V., and Hodgkin, P. D. (2003). Stochastic Model of T Cell Proliferation: A Calculus Revealing IL-2 Regulation of Precursor Frequencies, Cell Cycle Time, and Survival. *J. Immunol.* 170, 4963–4972. doi:10.4049/jimmunol.170.10.4963
- Dowling, M. R., Milutinović, D., and Hodgkin, P. D. (2005). Modelling cell lifespan and proliferation: is likelihood to die or to divide independent of age?. *J. R. Soc. Interf.* 2, 517–526. doi:10.1098/rsif.2005.0069

FUNDING

This project has received funding from the European Union's Horizon 2020 research and innovation programmed under the Marie Marie Skłodowska-Curie grant agreement No 764698. This publication has emanated from research supported in part by a research grant from Science Foundation Ireland (SFI) under Grant Number 16/RI/3399. This work was supported by the National Health and Medical Research Council of Australia (NHMRC) (Project Grant 1164800 and Investigator Grant 1176588 to P.D.H.), Victorian State Government Operational Infrastructure Support and the Australian Government NHMRC Independent Research Institutes Infrastructure Support Scheme (361646). J.M.M is supported by an NHMRC CJ Martin Fellowship. E.D.H is supported by NHMRC R.D. Wright Fellowship and project grants (1159488, 1140187, and 1165591) and grants from The Leukemia and Lymphoma Society (6552-18).

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: <https://www.frontiersin.org/articles/10.3389/fbinf.2021.723337/full#supplementary-material>

- Dowling, M. R., Kan, A., Heinzel, S., Zhou, J. H., Marchingo, J. M., Wellard, C. J., et al. (2014). Stretched Cell Cycle Model for Proliferating Lymphocytes. *Proc. Natl. Acad. Sci. U S A.* 111, 6377–6382. doi:10.1073/pnas.1322420111
- Downey, M. J., Jeziorska, D. M., Ott, S., Tamai, T. K., Koentges, G., Vance, K. W., et al. (2011). Extracting Fluorescent Reporter Time Courses of Cell Lineages from High-Throughput Microscopy at Low Temporal Resolution. *PLoS ONE* 6, e27886. doi:10.1371/journal.pone.0027886
- Duffy, K. R., and Hodgkin, P. D. (2012). Intracellular Competition for Fates in the Immune System. *Trends Cel Biol.* 22, 457–464. doi:10.1016/j.tcb.2012.05.004
- Duffy, K. R., and Subramanian, V. G. (2009). On the Impact of Correlation between Collaterally Consanguineous Cells on Lymphocyte Population Dynamics. *J. Math. Biol.* 59, 255–285. doi:10.1007/s00285-008-0231-x
- Duffy, K. R., Wellard, C. J., Markham, J. F., Zhou, J. H., Holmberg, R., Hawkins, E. D., et al. (2012). Activation-Induced B Cell Fates Are Selected by Intracellular Stochastic Competition. *Science* 335, 338–341. doi:10.1126/science.1213230
- Efron, B. (1979). Bootstrap Methods: Another Look at the Jackknife. *Ann. Statist.* 7, 1–26. doi:10.1214/aos/1176344552
- Gansov, V. V., Pilyugin, S. S., De Boer, R. J., Murali-Krishna, K., Ahmed, R., and Antia, R. (2005). Quantifying Cell Turnover Using CFSE Data. *J. Immunol. Methods* 298, 183–200. doi:10.1016/j.jim.2005.01.011
- Gerlach, C., Rohr, J. C., Perić, L., van Rooij, N., van Heijst, J. W., Velds, A., et al. (2013). Heterogeneous Differentiation Patterns of Individual CD8+ T Cells. *Science* 340, 635–639. doi:10.1126/science.1235487
- Gett, A. V., and Hodgkin, P. D. (2000). A Cellular Calculus for Signal Integration by T Cells. *Nat. Immunol.* 1, 239–244. doi:10.1038/79782
- Harris, T. E. (1963). *The Theory of Branching Processes*. Berlin: Springer-Verlag.
- Hasenauer, J., Schittler, D., and Allgöwer, F. (2012). Analysis and Simulation of Division- and Label-Structured Population Models: a New Tool to Analyze Proliferation Assays. *Bull. Math. Biol.* 74, 2692–2732. doi:10.1007/s11538-012-9774-5
- Hawkins, E. D., Turner, M. L., Dowling, M. R., van Gend, C., and Hodgkin, P. D. (2007). A Model of Immune Regulation as a Consequence of Randomized Lymphocyte Division and Death Times. *Proc. Natl. Acad. Sci. U S A.* 104, 5032–5037. doi:10.1073/pnas.0700026104

- Hawkins, E. D., Markham, J. F., McGuinness, L. P., and Hodgkin, P. D. (2009). A Single-Cell Pedigree Analysis of Alternative Stochastic Lymphocyte Fates. *Proc. Natl. Acad. Sci. U S A.* 106, 13457–13462. doi:10.1073/pnas.0905629106
- Hawkins, E. D., Turner, M. L., Wellard, C. J., Zhou, J. H., Dowling, M. R., and Hodgkin, P. D. (2013). Quantal and Graded Stimulation of B Lymphocytes as Alternative Strategies for Regulating Adaptive Immune Responses. *Nat. Commun.* 4, 2406. doi:10.1038/ncomms3406
- Heinzel, S., Binh Giang, T., Kan, A., Marchingo, J. M., Lye, B. K., Corcoran, L. M., et al. (2017). A Myc-dependent Division Timer Complements a Cell-Death Timer to Regulate T Cell and B Cell Responses. *Nat. Immunol.* 18, 96–103. doi:10.1038/ni.3598
- Hodgkin, P. D. (2018). Modifying Clonal Selection Theory with a Probabilistic Cell. *Immunol. Rev.* 285, 249–262. doi:10.1111/imr.12695
- Hoffman, M. D., and Gelman, A. (2014). The No-U-Turn Sampler: Adaptively Setting Path Lengths in Hamiltonian Monte Carlo. *J. Machine Learn. Res.* 15, 1593–1623.
- Hogquist, K. A., Jameson, S. C., Heath, W. R., Howard, J. L., Bevan, M. J., and Carbone, F. R. (1994). T Cell Receptor Antagonist Peptides Induce Positive Selection. *Cell* 76, 17–27. doi:10.1016/0092-8674(94)90169-4
- Horton, M. B., Prevedello, G., Marchingo, J. M., Zhou, J. H. S., Duffy, K. R., Heinzel, S., et al. (2018). Multiplexed Division Tracking Dyes for Proliferation-Based Clonal Lineage Tracing. *J. Immunol.* 201, 1097–1103. doi:10.4049/jimmunol.1800481
- Hyrien, O., and Zand, M. S. (2008). A Mixture Model with Dependent Observations for the Analysis of CFSE-Labeling Experiments. *J. Am. Stat. Assoc.* 103, 222–239. doi:10.1198/016214507000000194
- Hyrien, O., Chen, R., and Zand, M. S. (2010). An Age-dependent Branching Process Model for the Analysis of CFSE-Labeling Experiments. *Biol. Direct* 5, 41. doi:10.1186/1745-6150-5-41
- Jeffreys, H. (1961). *Theory of Probability*. 3 edn. Oxford: Oxford University Press.
- Kaech, S. M., Wherry, E. J., and Ahmed, R. (2002). Effector and Memory T-Cell Differentiation: Implications for Vaccine Development. *Nat. Rev. Immunol.* 2, 251–262. doi:10.1038/nri778
- Lee, H. Y., Hawkins, E., Zand, M. S., Mosmann, T., Wu, H., Hodgkin, P. D., et al. (2009). Interpreting CFSE Obtained Division Histories of B Cells *In Vitro* with Smith-Martin and Cyton Type Models. *Bull. Math. Biol.* 71, 1649–1670. doi:10.1007/s11538-009-9418-6
- Luzyanina, T., Roose, D., Schenkel, T., Sester, M., Ehl, S., Meyerhans, A., et al. (2007). Numerical Modelling of Label-Structured Cell Population Growth Using CFSE Distribution Data. *Theor. Biol. Med. Model.* 4, 26. doi:10.1186/1742-4682-4-26
- Lyons, A. B., and Parish, C. R. (1994). Determination of Lymphocyte Division by Flow Cytometry. *J. Immunol. Methods* 171, 131–137. doi:10.1016/0022-1759(94)90236-4
- Marchingo, J. M., Kan, A., Sutherland, R. M., Duffy, K. R., Wellard, C. J., Belz, G. T., et al. (2014). T Cell Signaling, Antigen Affinity, Costimulation, and Cytokine Inputs Sum Linearly to Amplify T Cell Expansion. *Science* 346, 1123–1127. doi:10.1126/science.1260044
- Marchingo, J. M., Prevedello, G., Kan, A., Heinzel, S., Hodgkin, P. D., and Duffy, K. R. (2016). T-cell Stimuli Independently Sum to Regulate an Inherited Clonal Division Fate. *Nat. Commun.* 7, 13540. doi:10.1038/ncomms13540
- Markham, J. F., Wellard, C. J., Hawkins, E. D., Duffy, K. R., and Hodgkin, P. D. (2010). A Minimum of Two Distinct Heritable Factors Are Required to Explain Correlation Structures in Proliferating Lymphocytes. *J. R. Soc. Interf.* 7, 1049–1059. doi:10.1098/rsif.2009.0488
- Marquardt, D. W. (1963). An Algorithm for Least-Squares Estimation of Nonlinear Parameters. *J. Soc. Ind. Appl. Math.* 11, 431–441. doi:10.1137/0111030
- Mazzocco, P., Bernard, S., and Pujo-Menjouet, L. (2017). Estimates and Impact of Lymphocyte Division Parameters from CFSE Data Using Mathematical Modelling. *PLOS ONE* 12, e0179768. doi:10.1371/journal.pone.0179768
- McElreath, R. (2020). *Statistical Rethinking*. 2 edn. Boca Raton, Florida: CRC Press/Taylor & Francis Group. doi:10.1201/9780429029608
- Miao, H., Jin, X., Perelson, A. S., and Wu, H. (2011). Evaluation of Multitype Mathematical Models for CFSE-Labeling experiment Data. *Bull. Math. Biol.* 74, 300–326. doi:10.1007/s11538-011-9668-y
- Mitchell, S., Roy, K., Zangle, T. A., and Hoffmann, A. (2018). Nongenetic Origins of Cell-To-Cell Variability in B Lymphocyte Proliferation. *Proc. Natl. Acad. Sci. U S A.* 115, E2888–E2897. doi:10.1073/pnas.1715639115
- Newville, M., Stensitzki, T., Allen, D. B., and Ingargiola, A. (2014). *LMFIT: Non-linear Least-Square Minimization and Curve-Fitting for Python*. [Dataset]. doi:10.5281/zenodo.11813 Available at: <https://zenodo.org/record/11813>
- Nordon, R. E., Nakamura, M., Ramirez, C., and Odell, R. (1999). Analysis of Growth Kinetics by Division Tracking. *Immunol. Cel Biol.* 77, 523–529. doi:10.1046/j.1440-1711.1999.00869.x
- Quah, B. J., and Parish, C. R. (2012). New and Improved Methods for Measuring Lymphocyte Proliferation *In Vitro* and *In Vivo* Using CFSE-like Fluorescent Dyes. *J. Immunol. Methods* 379, 1–14. doi:10.1016/j.jim.2012.02.012
- Revy, P., Sospedra, M., Barbour, B., and Trautmann, A. (2001). Functional Antigen-independent Synapses Formed between T Cells and Dendritic Cells. *Nat. Immunol.* 2, 925–931. doi:10.1038/ni713
- Sakaue-Sawano, A., Kurokawa, H., Morimura, T., Hanyu, A., Hama, H., Osawa, H., et al. (2008). Visualizing Spatiotemporal Dynamics of Multicellular Cell-Cycle Progression. *Cell* 132, 487–498. doi:10.1016/j.cell.2007.12.033
- Salvatier, J., Wiecki, T. V., and Fonnesbeck, C. (2016). Probabilistic Programming in Python Using PyMC3. *PeerJ Comput. Sci.* 2, e55. doi:10.7717/peerj-cs.55
- Schindelin, J., Arganda-Carreras, I., Frise, E., Kaynig, V., Longair, M., Pietzsch, T., et al. (2012). Fiji: an Open-Source Platform for Biological-Image Analysis. *Nat. Methods* 9, 676–682. doi:10.1038/nmeth.2019
- Shokhirev, M. N., and Hoffmann, A. (2013). FlowMax: A Computational Tool for Maximum Likelihood Deconvolution of CFSE Time Courses. *Plos One* 8, e67620. doi:10.1371/journal.pone.0067620
- Shokhirev, M. N., Almaden, J., Davis-Turak, J., Birnbaum, H. A., Russell, T. M., Vargas, J. A., et al. (2015). A Multi-Scale Approach Reveals that NF-Kb cRel Enforces a B-Cell Decision to divide. *Mol. Syst. Biol.* 11, 783. doi:10.15252/msb.20145554
- Smith, J. A., and Martin, L. (1973). Do cells Cycle?. *Proc. Natl. Acad. Sci. U S A.* 70, 1263–1267. doi:10.1073/pnas.70.4.1263
- Subramanian, V. G., Duffy, K. R., Turner, M. L., and Hodgkin, P. D. (2008). Determining the Expected Variability of Immune Responses Using the Cyton Model. *J. Math. Biol.* 56, 861–892. doi:10.1007/s00285-007-0142-2
- Turner, M. L., Hawkins, E. D., and Hodgkin, P. D. (2008). Quantitative Regulation of B Cell Division Destiny by Signal Strength. *J. Immunol.* 181, 374–382. doi:10.4049/jimmunol.181.1.374
- Vehtari, A., Gelman, A., and Gabry, J. (2017). Practical Bayesian Model Evaluation Using Leave-One-Out Cross-Validation and WAIC. *Stat. Comput.* 27, 1413–1432. doi:10.1007/s11222-016-9696-4
- Veiga-Fernandes, H., Walter, U., Bourgeois, C., McLean, A., and Rocha, B. (2000). Response of Naive and Memory CD8+ T Cells to Antigen Stimulation *In Vivo*. *Nat. Immunol.* 1, 47–53. doi:10.1038/76907
- Wagenmakers, E. J., Verhagen, J., and Ly, A. (2016). How to Quantify the Evidence for the Absence of a Correlation. *Behav. Res. Methods* 48, 413–426. doi:10.3758/s13428-015-0593-0
- Watanabe, S., and Opper, M. (2010). Asymptotic Equivalence of Bayes Cross Validation and Widely Applicable Information Criterion in Singular Learning Theory. *J. Machine Learn. Res.* 11, 3571–3594.
- Wellard, C., Markham, J., Hawkins, E. D., and Hodgkin, P. D. (2010). The Effect of Correlations on the Population Dynamics of Lymphocytes. *J. Theor. Biol.* 264, 443–449. doi:10.1016/j.jtbi.2010.02.019
- Wellard, C., Markham, J. F., Hawkins, E. D., and Hodgkin, P. D. (2011). *Mathematical Models and Immune Cell Biology*. New York: Springer. doi:10.1007/978-1-4419-7725-0
- Yates, A., Chan, C., Strid, J., Moon, S., Callard, R., George, A. J., et al. (2007). Reconstruction of Cell Population Dynamics Using CFSE. *BMC Bioinf.* 8, 196. doi:10.1186/1471-2105-8-196
- Yates, C. A., Ford, M. J., and Mort, R. L. (2017). A Multi-Stage Representation of Cell Proliferation as a Markov Process. *Bull. Math. Biol.* 79, 2905–2928. doi:10.1007/s11538-017-0356-4
- Zhou, J. H. S., Markham, J. F., Duffy, K. R., and Hodgkin, P. D. (2018). Stochastically Timed Competition between Division and Differentiation Fates Regulates the Transition from B Lymphoblast to Plasma Cell. *Front. Immunol.* 9, 2053. doi:10.3389/fimmu.2018.02053

Zilman, A., Ganusov, V. V., and Perelson, A. S. (2010). Stochastic Models of Lymphocyte Proliferation and Death. *Plos One* 5, e12775. doi:10.1371/journal.pone.0012775

Conflict of Interest: Author SJD is employed by AstraZeneca.

The remaining authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Publisher's Note: All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of

the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

Copyright © 2021 Cheon, Kan, Prevedello, Oostindie, Dovedi, Hawkins, Marchingo, Heinzl, Duffy and Hodgkin. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.